UNIL | Université de Lausanne

Unicentre
CH-1015 Lausanne
http://serval.unil.ch

Year : 2018

The impact of mode of data collection on measures of subjective wellbeing

Sánchez Tomé Rosa

Sánchez Tomé Rosa, 2018, The impact of mode of data collection on measures of subjective wellbeing

Originally published at : Thesis, University of Lausanne

Posted at the University of Lausanne Open Archive http://serval.unil.ch
Document URN : urn:nbn:ch:serval-BIB_F89D8660FBE78

Droits d'auteur
L'Université de Lausanne attire expressément l'attention des utilisateurs sur le fait que tous les documents publiés dans l'Archive SERVAL sont protégés par le droit d'auteur, conformément à la loi fédérale sur le droit d'auteur et les droits voisins (LDA). A ce titre, il est indispensable d'obtenir le consentement préalable de l'auteur et/ou de l'éditeur avant toute utilisation d'une oeuvre ou d'une partie d'une oeuvre ne relevant pas d'une utilisation à des fins personnelles au sens de la LDA (art. 19, al. 1 lettre a). A défaut, tout contrevenant s'expose aux sanctions prévues par cette loi. Nous déclinons toute responsabilité en la matière.

Copyright
The University of Lausanne expressly draws the attention of users to the fact that all documents published in the SERVAL Archive are protected by copyright in accordance with federal law on copyright and similar rights (LDA). Accordingly it is indispensable to obtain prior consent from the author and/or publisher before any use of a work or part of a work for purposes other than personal use within the meaning of LDA (art. 19, para. 1 letter a). Failure to do so will expose offenders to the sanctions laid down by this law. We accept no liability in this respect.

UNIL | Université de Lausanne

FACULTÉ DES SCIENCES SOCIALES ET POLITIQUES

INSTITUT DES SCIENCES SOCIALES

The impact of mode of data collection on measures of subjective wellbeing

THÈSE DE DOCTORAT

présentée à la

Faculté des sciences sociales et politiques
de l'Université de Lausanne

pour l'obtention du grade de
Docteure ès sciences sociales

par

Rosa Sánchez Tomé

Directrice de thèse
Professeure Caroline Roberts

Co-directeur de thèse
Professeur Dominique Joye

Jury de thèse
Professeur André Berchtold, Professeur à l'Université de Lausanne
Professeure Annette Jäckle, Professeure à l'Université d'Essex
Professeur Michel Oris, Vice-recteur à l'Université de Genève

LAUSANNE
2018

UNIL | Université de Lausanne

FACULTÉ DES SCIENCES SOCIALES ET POLITIQUES

INSTITUT DES SCIENCES SOCIALES

The impact of mode of data collection on measures of subjective wellbeing

THÈSE DE DOCTORAT

présentée à la

Faculté des sciences sociales et politiques
de l'Université de Lausanne

pour l'obtention du grade de
Docteure ès sciences sociales

par

Rosa Sánchez Tomé

Directrice de thèse
Professeure Caroline Roberts

Co-directeur de thèse
Professeur Dominique Joye

Jury de thèse
Professeur André Berchtold, Professeur à l'Université de Lausanne
Professeure Annette Jäckle, Professeure à l'Université d'Essex
Professeur Michel Oris, Vice-recteur à l'Université de Genève

LAUSANNE
2018

UNIL | Université de Lausanne
Faculté des sciences
sociales et politiques

IMPRIMATUR

Le Décanat de la Faculté des sciences sociales et politiques de l'Université de Lausanne, au nom du Conseil et sur proposition d'un jury formé des professeurs

- Caroline ROBERTS, directrice de thèse, Professeure à l'Université de Lausanne
- Dominique JOYE, co-directeur de thèse, Professeur à l'Université de Lausanne
- André BERCHTOLD, Professeur à l'Université de Lausanne
- Annette JÄCKLE, Professeure à l'Université d'Essex
- Michel ORIS, Vice-recteur à l'Univesité de Genève

autorise, sans se prononcer sur les opinions de la candidate, l'impression de la thèse de Madame Rosa SANCHEZ TOME, intitulée :

« The impact of mode of data collection on measures of subjective wellbeing »

Jean-Philippe LERESCHE
Doyen

Lausanne, le 2 février 2018

Abstract
Measures on quality of life are now widely available and questions such as life satisfaction and happiness are common in social surveys, providing invaluable information to researchers and policy makers across the world. However, surveys are subject to error and it is important to evaluate the effect of survey design on data quality. In particular, some survey methodologists have focused on the mode of data collection, as there are concerns about nonresponse bias in traditional single-mode surveys. As a consequence, alternative ways of collecting information have become popular, such as the combination of multiple modes of data collection (e.g. telephone and web) as a way of attracting people that would otherwise not respond. However, collecting data using different modes may render data incomparable, which could pose a problem for those researchers that rely on data collected using different modes.

This thesis attempts to provide some answers about the potential drawbacks of using data on the topic of wellbeing that come from different modes of data collection. Analysing data from a mode comparison experiment implemented in Switzerland, I compared responses to twenty-seven subjective wellbeing questions in telephone, web and mail. The results of the study demonstrate that mode has an effect on who responds and how responses are given. Mode affected measures such as happiness and job satisfaction, responses to open-ended questions about life events, and the relationship between subjective wellbeing and its predictors. However, the latent measure of general subjective wellbeing was equivalent across modes.

Resumé
Des mesures sur la qualité de vie sont maintenant largement disponibles et des questions telles que la satisfaction de la vie et le bonheur, qui fournissent des informations inestimables aux chercheurs et aux décideurs politiques à travers le monde, sont courantes dans les enquêtes sociales. Cependant, les enquêtes sont sujettes à des erreurs et il est important d'évaluer l'effet de la façon dont l'enquête est conçue sur la qualité des données. Certains méthodologistes d'enquête se sont concentrés sur le mode de collecte des données étant donné qu'il existe des préoccupations concernant les biais de non-réponse dans les modes d'enquêtes traditionnelles En conséquence, d'autres méthodes de collecte d'informations sont devenues populaires, telles que la combinaison de plusieurs modes de collecte de données (par exemple, le téléphone et le Web) pour attirer des personnes qui autrement ne répondraient pas. Cependant, la collecte de données en utilisant différentes méthodes peut rendre les données incomparables, ce qui pourrait poser un problème aux chercheurs qui s'appuient sur des données collectées en utilisant différentes méthodes.

Cette thèse tente de fournir quelques réponses sur les inconvénients de l'utilisation de données sur le thème du bien-être provenant de différentes méthodes de collecte de données. Analysant les données d'une expérience de comparaison de méthode mise en œuvre en Suisse, j'ai comparé les réponses à vingt-sept questions subjectives sur le bien-être posées par téléphone, sur le Web et par courrier. Les résultats de l'étude démontrent que la méthode a un effet sur qui répond et comment les réponses sont données. La méthode a affecté des mesures telles que le bonheur et la satisfaction au travail, les réponses aux questions ouvertes sur les événements de la vie, et la relation entre le bien-être subjectif et ses indicateurs. Cependant, la mesure latente du bien-être subjectif général était équivalente d'une méthode à l'autre.

Acknowledgments

There are many people I would like to thank (or blame), without whom I would have never started nor being on my way to finish this thesis.

First of all, thank you LIVES and the SNF for their funding, and to the respondents that filled the survey.

Special thanks to Caroline Roberts and Dominique Joye for trying so hard to make of me a competent researcher, and Michele Ernst Stähli for her support and helpful comments during these 4 years and her collaboration with the survey experiment, all of you made my PhD work much easier that it could have been. I would also like to thank Jon Burton, Peter Lynn and Nick Allum: I wouldn’t have come to Lausanne without your encouragement and support.

Thanks also to Annette Jäckle, André Berchtold and Michel Oris, for taking the time to read my thesis and for your helpful comments and suggestions on a previous version of this thesis.

Thank you to all those researchers that provided food for thought during the LIVES’ and LINES’ doctoriales, the LINES’ and the different conferences I have been to, particularly to Davide Morselli for his comments on one of my chapters. Also to LIVES, LINES, and FORS friends and colleagues, for providing a great and friendly atmosphere and enough parties to survive this PhD process: Romina, Gina, Emanuele, Emilio, Hannah, every Anna, Florence, Minja, Robert, Matteo, Pablo, Sara, Nora, Andrés, Claudia, Gaëlle, Nicolas, and many more. I am particularly grateful to those who I’ve shared an office with: Ana, who introduced me to everyone in LIVES when I arrived, lost and without a place to stay; to Maïlys, Rafael, Isabel, Uta, and Patrick- I would have probably given up without your support- and definitely I would have never learnt French without you! Thanks also to my Spanish-speaking friends in Lausanne (Lorena, Paula, Clara, Peter, Florian, you helped me a lot too) and to one of my best friends in Lausanne, Suzanne, an unexpected friend with whom I've spent great moments and beer and shared our interest in languages, books, travelling and complaints about the work life. I am also thankful to my friends in Béjar, my hometown (Mada, Sara, Virgi, Sheila, Patri H., Patri S., Teresa), and Miren, Eva, Ana and Celia: you all increased my level of wellbeing during various holidays and weekends, which helped me to get this thesis finished.

And I cannot avoid thanking my parents, Rosa and Javier, for their unconditional support (both psychological and financial) during every year of my life- I am particularly amazed at my mum's capacity to make everything seem easy to do and not give importance even to those events and situations in which I felt I could not keep going. This involves both getting a thesis written, but also my wisdom teeth removed. I also want to thank my grandparents Maria Dolores and Antonio, for being supportive and inspiring.

And especially to Ben Hurrell, I just would not be alive without you, as I would have died after eating too many pot noodles.

Table of Contents

List of tables

List of figures

CHAPTER 1. INTRODUCTION

Quality of life indicators have become increasingly valuable in the last fifty years (Diener, 1984; Ferrer-i-Carbonell & Frijters, 2004). This trend has been facilitated by the consensus that traditional economic measures such as Gross Domestic Product (GDP) do not satisfactorily explain societal progress (Paul Dolan & Kavetsos, 2012; Kubiszewski et al., 2013). Accordingly, politicians, policy makers and academics have shown great interest in learning about the way in which the population perceives their lives (Fleche, Smith, & Sorsa, 2012), and have found that subjective indicators of life quality are useful for their tasks related to decision making and policy evaluation. Particularly, governments across the world aim to find ways to tackle the problems and challenges endured by different subgroups of their populations, such as the elderly, women, children, immigrants, or the working population (Stone et al., 2013). For this reason, comparisons of how people evaluate their lives, both across countries (OECD, 2015) and across different population subgroups within the same country (Stone, 2012) are widely used.

However, the field of wellbeing and quality of life studies has been marred by difficulties with regard to reaching an agreement across disciplines, such as economics, environmental studies, psychology and sociology, in defining what quality of life is (Bahadur, Ibrahim, & Tanner, 2010). The lack of consensus in this area stems from how to measure, collect, and analyse the data (Sumner & Mallett, 2011). In fact, the debate has generated a large number of journal articles, books and reports dedicated to concepts such as happiness, vulnerability and wellbeing (Patrick, Wild, Johnson, Wagner & Martin, 1994; Gough & McGregor, 2007; Diener, 2009; Dodge, Daly, Huyton, & Sanders, 2012; Fleche et al., 2012; OECD, 2013).

These studies about quality of life are often multidisciplinary and use the concept of subjective wellbeing as a multidimensional measure that reunites such aspects of people's lives that relate to feelings, social life and support, income, work, ability to overcome life difficulties, or environmental context. Throughout this thesis, someone with a high level of subjective wellbeing is conceived as a person who is satisfied with their life, who finds pleasure and enjoyment in life, and who rarely experiences "unpleasant emotions such as sadness or anger" (Diener, Suh, & Oishi, 1997: 25).

Currently, researchers and policy makers often work with information collected through large scale surveys – such as the World Values Survey, the Gallup World Poll, the European Social Survey, the German Socio-Economic Panel, or the Quality of Life in Europe project by Eurostat – that include measures related to happiness and life satisfaction (Kapteyn, Lee, Tassot, Vonkova, & Zamarro, 2015). These studies may also include information on other aspects of wellbeing, such as measures of anxiety, loneliness or life achievements. In Switzerland, national surveys such as the Swiss Household Panel, the MOSAiCH and the Swiss Labour Force Survey provide information on subjective wellbeing (FORS, 2016).

The important role that measures of subjective wellbeing can have in research and policy-making explains the abundant deliberation about the best way of defining, measuring, collecting and analysing information related to both vulnerability and wellbeing (Sumner & Mallett, 2011). However, there has been little research dedicated to examining the impact that survey design has on the information obtained (Conti & Pudney, 2011a). This question is of particular relevance at a moment in time when traditional surveys face important challenges in their ability to obtain good quality data. These challenges include decreasing response rates (especially for some groups of the population), growing public distrust towards surveys, increases in the costs associated with implementing surveys (Hox, de Leeuw, & Zijlmans, 2015; Dillman, 2017) and difficulties gathering information by the traditional means of data collection designs, such as telephone, face-to-face or mail (Dillman et al., 2009). In Switzerland, where telephone surveys have traditionally been the principal means of collecting data (Joye, Pollien, Sapin, & Ernst Stähli, 2012), it is particularly worrying that some groups of the population, for example foreigners, are underrepresented in telephone surveys (Lipps, 2016).

In this context, there is a debate surrounding how to obtain good quality data without raising costs (Tourangeau, 2017). Out of the different options that can help manage these challenges, survey designs based on mixing different modes of data collection are growing in popularity (de Leeuw, Hox, & Dillman, 2008). This type of design involves survey participants responding to questions using different response modes (de Leeuw, 2005): for example, respondents may answer by web, face-to-face or telephone. Using different modes of data collection can help reach different types of people, improving the coverage and representativeness of the population, because the choice of mode determines to some extent who can be reached and who decides to participate. This effect is referred to as the mode selection effect, whether it be due to possibility, or willingness (Bowling, 2005). For instance, an elderly person without the Internet might be more willing and available to respond by telephone than in a different mode. However, mixing modes can also create problems because mode of data collection also has an effect on how questions are answered (de Leeuw, 2005). This is known as the mode measurement effect. This type of effect can be partly avoided at the design stage of the survey by designing equivalent questionnaires (Dillman, Smyth, & Christian, 2014), but it can also be inherent to the different modes' characteristics (Hox, de Leeuw, & Zijlmans, 2015). The presence or answer of an interviewer – whether the question is presented visually or aurally – together with differential cognitive efforts necessary to respond to questions in different modes (Martin & Lynn, 2011), mean that different modes are associated with different types of error (Revilla, 2010). Both selection and measurement mode effects mean that survey estimates may differ due to the mode of data collection, and they are a cause of concern that worry survey researchers with regard to the comparability of such estimates. While selection effects are often sought after (in the case of mixed-mode surveys, because of the potential to improve the representativeness of survey samples) and easier to control for if they relate to socio-demographic characteristics, estimates can be affected by both effects confounded, complicating the task of deciding whether the combination of data that come from different modes is recommendable (Vannieuwenhuyze & Loosveldt, 2013). The situation gets even more complicated as different types of respondent, who already have different levels of probability to participate in certain modes, may be more or less susceptible to different sources of measurement error. Differences in age, education, or sex, have been previously identified as certain groups that appear to give responses of different quality (Revilla, 2012). Underreporting certain opinions or behaviours, or having difficulties understanding the questions because of language barriers or reading comprehension may well affect some sectors of the population more than others. If different types of respondents are attracted to different modes, and they react differently giving answers to different modes, the results from comparisons across the population, and from analyses of particular populations that are of interest to subjective wellbeing (such as women, the unemployed, or the elderly) could be negatively affected.

Measures of subjective wellbeing are, like all survey measures, subject to measurement error. Mood and interpersonal relations can influence responses, and respondents can feel that the questions are intrusive to their lives, which means that the context in which the survey is being responded to and the mode of data collection may affect the responses. Researchers such as Conti and Pudney (2011) or Sarracino, Riillo and Mickucka (2017) have previously looked at the impact of the mode of data collection in measures of subjective wellbeing and found contradictory results on the equivalence of subjective wellbeing data across modes. Moreover, mode effects in individual items of subjective wellbeing have also been found to impact the relationship between the variables and their predictors (Holford & Pudney, 2015a).

By combining substantive research on wellbeing with survey methodological research on mode effects, this thesis aims to contribute to the debate about the importance of survey design for measuring quality of life by studying the impact that different types of survey error may have on the study of wellbeing and vulnerability, paying special attention to the comparability of data across modes in the field of quality of life research. Investigating the effect of mode in measures of subjective wellbeing is of particular interest in the context in which this thesis was developed, the Swiss National Science Foundation research project called 'LIVES – Overcoming Vulnerability: Life Course Perspectives', hosted by the Universities of Lausanne and Geneva, in which researchers from multiple disciplines study the life course of people in Switzerland and elsewhere (Oris, Roberts, Joye, & Ernst Stähli, 2016), often focusing on specific parts of the population such as single mothers (Struffolino, Voorpostel, & Bernardi, 2015), foreigners (Galhano, 2016), the elderly (Ihle et al., 2015) or young people (Brändle, 2017).

Although both quantitative and qualitative research methods are used in LIVES, and often combined, an important basis for vulnerability and wellbeing research is survey data (Oris et al., 2016). Some of the LIVES research uses data from national surveys such as the Swiss Household Panel, while other researchers design their own survey instruments to collect their own information in LIVES (Knecht, Wiese & Freund, 2016; Paggi, Jopp & Hertzog, 2016; Ruch, Martínez-Martí, Heintz, & Brouwers, 2014).

For the present thesis, the data analysed come from a methodological experiment implemented by LIVES in collaboration with FORS (Swiss Centre of Expertise in the Social Sciences) in 2012 and 2013, designed to test how different modes of data collection (web, mail and telephone), used on their own or in combination in mixed-mode designs, influence a series of measures of subjective wellbeing. Motivated by the findings of previous research on the topic of subjective wellbeing and survey methodology, this study sought to answer three research questions:

RQ1. Do different modes of data collection differentially affect the quality of survey estimates of subjective wellbeing?

RQ2. Do mode effects on measurement affect all respondents equally?

RQ3. Do mode effects on measures of subjective wellbeing impact the results of substantive research into the predictors and correlates of subjective wellbeing measures?

After this chapter, I introduce the main concepts and ideas that are important to understand the remainder of the thesis, combining the topics of survey methodology and subjective wellbeing and discussing some of the main challenges that had to be confronted in the research undertaken. Afterwards, in chapter 3, I describe the data that I use, and the design of the methodological experiment from which they came. From chapter 4 onwards, the thesis is composed of five empirical studies aimed at answering the three main research questions.

The first research question consists of investigating the effect of mode in a series of items commonly used to measure subjective wellbeing. Because I look at both closed-ended questions and open-ended questions, for which I test for mode effects in different ways looking at different estimates and using different methodologies, I divided the research into two independent studies. Study 1 consists of the study of the extent of mode effects in closed-ended questions about different aspects of wellbeing (from general wellbeing measures, such as happiness and life satisfaction, work satisfaction, social support to measures of negative emotions). The focus of the research is to identify the presence of mode effects and its sources, by aiming to disentangle selection from measurement effects. Study 2 looks at different quality indicators and responses to open-ended questions about important life events which have been used to measure risk factors of vulnerability: response length, item-nonresponse, and the theme and positivity of the answers are examined. As there are concerns about mode effects not affecting everyone in the same way (Conti & Pudney, 2011; Revilla, 2012), for example, respondents with different levels of education or different ages may give responses of different quality depending on response mode, Study 3 involves looking at the interaction between mode and demographic characteristics in responses to the subjective wellbeing questions, in order to investigate the extent to which population subgroup comparisons may be affected by mode effects. The last two studies focus on the extent to which the observed mode effects affect results of common analyses of subjective wellbeing, that is, whether mode effects matter for the types of substantive research typically undertaken with such data. As different researchers use different analytical techniques, often dependent on available measures of wellbeing, I compare the results from various types of statistical analysis across modes. In Study 4, I present the analysis of the relationship between individual subjective wellbeing measures and their predictors, to see whether such relationships differ across modes, replicating common types of analyses used by researchers using quality of life indicators. The last empirical chapter, Study 5, consists of an examination of the results from more complex analyses that involve the use of multi-dimensional latent measures of subjective wellbeing. Lastly, the final chapter summarizes the main findings of the thesis and discusses its contributions, particularly for users of survey data about subjective wellbeing, and the study limitations.

CHAPTER 2. THEORETICAL FRAMEWORK: SURVEY METHODS AND SUBJECTIVE WELLBEING RESEARCH

2.1. Introduction

Surveys are the most commonly used method for collecting information in social science research (De Vaus, 2012). Using interviews and questionnaires distributed to a part of the population of study, researchers are able to obtain data about people's attitudes, feelings or behaviours (Saris & Revilla, 2016, p. 1006).

The information collected from a survey is used to create indicators (Groves et al., 2009), known as *statistics,* that summarize the information obtained to facilitate its interpretation. Statistics can be *descriptive*, when they summarize data – for example the proportion of participants that have one good friend – or *inferential,* to make predictions about a whole population based on a sample of the population, e.g., predicting the percentage of people that have one good friend in a whole country based on the information reported in a survey (Agresti & Finlay, 2009).

The quality of the inferences based on sample survey data is related to the precision of the responses to the survey questions (Groves et al., 2009) and to how well the sample of the population responding the questions represents the population (Agresti & Finlay, 2009). In reality, all statistics are subject to *survey error*, potentially failing to accurately describe the population of study (Agresti & Finlay, 2009). For this reason, the objective when implementing survey research is to minimize survey error as much as possible. Survey error is inherent to surveys (de Leeuw et al., 2008) but it is not necessarily a cause of concern unless survey error is not random, but systematic, also known as bias (Groves et al., 2009; Niemi, 1993). I will explain the differences between these two types of error in the next section.

The field of survey methodology focuses greatly on the concept of survey error, assessing how different factors associated to the different survey implementation steps – the design of the survey research tools, the data collection process and the analysis of the data – can provoke survey error and impact statistics (Groves et al., 2009), and providing advice on how to describe the population as precisely as possible with the available resources. The objective in survey research design is to minimize the mean square error (MSE), in other words, to provide precise estimates (de Leeuw, Hox & Dillman, 2008). Survey methodology, though, has a wider scope and alternative indicators of survey quality, such as factors related to the survey question's importance, convenience in terms of time and ease of access to the data, consistency and the possibility to contrast and compare the information to other surveys, or costs (Groves et al., 2009; Groves & Lyberg, 2010). In spite of the various possibilities to study survey quality, in this thesis I focus on the concept of survey error.

2.2. Total survey error and the survey process

Survey error is, therefore, associated to the decisions made about the survey process (see figure 1). This means that there are different types of error that can affect survey estimates at the same time. Some of these errors are related to the way in which the survey was designed and can be minimised by researchers, and others have their origin in external sources that cannot be controlled for at the design stage.

The aggregate of the different types of survey error is called *total survey error* (TSE). The total survey error concept was developed to help understand the concept of data quality in surveys and it is related to how well the survey indicators describe the population. Simply put, it can be defined as the difference between the survey estimate – or statistic – and the true value of the population characteristic that is measured through the survey (Biemer & Lyberg, 2003). To evaluate the amount of error in a survey, it is theoretically possible to use statistical techniques that indicate the components of total survey error. However, this is rarely possible as it would be necessary to have “an estimate of the parameter that is essentially error free” (Biemer, 2010, p. 826).

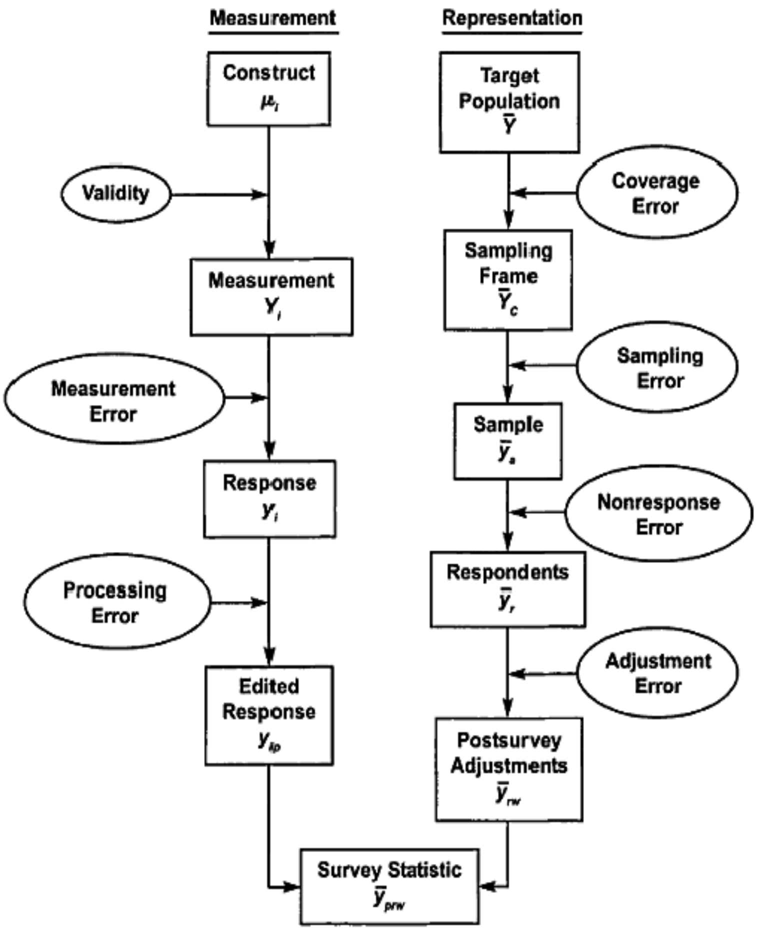


Figure 1. Survey life cycle from a quality perspective (Groves et al., 2009, p. 48).

The figure above shows the way in which survey estimates are obtained from the population of study and the different types of error that may occur during each stage. The whole set of errors are part of the 'total survey error' (see figure 1).

In figure 1, the first column corresponds to the measurement aspect and the information collected through the survey. The figure starts with the concept of construct validity, which relates to how well the question measures the theoretical construct of interest (de Leeuw, Dillman & Hox, 2007). The measurement error appears when "a respondent's answer to a question is inaccurate, departs from the 'true' value" (de Leeuw et al., 2007, p. 7). The last type of error in the figure affecting the measurement of the survey statistic is the processing error, which consists of errors in the editing of the data, introducing the data into the computer, the coding, the application of survey weights and “tabulation of survey data” (Biemer, 2010, p. 824).

The second column refers to the representation of the population of study in the survey and nonobservation. The coverage error happens when there are members of the population that have no probability of being chosen to participate in the survey, while the sampling error is due to the fact that only some people from the total sample frame are selected to participate. Finally, nonresponse error appears when there is a failure to collect data from some of the sampling units that were selected to participate in a survey. It can also result from when a respondent participates but fails to answer to a particular item of the questionnaire.

There are alternative ways of grouping these types of errors but they are often classified under the names of *sampling error* and *non-sampling error* (Biemer & Lyberg, 2003). Sampling, noncoverage, and nonresponse errors (see figure 1) are also known as nonobservation errors. Measurement error is a type of observation error, and refers to the discrepancy between the “true” response and the survey response (Dillman, 1991).

In this thesis I focus on non-sampling errors. Biemer and Lyberg (2003) identify five components of this type of error: *specification error, frame or coverage error, nonresponse error, measurement error* and *processing error*. Within these, measurement and nonresponse error have received particular attention (Dillman et al., 2009; Groves et al., 2009; de Leeuw et al., 2008).

2.2.1. Nonresponse error

Nonresponse results when someone refuses to answer, when he or she could not be contacted or was unable to answer, for example due to illness, due to a survey process mistake or to not being able to speak the questionnaire language, among other possibilities. Nonresponse may also happen when someone answered but the questionnaire was lost (de Leeuw et al., 2007; Groves & Peytcheva, 2008). When nonresponse error is systematic, it is called nonresponse bias or selection bias. When nonrespondents are significantly different to respondents, their characteristics will also be underrepresented (de Leeuw et al. 2008). This situation can lead to survey estimates that hinder the process of extrapolation that allows the description of the whole population based on the subset that answered the survey.

2.2.2. Measurement error

Within the total survey error concept, measurement error is "the observational gap between the ideal measurement and the response obtained" (DeCastellarnau, 2017, p. 2). As with nonresponse error, systematic deviations in responses could lead to biased survey estimates. An example of this would be if some respondents systematically interpreted a question in different ways.

Measurement error can result from a combination of the design of the questionnaire, the mode of administration, and the characteristics of the respondent, factors which influence the cognitive processes involved in responding to questions (Groves et al., 2009; de Leeuw, 2005). Within the sources of measurement error, previous studies have found that the questions' characteristics are a particularly important source of measurement error (DeCastellarnau, 2017; Saris & Gallhofer, 2007).

Understanding the question, being able to give a response, and doing it honestly are key aspects that affect the level of measurement error in answers to surveys (Collins, 2003). From the cognitive perspective – and related to the capacity of the respondent to answer the questions – the respondent has to take a series of steps to answer a question: interpreting what the question is asking from him or her, retrieving the information related to the question, making a judgment on how to give this information, and finally formulating the answer (Groves et al., 2009; Tourangeau, Rips & Rasinski, 2000).

Dillman (1991) summarizes the sources that can generate measurement error: respondents not being able or not wanting to provide truthful information, questions phrased in a way that does not allow the choosing of the correct response option, or the order in which questions are presented. Such sources of measurement error can be prompted by the content – for example, social desirability due to a sensitive topic that people do not like talking about – or by the format of the question, such as the number of response alternatives available: if there are many options, for example, respondents in telephone surveys may choose the option they remember rather than the most accurate one (Roberts, 2016).

When investigating the impact of survey design on measurement bias, previous research has mainly focused on two ways of responding: satisficing and social desirability bias (Holbrook, Green & Krosnick, 2003). In fact, some authors argue that there is an overlap between social desirability and satisficing response styles that can be difficult to distinguish in response scales (Turner, Lessler & Gfroerer, 1992). However, while satisficing appears to be closely linked to the question format, social desirability is related to the content of the questions.

Satisficing response style happens when the respondent does not give carefully considered answers (Holbrook, Green, & Krosnick, 2003; Krosnick, 1991) but makes the minimum effort possible to answer a question, affecting the quality of the given information. There are different types of satisficing according to the strength of the response style. The two strongest types of satisficing involve selecting a no-opinion response category and non-differentiation, that is to say, selecting the same response alternative when responding a battery of questions, also known as "straight-lining" (Holbrook et al., 2003). In addition, respondents may select the first "reasonable" answer without attempting to find the optimal option (Barge & Gehlbach, 2012). Overall, and to summarize, satisficing is related to the fact that respondents might make little mental effort in responding to questions: not attempting to recall, or not paying attention to all the response alternatives and therefore potentially missing relevant information (Bertrand & Mullainathan, 2001a).

Acquiescent responding is another responding tendency that has been identified as a type of satisficing (Holbrook et al., 2003; Krosnick, 1991; Krosnick, Villar & MacInnis, 2011). It involves agreeing with a statement without taking into account its content (Holbrook et al., 2003). Acquiescent responding, however, can also be related to respondents finding the action of disagreeing very unpleasant, or to respondents perceiving that they should "defer to individuals of higher social status" (Krosnick & Presser, 2010, p. 275).

On the other hand, socially desirable responding is an intentional response style that consists of respondents distorting the information so that they can offer a positive image of themselves (Holbrook et al., 2003). It is therefore related to the question content, especially occurring when the questions implicate giving information on sensitive, personal topics (de Leeuw et al., 2008). However, sensitive questions are not the only affected ones: "even ordinary questions that seem on the surface to have little social desirability consistently exhibit this effect" (Dillman et al., 2009, p. 313).

Studies on social desirability (Davis, Thake & Vilhena, 2010; DeMaio, 1984; Paulhus, 1984; Paulhus & Reid, 1991; Tourangeau & Yan, 2007) identify different approaches to the concept: either as a response strategy, as a personality characteristic or as a characteristic specific to certain questions. In addition, Paulhus (2002) describes different types of social desirability depending on whether they involve exaggerating personal and social status or exaggerating moral qualities. Previous studies show that the construction of a favourable image is inherent to social interactions (Holbrook et al., 2003), and for this reason, social desirability is linked to the fact that an interview is a type of social interaction (Sudman & Bradburn, 1974), with survey respondents perceiving the necessity to behave according to the social norms and values that they use in other social interaction settings. This means that respondents control the image they give and present a positive version of themselves, leading to deviations from the truth in the information collected (Collins, 2003). This tendency can even be an unconscious process to which people are accustomed in everyday life situations, such as when looking for a job or trying to avoid arguments (Tourangeau & Yan, 2007). For this reason, it has been shown that survey participants are more willing to reveal socially undesirable information when they perceive a high level of anonymity, or when respondents think that researchers can access to the same information through other means (Holbrook et al., 2003)

2.2.3. Is it possible to measure the "true" value of subjective survey data?

By using the concept of measurement error and measurement bias, there is the assumption that a 'true' value exists which can be different to the answer given by the respondent. However, difference between the true and the collected information can also be caused involuntarily by the respondent misunderstanding a question or by the impossibility of accurately recalling the information a question is asking for (Collins, 2003). In the case of opinions, evaluations and subjective information in general, it is even more difficult to know what is an accurate response – or the "truth" – is, as questions may ask for information that the respondent had not thought about before and therefore does not have a formed opinion. In fact, attitudes "may not exist in a coherent form" (Bertrand & Mullainathan, 2001b, p. 4): they may not be stable over time, responses may even be affected by cognitive dissonance as respondents report as fact what may not be consistent with the "truth".

In addition, the answers given by respondents can change depending on the context in which they are given, for example on the interview or questionnaire response situation. Indeed, "contextual factors such as norms of politeness, expectations of what actions the person asking might take given a certain answer" (Smyth, Dillman & Christian, 2008, p. 1) affect the way people answer questions. An illustration of this is how such context effects restrict the usefulness of self-reported subjective measures, such as subjective wellbeing (Deaton & Stone, 2013), or modify responses depending on which questions have been previously asked (Tourangeau & Smith, 1996b): Deaton and Stone, for example, found that wellbeing reports depend on the previous questions being asked. Context effects can also relate to the presence of an interviewer, or another person present while completing the survey.

Questions that ask for subjective information complicate the task of evaluating data quality, although this may be done by examining at response latencies or response times. Yan (2015) explains that not only is it difficult for researchers to evaluate the quality of attitudinal data, but it is also difficult for respondents to answer such questions. He gives the example of a question measuring life satisfaction: the respondent has to understand the question, then they have to think over and recall all relevant aspects of their lives, give a judgment based on the information they have retrieved, and translate the information into the response alternatives offered. In fact, responses can be vague and the respondent has to decide what it means to be "somewhat satisfied" or "not very satisfied". In this complex process, there is the possibility of things going wrong: misunderstandings, failing to retrieve the information, inability to adequately make an evaluation about the information they could retrieve or map their response into the response categories, and so on. All of this process requires cognitive effort and capacity and willingness that can make the respondent satisfice and choose a not-so-optimal response, if not skip the question completely.

Depending on the type of question, some respondents can find survey questions difficult because they have conflicting ideas on the matter. While some people find it easy to form consistent ideas about their opinions and attitudes, asking other respondents to express a single coherent position can lead them to give a random answer (Bem, 1972). This may depend on the question format and the set of categories available, as some respondents may feel represented by more than one response category and not feel able to choose.

2.3. The importance of the mode of data collection

Data from surveys can be gathered in many different ways. Although some decades ago most surveys involved "traditional" modes of data collection such as face-to-face, telephone or mail, nowadays new ways of collecting the information such as using the Internet or mobile phone surveys – or a combination of multiple modes, both new and traditional – are increasingly popular, (Dillman, 2017; Mohorko, de Leeuw & Hox, 2013). However, these modes have different characteristics that can lead to different “coverage, sampling, non response, and measurement errors” (Melanie A. Revilla & Saris, 2013, p. 242). These characteristics that distinguish the different modes have a differential impact on the way respondents understand, interpret and respond to the questions (Esposito & Jobe, 1991) leading to differences in responses. The differences in the information obtained from the survey due to mode can be due to different sample compositions and ways of responding, and are known as a *mode effects* (de Leeuw, 2005).

For example, an important characteristic is whether there is an interviewer involved in gathering the data or not. Although their presence can improve the survey response rate, researchers such as Kreuter, Presser and Tourangueau (2008) show how the use of interviewers results in a larger amount of social desirability bias compared to when the questionnaire is self-completed. Another type of mode-effect may arise caused by the difference between hearing and reading the questions: auditory compared to visual transmission of data. Telephone surveys are normally carried out by interviewers that read the questions along with the response options, while respondents in mail or Internet surveys read both questions and response options themselves, requiring different response processes (Dillman & Christian, 2005). Moreover, Krosnick and his colleagues (2012) found that respondents express their non-opinion to a larger extent in private questionnaires than in questionnaires that are completed orally with an interviewer. Hope and colleagues (2014) describe how there were response differences in a series of agree/disagree questions and that acquiescence was less prominent in the web mode, while telephone respondents were more likely to choose the last response option instead of the optimal choice. However, web respondents tended to satisfice by choosing the mid-point options to a higher extent than telephone ones.

Previous studies found differences between the interviewer-based modes and self-completed modes in terms of social-desirability bias (Hope, Campanelli, Nicolaas, Lynn & Jäckle, 2014). The presence or absence of an interviewer determines the way in which the questions are presented – heard or read – and the responses recorded – by the interviewer or by the respondent him or herself (Jans, 2008). Holbrook and colleagues (Holbrook et al., 2003) explain that respondents' honesty increases with social distance or "the physical and psychological proximity of one conversational partner to another" (Holbrook et al., 2003, p. 86): the presence of an interviewer, whether in a face-to-face interview or by telephone, can make respondent perceive a risk of being disapproved by the interviewer.

2.3.1. Mode effects: what happens when modes are mixed?

The interest in mixed-mode surveys (combining more than one mode of data collection to obtain information from the same population of study) has been steadily growing during the last decades, hand in hand with interest in investigating the role of questionnaire mode as a source of measurement error (Dillman, 1991). The advantage of using mixed-mode survey designs is the potential to be able to reduce problems of coverage and nonresponse (Hox, de Leeuw & Klausch, 2017), but there is a risk that the information collected through different methods can have information that cannot be compared. The ideal situation would be such that, differential response (for example, older respondents answering by telephone, and young respondents by web, and foreigners by mail) improves the representativeness of the sample, but without the differential response effect introduced by mode (respondents to telephone giving more socially desirable answers than mail and web respondents, for example). In other words, selection effects are considered a positive indicator, meaning that different modes attract different types of respondent. On the other hand, measurement effects may prevent comparability across modes of data collection and cancel out any favourable selection effects.

Survey methodologists have investigated ways of calculating the extent of the bias introduced due to mixing modes of data collection in recent years (for example, Lugtig, Lensvelt-Mulders, Frerichs & Greven, 2011; Revilla, 2013; Vannieuwenhuyze & Loosveldt, 2012), and while some of them have found evidence that using mixed modes has repercussions on the comparability of results, mode effects are not always found to impact research conclusions if differences are due to differences in selection errors (Vannieuwenhuyze & Loosveldt, 2013). Selection effects are related to differences within the composition of the different modes' samples due to the fact that different modes attract different types of respondent or due to coverage/or nonresponse as a result of not being able to participate in a particular mode, whereas measurement effects occur due to people giving different responses depending on the mode used (Dillman, Smyth & Christian, 2014). Up to now, researchers have used three main techniques to separate selection and measurement effects (Tourangeau, 2017). I will explain this in detail in chapter 4.

Mode effects, which have been widely studied in survey methodology, have not received as much attention in terms of possible impact in substantive research, as has been the case in subjective wellbeing studies (Ferrer-i-Carbonell & Frijters, 2004; Pudney, 2010). In spite of this, there have been some studies that have looked at whether mode effects matter for the study of quality of life measures. The following section of this chapter is dedicated to the substantive objective of the thesis, and I present some of these studies along with a brief context on what subjective wellbeing and vulnerability are.

2.4. Wellbeing and vulnerability

Measuring subjective perceptions of how citizens live involves different dimensions that are interrelated and that come from a variety of research fields, such as economics, psychology, social sciences and environmental change studies (Sumner & Mallett, 2011). Two of these measures are vulnerability and subjective wellbeing. These two concepts have become increasingly popular and an objective of governments' policies from the last half of the twentieth century, trend that has led to a proliferation of studies on wellbeing and vulnerability (Oris, 2017).

Vulnerability and subjective wellbeing are non-observable, latent measures that can be defined in different ways depending on the researchers' background. In the next lines I define the concepts of wellbeing and vulnerability emphasising the relationship between them, as well as their dimensions.

The concept of *vulnerability* can have different definitions depending on field of research (Proag, 2014). There are, however, some defining characteristics (Bahadur, Ibrahim & Tanner, 2010) that combine physical, social, environmental and governmental aspects (Sumner & Mallett, 2011). It is a multi-faceted and multidisciplinary construct whose definition is still being debated (Birkmann, 2013; Spini, Hanappi, Bernardi, Oris & Bickel, 2013). In spite of this, Spini and his colleagues (2017: 2) define vulnerability as:

"A lack of resources in one or more life domains, which given a specific context, places individuals or groups at a major risk of experiencing (1) negative consequences related to sources of stress, (2) the inability to cope effectively with stressors, and (3) the inability to recover from the stressor or to take advantage of opportunities before a given deadline."

Someone is vulnerable when they are at risk of suffering from some type of adverse perturbation. The source can be social, economic, political, psychological or physical, but is usually due to a combination of different factors. The ability that someone has to overcome vulnerability is determined by the coping capacity to overcome risk, which depends greatly on the underlying circumstances, such as surrounding family, work, friendships and socioeconomic position (Carreño, Cardona & Barbat, 2007; Spini et al., 2017).

Vulnerability research has put a focus on individuals, as the object of study, and focuses on the capacities of such individuals (Oris, 2017). However, Oris argues, studying vulnerability involves individuals that are connected to others and to their own conditions and resources –which are not equally distributed across the population (Oris, 2017: p. 42). Even thought the focus is on the individual, the risks and resources to overcome such risks also involve the context, the history, the social links, and other dimensions that conform their life course.

The second concept of interest for this thesis is *subjective wellbeing* (SWB). The use of the concept of subjective wellbeing notes the non-material dimension *quality of life*, in contrast to material definitions such as poverty and material vulnerability (Sumner & Mallett, 2011). As for vulnerability, there are different ways of defining subjective wellbeing and a lack of agreement on how to define and operationalize the concept (Fleche, Smith & Sorsa, 2012), but it is generally understood in terms of emotional feelings and evaluative responses (Veenhoven, 2000). This way, someone whose level of subjective wellbeing is high "experiences life satisfaction and frequent joy, and only infrequently experiences unpleasant emotions such as sadness or anger" (Diener, Suh & Oishi, 1997: 25). Conversely, life dissatisfaction, rare joy and negative feelings (e.g. anxiety or depression) indicate a low level of wellbeing. O'Brien and her colleagues state that the reduction of vulnerability is a prerequisite to promote wellbeing (2004), although it is not the only source.

Diener and his colleagues understand that measuring subjective wellbeing means investigating both cognitive and affective life evaluations (Diener, Oishi & Lucas, 2002: 63). Affective wellbeing is related to how intense and frequent of emotions (both negative and positive), while cognitive wellbeing involves evaluations of specific life domains such as work satisfaction, or general life satisfaction (Diener, 1984). Both types of wellbeing can be separated, although they are related, and they are measured in different ways, and their predictors are also different (Luhlmann, Hawkley, Eid and Cacioppo, 2012). An example of this is that affective wellbeing is often researched by asking frequency questions related to emotions, while cognitive wellbeing questions do not ask about a particular time frame. As a multi-dimensional measure, wellbeing does not only include vulnerability indicators, but also subjective ones such as happiness. An effective measurement of wellbeing is essential to understand the dynamics of vulnerability and its impact.

There are two main approaches to understand the concept of wellbeing: the hedonic approach and the eudemonic approach (Deci & Ryan, 2008). The hedonic approach considers that wellbeing is based on feelings and is closely related to the concept of happiness and the absence of negative feelings. The second view, the eudemonic approach, considers the functioning of a person, focusing on them living well. This approach is less interested in feelings and more in the assessment of the conditions of living.

Veenhoven (2008) discusses the sociological aspects of subjective wellbeing, arguing that wellbeing studies offer important information about the quality of the social systems where citizens live. She adds some information to the understanding of eudemonic wellbeing, which she calls cognitive and places emphasis on the role that the comparison – people comparing their own life to the expectations they have of what a good life consists of – in establishing wellbeing. According to this approach, subjective wellbeing is based on a feeling of contentment with your life. She argues that both approaches of wellbeing play an important part in understanding SWB. However, people appear to evaluate their lives taking their mood into account, and not so much the socially constructed notion of what a good life is (Schwarz & Strack, 1991). This idea is contested by other researchers that argue that variables such as inequalities or social norms are essential to the study of wellbeing (Austin, 2016).

Such expositions show that subjective wellbeing is, therefore, a complex multi-dimensional concept that can be studied from different approaches. For this reason, Huppert and So (2013: 843) developed a new inclusive framework of wellbeing that takes into account the following features: feeling accomplished, emotionally stable, engaged, valuable, optimistic, absence of negative emotions, having supporting relationships, self-esteem and vitality. Figure 2, created by Huppert and her colleagues (Huppert, Marks, Michaelson, Vázquez & Vittersø, 2013), illustrates the different components of subjective wellbeing.

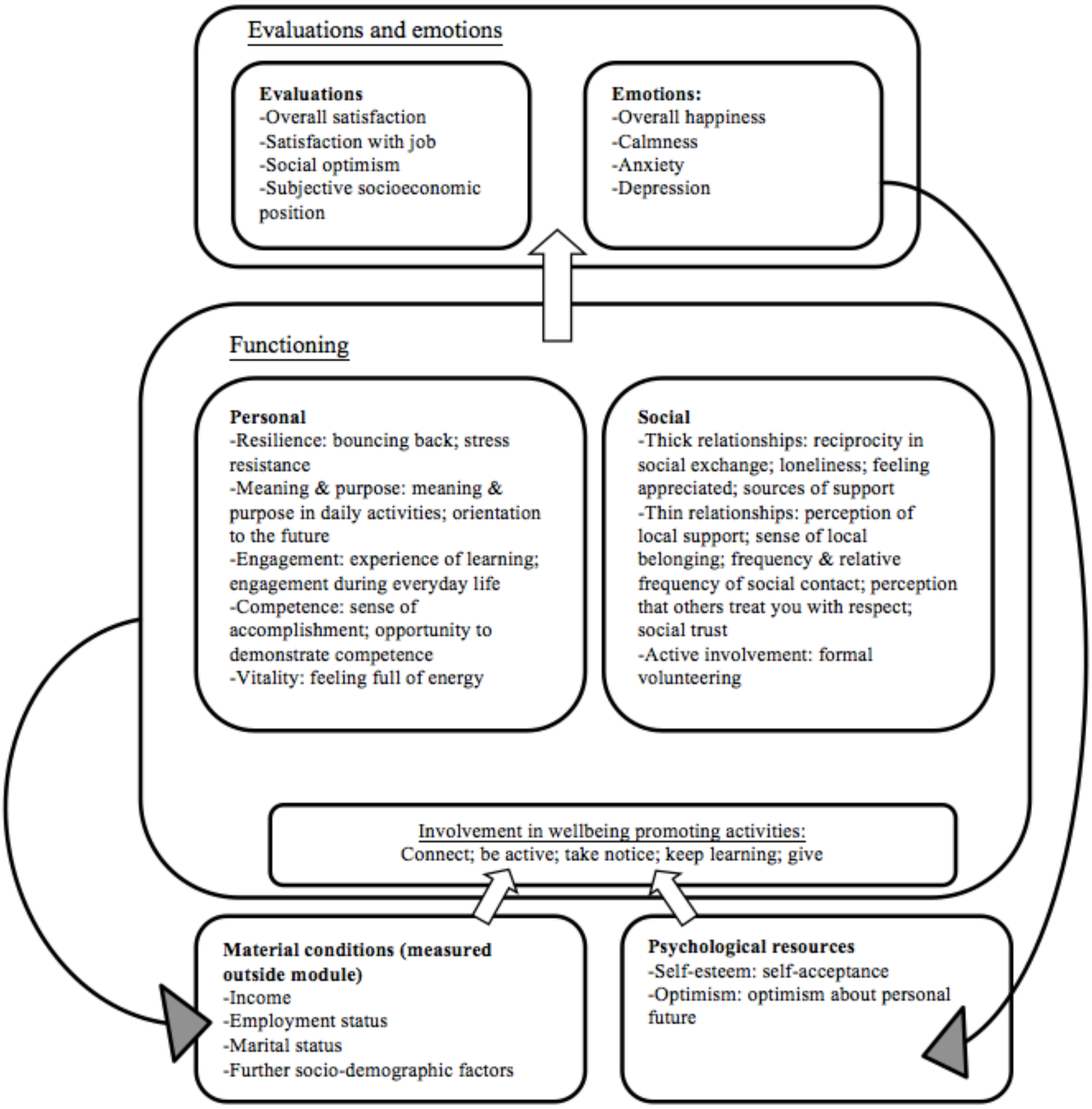


Figure 2. Components of subjective wellbeing. From “Round 6 Module on Personal and Social Wellbeing- Final Module Template”, by Huppert et al. (2013).

2.4.1. Measuring vulnerability and subjective wellbeing

Studies that measure vulnerability require information about the context of the individual and aim to measure the potential triggering situations that make someone vulnerable, which are often stressful life events (Billings & Moos, 1981).

When measuring subjective wellbeing, life satisfaction and/or overall happiness questions are particularly popular (Diener, 2009; Dolan, Peasgood & White, 2008). In spite of this, such measures of overall life quality are also subject to criticism. For example, Austin (2016) points out that a serious issue within wellbeing research is the problem of "adaptive preferences" (130). She argues that more vulnerable people, who often have the least resources to cope with stressors, may report high levels of wellbeing. She affirms the need to take into account deprivation indicators that balance the result of the high level of happiness that might be due to the capacity for adaptation and contentment that people have. For this reason, more studies have started including additional measures such as those recommended by Huppert and colleagues (2013).

Subjective wellbeing measurement involves latent attitudes that cannot be observed directly. Attitudinal questions make the respondent choose between positive and negative dimensions on how they feel about something, for example the level of approval or satisfaction (Alwin & Krosnick, 1991). The extent of error on attitudinal measures depends on the topic of the question, the characteristics of the population being surveyed, the design of the questions and the conditions of the measurement (for example, the mode of administering the questionnaire, the social situation in which the interview takes place).

At the same time, respondents can perceive questions on vulnerability and wellbeing as sensitive. Sensitive questions are identified by three characteristics (Tourangeau et al., 2000). First of all, they may be perceived as intrusive and disrespectful of respondents' privacy, these questions being strongly affected by the normative context surrounding how people talk about their mental health, no matter what the answer option chosen is. Secondly, they may provoke questions about privacy rules being correctly followed, which may deter them from giving sincere answers. Finally, they ask for attitudes or behaviours that society sees as negative or positive, and the respondent perceives that choosing certain response options will give a bad impression about him or herself (Tourangeau & Yan, 2007). To summarize, by asking sensitive questions researchers aim to obtain accurate responses, also when the "true" value is considered to be socially undesirable.

2.4.2. The impact of survey design on measures of quality of life

Subjective wellbeing (SWB) research is often based on surveys that collect information by asking a part of the population to self-evaluate their level of happiness – emotional and variable aspect (Brockmann, Delhey, Welzel & Yuan, 2009) – and life satisfaction – representing the cognitive and less variable part of wellbeing (Diener, Lucas, Schimmack & Helliwell, 2010). These surveys have been under scrutiny, as there is some evidence that subjective wellbeing is a multi-dimensional concept that cannot be measured with a single question (Halleröd & Seldén, 2013) .

Although methodology research for wellbeing studies has been centred on the study of reliability issues (Krueger & Schkade, 2008), some researchers interested in survey methodology and wellbeing studies (Pudney, 2010; Dolan & Kavetsos, 2016) have addressed the question of how the mode used to gather the information impacts data quality and the statistical analysis on the topic of quality of life (Fleche et al., 2012).

The concerns about mode of data collection are specifically related to the implications that measurement and selection effects may have for the comparability of the estimates obtained from different modes' samples, whether from single or mixed-mode studies. The same wellbeing variables, measured with the same questions, can be sensitive to mode of data collection, affecting the information obtained in a significant way (Turner et al., 1992)

Researchers such as Conti and Pudney (2011) and Dolan and Kavetsos (2012) have studied the effect of mode on subjective wellbeing measures in mixed mode surveys, and warn survey data users about the fact that different collection modes show different wellbeing scores (OECD, 2013). In particular, face-to-face respondents scored lower on wellbeing than telephone respondents (Paul Dolan & Kavetsos, 2012). In addition, when comparing respondents who auto-completed the questionnaire (for example, through a paper questionnaire) with respondents that answered with the help of an interviewer, results show that interviewer-based modes show higher levels of wellbeing (Hanmer, Hays & Fryback, 2007).

2.4.3. Studying vulnerable populations

Studies of subjective wellbeing and vulnerability often include group comparisons and examining the situation of minorities and other sub-groups at risk. One of the challenges in doing so is that often it is those vulnerable people who are more difficult to reach and are less willing to provide information about their lives. The problem is that, being significantly different from the rest of the population, obtaining biased information on them can lead to wrong conclusions and, when mixed-mode designs are involved in the study of sub-population analysis, the comparability of the groups becomes difficult. Moreover, nonrespondents are often shown to be the sectors of the population with the lowest socio-economic status, with the lowest level of education, and foreigners. They are social groups that are at risk of being excluded from the rest of society. Rothenbühler and Voorpostel (2016) explain how certain vulnerable people do not have the resources needed to respond to surveys in a pleasurable way. In Switzerland, there are some of the characteristics that can be identified in reluctant respondents and nonrespondents: age (aging is identified with increasing the likelihood of responding, up to 55 years old, when the tendency changes), education (higher education corresponds to a greater likelihood of responding), nationality, working status, income and health condition (Rothenbühler & Voorpostel, 2016).

Overall, these studies suggest that differences in response styles due to respondents' characteristics and mode effects can affect substantive conclusions drawn from surveys, and that if these differential response effects are not taken into account, they may lead to biased estimates (Tutz & Berger, 2016).

In this chapter I have attempted to provide a brief summary of the literature relating to both survey methodology and subjective wellbeing studies and I have shown that mode of data collection is an important aspect of the survey design that, due to differential errors across modes, can lead to biased conclusions in surveys that use data from multiple modes. The next chapter describes the procedures and methods used in this investigation to address the research questions that were presented in the introduction.

CHAPTER 3. DATA AND METHODS

The purpose of the thesis was to study the effect that mode of data collection has in responses to subjective wellbeing questions. In particular, my aim was to answer the research questions presented in Chapter 2, which are:

RQ1. Do different modes of data collection differentially affect the quality of survey estimates of subjective wellbeing?

RQ2. Do mode effects on measurement affect all respondents equally?

RQ3. Do mode effects on measures of subjective wellbeing impact the results of substantive research into the predictors and correlates of subjective wellbeing measures?

To do this, I used cross-sectional data from methodological study designed to compare different single and mixed mode survey designs on the topic of subjective wellbeing and vulnerability. In this chapter, I present the details of the data source used to implement the thesis' analyses, such as the sample composition and the data collection methods and provide some information about the methods I used. I finish by presenting some of the potential weaknesses of the chosen methodological approach. The chapter offers a general overview of the methods used to answer the research questions; it is possible to find a more detailed overview of the specific analytical approaches adopted in the methods sections of each empirical chapter.

3.1. Data

This thesis presents a statistic analysis of quantitative data from a survey that was conceived as a methodological experiment designed to assess the impact that using different modes of data collection – whether singly or in combination – has on different sources of survey error. The survey experiment was carried out by LIVES' Individual Project 15 and FORS International Surveys group in 2012 and 2013 on the topic of wellbeing (Roberts, Joye & Stähli, 2016). The Economic and Social Research Institute M.I.S. Trend was responsible for the collection of the data. The population of study included those persons that reside in the French-speaking part of Switzerland, aged 15 and above. To select the survey participants, the Swiss Federal Statistical Office drew the sample using simple random sampling based on the registers of municipalities (Roberts, Joye & Stähli, 2016) in the French-speaking part of Switzerland. The gross sample for the study included 3919 individuals, from cantons such as,Geneva, Jura, Neuchâtel, and some parts of Fribourg, Valais and Bern – for which only French speakers were selected (Roberts et al., 2016).

In addition to the information gathered by the survey, the sampling frame contained socio-demographic information from the municipal registers, including sex, age, marital status, nationality, country of birth, residence permit, household size and urbanization. Telephone numbers of those survey participants whose telephone number was listed in the telephone directory maintained by Swisscom (telephone and Internet provider), and also used for sampling purposes by the Federal Statistical

Office, provided additional information that was used as a basis for the experimental design.

3.2. The experimental design

The survey experiment consisted of a single mode mail survey; a mixed-mode sequential survey that involved a web survey, a paper questionnaire in case of nonresponse and finally, in case of not having answered the mail questionnaire, a telephone-call (if the person had a listed fixed phone number) or a face-to-face interview (if the person did not have a listed fixed phone number); and another mixed mode sequential survey in which the first mode of data collection used was telephone, followed by a mailed paper questionnaire in case of nonresponse (see figure 3).

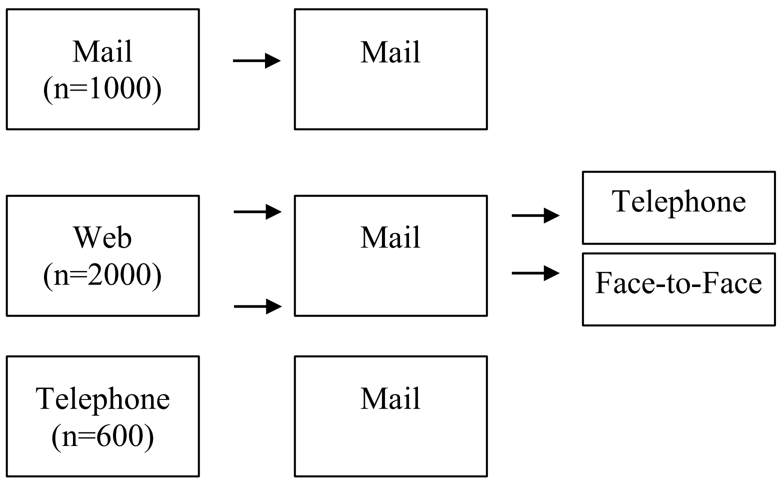


Figure 3. The different stages of the data collection process

The questionnaire was adapted for each mode of data collection, there being four versions of the questionnaire for each mode: telephone, mail, web and face-to-face. Adaptations were related to the showcards used in face-to-face versions of the questions on the European Social Survey (ESS), that had to be addressed in the telephone mode by having the interviewer explain a certain type of scale, breaking down the question into smaller sections, or converting a closed-ended into an open-ended question. Adaptations were also necessary for the web version of the survey, which were based on the European Social Survey guidelines (see Roberts et al., 2016). Survey participants were sent a pre-notification letter including an unconditional cash incentive of 10 Swiss Francs before the interview took place for telephone respondents, and before the paper questionnaire or the link to fill the web survey were mailed to them, for those invited to respond to the mail or web version.

Of the 3919 individuals, 3600 were allocated to one of the survey modes of data collection used depending on whether they had a listed telephone number or not, and on the expected response rates for each mode of data collection. The other 319 were used to form a 'reserve' sample that I do not use in this thesis (see table 1):

Table 1. Sample size by mode of data collection and telephone information

	Main sample		**Reserve sample**
Mode	With listed telephone number	Without listed telephone number	
CATI	600	-	-
PAPI	500	500	319
CAWI	1000	1000	-

Moreover, the survey design involved various stages of data collection, in which different methods were used to follow up each prior phase's non-respondents. This method was dependent on the mode of data collection to which the person had been allocated in the first stage (see figure 3). The design of the methodological design consisted of a first concurrent phase in which individuals were allocated to either mail (PAPI), web (CAWI) or telephone (CATI) modes, and a sequential phase for those that had not responded in previous stages.

Given these features of the study's design, it is possible to find three different designs, in which the 'assigned' mode can be different to the 'responding' mode. It is possible to find more details about the different reminders and contact attempts in the research report by Roberts and her colleagues (2016, p. 13).

The response rates for each mode of data collection were calculated by Roberts and her colleagues (*ibid.*) as recommended by the American Association for Public Opinion Research, in which nonrespondents are those eligible cases that are not interviewed (further details are available in the study's methodological report; *ibid.*). Table 2 (below) provides an overview of the response rates per assigned mode of data collection. The first column of the table shows the response rate regardless of the response mode, while the second column refers exclusively to those survey participants that responded to their assigned mode. The following two columns present the sample size for each mode of data collection for respondents to their allocated modes, the last one presenting the information only for respondents with a listed telephone number.

Table 2. Response rate by assigned mode and response mode

	Overall response rate	Response rates to assigned mode	(n)	(n, with telephone number)
PAPI (n=1000)	70.2%	65.4%	(654)	(351)
CAWI (n= 2000)	71.4%	44.5%	(889)	(457)
CATI (n=600)	70.0%	60.7%	(364)	(364)
Total			1907	1172

The questionnaire: "Bien-être et Mal-être en Suisse romande"

Translated into English, the title of the questionnaires was "Well-being and Unease in French-speaking Switzerland", so survey participants did not know that the survey was a methodological experiment, as they were informed that the aim of the study was exclusively to research the wellbeing of the population in the French speaking part of Switzerland. The questionnaire contained approximately 125 questions: 44 measuring socio-economic background and 41 measuring different aspects of subjective wellbeing obtained from the European Social Survey on personal and social wellbeing (European Social Survey, 2013) and previous LIVES and FORS surveys (such as MOSAiCH). The rest of the questions (40) were about opinions and perceptions about society in general (such as attitudes towards immigration) and measures designed to investigate methodological aspects such as mode preferences or social desirability bias (Roberts et al., 2016, p. 15).

Completing the questionnaire took respondents between 25 and 30 minutes of their time. During this time, participants had to provide information related to socio-demographic questions and answer questions that allow the measurement of various topics related to the NCCR LIVES's interest in vulnerability across the life course such as life satisfaction, health, happiness, social support, personal relations, individual wellbeing and important life events, among others.

It is possible to find the whole questionnaire used to collect the data in the mail survey in the appendix of this thesis.

3.2.1. Survey participants

In this thesis, I used data from respondents of the concurrent mixed-mode part of the survey, which includes those respondents that answered through their allocated mode. That is, mail respondents that had been allocated to mail, web respondents that had been allocated to web, and telephone respondents that were allocated to telephone.

In addition, I restricted my analysis by using only information from respondents that answered in the modes they were allocated to in the first place, allowing the comparison of estimates and statistical results across modes as it is easier to isolate the source of the differences than when there are multiple confounding survey design differences, such as the different timing and trials. In addition, for some parts of the analyses I used only those individuals that have a listed fixed telephone number, with the aim of having sample compositions in the web and mail modes that were as similar as possible to the telephone mode. Respondents that have landline telephones had significantly different characteristics in terms of age, sex or place of birth, from those individuals that do not have a fixed telephone and such differences could further complicate the analysis and identification of the sources of error.

In table 3 (below), I present some socio-demographic characteristics based on data available from the population register that describe the survey participants of each mode of data collection, and compare these characteristics to those of the gross sample, for those who have listed telephone numbers.

Table 3. Composition of responding samples in primary assigned mode (sample members with telephone numbers), items from the registry

	Gross sample	**CAWI**		**PAPI**		**CATI**	
	(n=2100)	(n=457)		(n=351)		(n=364)	
Auxiliary Variables	% (Std.Err.)	% (Std.Err.)	p	% (Std.Err.)	p	% (Std.Err.)	p
Male	47.4 (1.1)	49.0 (2.3)		45.9 (2.7)		47.0 (2.6)	
Age (mean in years)	50.3 (0.4)	45.1 (0.8)		50.3 (1.0)		48.5 (1.0)	
Age group			***				*
<30	18.5 (0.9)	24.5 (2.0)		16.2 (2.0)		21.2 (2.1)	
30-44	20.8 (0.9)	22.3 (2.0)		23.4 (2.3)		18.7 (2.1)	
45-64	33.9 (1.0)	38.7 (2.3)		35.0 (2.6)		37.6 (2.5)	
65+	26.9 (1.0)	14.4 (1.7)		25.4 (2.3)		22.5 (2.2)	
Marital Status			*		**		
Single	28.9 (1.0)	34.1 (2.2)		24.8 (2.3)		28.8 (2.4)	
Married[1]	56.9 (1.1)	54.3 (2.3)		64.4 (2.6)		58.0 (2.6)	
Nationality							*
Swiss	79.2 (0.9)	81.6 (1.8)		81.5 (2.1)		86.3 (1.8)	
Bordering	8.5 (0.6)	9.2 (1.4)		7.7 (1.4)		6.6 (1.3)	
Other	12.3 (0.4)	9.2 (1.4)		10.8 (1.7)		7.1 (1.4)	
Household			**				
1	16.0 (0.8)	10.7 (1.5)		14.2 (1.9)		14.6 (1.9)	
2	30.6 (1.0)	28.0 (2.1)		30.8 (2.5)		27.7 (2.4)	
3	18.6 (0.9)	18.6 (1.8)		19.7 (2.1)		20.6 (2.1)	
4+	34.8 (1.0)	42.7 (2.3)		35.3 (2.6)		37.1 (2.5)	
Urbanisation					†		
City/town	28.4 (1.0)	26.3 (2.1)		23.1 (2.3)		26.9 (2.3)	
City/town	42.8 (1.1)	45.3 (2.3)		43.3 (2.7)		42.6 (2.6)	
Isolated town	1.10 (0.2)	0.7 (0.4)		0.9 (0.5)		1.6 (0.7)	
Rural	27.7 (1.0)	27.8 (2.1)		32.8 (2.5)		28.8 (2.4)	

Notes: Non-parametric one-sample chi-square tests of goodness of fit: *** p<0.001, **p<0.01, *p<0.05, †p<0.10

In table 4, I present additional socio-demographic information about the respondents obtained from the questionnaire, for which there is no register information available, to show how the characteristics of health, longstanding disability, work status, marital status, education, whether French is spoken at the household, use of the internet and, for foreign respondents, the amount of years lived in Switzerland vary across modes.

[1] Includes only married people (not in partnership)

[2] Based on the ordered logistic regression without the matching, with the "omodel" test (not available

The table shows that there are statistically significant differences between the three modes in terms of disability, education, years lived in Switzerland for foreigner respondents, whether French is spoken at home and use of the Internet.

Table 4. Composition of responding samples in primary assigned mode (sample members with telephone numbers), questionnaire items

	CAWI		PAPI		CATI		
Questionnaire socio-demographics	(n=457)		(n=351)		(n=364)		
	%	(Std.Err.)	%	(Std.Err.)	%	(Std.Err.)	p
Bad health	12.0	(1.5)	17.1	(2.0)	11.8	(1.7)	†
Longstanding illness	15.8	(1.7)	31.6	(2.6)	18.7	(2.0)	***
In paid work	67.4	(2.1)	60.6	(2.6)	58.8	(2.6)	*
Partner							
Yes, living together	62.3	(2.3)	67.2	(2.5)	64.3	(2.5)	
Yes, not living together	10.1	(1.4)	10.7	(1.6)	7.42	(1.4)	
No	27.6	(2.1)	22.0	(2.2)	28.3	(2.4)	†
Education							
None	0.24	(0.02)	1.5	(0.6)	0.0	-	*
Basic	19.5	(1.9)	19.5	(2.1)	25.8	(2.3)	*
General training	16.9	(1.8)	10.9	(1.7)	9.7	(1.6)	***
Professional	38.3	(2.3)	47.2	(2.7)	37.5	(2.6)	***
University	25.0	(2.1)	20.9	(2.2)	26.9	(2.3)	*
Years in Switzerland							
1-3	4.7	(2.1)	9.60	(3.1)	3.4	(1.9)	
4-10	14.2	(3.4)	17.2	(3.9)	11.4	(3.4)	*
+11	81.1	(3.8)	73.1	(4.6)	85.2	(3.8)	**
French spoken at home	96.7	(0.8)	91.4	(1.5)	96.7	(0.9)	***
No access/little use of Internet	3.50	(0.8)	16.2	(2.0)	18.2	(2.0)	***

Notes: Pearson's chi-squared test: *** $p<0.001$, **$p<0.01$, *$p<0.05$, †$p<0.10$

Respondents suffering from a longstanding disability appear to be significantly more numerous in the mail sample than in web and telephone samples, while the percentage of respondents that have a paid job is higher in web than in mail and telephone. In addition, the modes seem to attract respondents with different educational levels: mail has a higher level of respondents with professional training or an apprenticeship; the web sample contains a higher level of respondents whose highest educational level is general training. At the same time, there are no respondents without formal education answering the telephone survey, while there are in the self-completion samples. Related to both the number of years lived in the country and whether French is spoken at home, it is possible to see how those respondents that do not speak French at home and those that have lived in Switzerland for less than 10 years are more likely to respond the mail survey. People who have been living in the country for more than 10 years are more likely to respond by phone, followed by web, and are less likely to respond by mail.

3.3. Analytical approach

The analytical approach for the thesis consisted of three main steps of statistical analysis aimed at answering the three main research questions presented in the first chapter. In the next lines, I recap these questions, and describe the methods used to answer them in each of the empirical studies.

R.Q.1: Do different modes of data collection differentially affect the quality of survey estimates of subjective wellbeing?

To answer the first research question, in Study 1, I compare the effect of mode on the distribution of the responses and estimated means and implemented a series of regression analyses in which I examined the relationship between response outcome (each one a measure of subjective wellbeing) and response mode, which is the independent variable alongside a socio-demographic control. In addition, I look at the effect of mode in the means of the same measures. It is in this analysis where one of the issues relevant to the test for mode effects appears: the confounding of selection and measurement effects. The key priority of this study was to be able to distinguish between selection and measurement mode effects for a series of closed-ended questions on different personal, social and work wellbeing measures. On the one hand, selection effects due to mode are related to the fact that different modes attract different types of respondent, meaning that some types of respondents are less willing –or less able– to participate in surveys conducted in some modes, potentially increasing the likelihood of some groups of the population responding to a larger extent than others. On the other hand, measurement effects are related to people giving different responses to different modes, potentially making responses incomparable across modes (Dillman, Smyth & Christian, 2014).

In this first empirical chapter, I explain this problem in detail and the different techniques researchers have used to separate selection and measurement effects (Tourangeau, 2017). To overcome the problem of confounding effects, I chose a similar procedure to previous studies that study the effects of mode of data collection in which respondents are randomly allocated to different modes of data collection, and coverage and nonresponse differences across the modes can be controlled using administrative data or socio-demographic information (Heerwegh & Loosveldt, 2011; Lugtig, Lensvelt-Mulders, Frerichs & Greven, 2011; Vannieuwenhuyze & Loosveldt, 2013). In particular, I looked at three alternative ways of adjusting for differences in socio-demographic characteristics across the modes. The idea behind the analysis is that, if the different modes' samples are rendered as similar as possible, remaining differences in means and distributions will be due to differences in how people respond to different modes.

In addition, in Study 2, I focused on the effect of mode in responses to open-ended questions about important life events. Although the aim is the same as in the previous chapter, the fact that open-ended questions require a different type of statistical analysis, I dedicate a separate chapter to focus on this type of question. In particular, I look at different indicators of the quality of responses such as item-nonresponse, response length, and differences across modes of data collection in the types of life events mentioned by respondents and whether their impact was considered to be positive or negative.

R.Q.2: Do mode effects on measurement affect all respondents equally?

The aim of the third study was to examine whether the effect of mode is the same for all respondents or whether response differences across modes vary for certain population subgroups. To do this, I replicated the regression analyses from the first study, including interaction terms between a series of respondent characteristics, including age, sex, education, nationality, and motivation to respond and the mode of data collection. Before implementing the regressions, I implemented matched the samples using coarsened exact matching to render the samples of each mode as similar as possible.

R.Q.3: Do mode effects on measures of subjective wellbeing impact the results of substantive research into the predictors and correlates of subjective wellbeing measures?

I dedicate the last two empirical studies (4 and 5) to the question about the importance of mode effects on measurement at the item level for substantive researchers conducting multivariate analyses. In order to do this, I implemented statistical analyses techniques that are widely used in social science research as a way of illustrating whether researchers using the same statistical analysis would obtain different results depending on the mode of data collection. As researchers use both single items of wellbeing, such as happiness or life satisfaction, and composed measures that take into account the different aspects of the concept, I aimed to respond to this research question in two parts: using both uni- and multi-dimensional measures of subjective wellbeing in studies 4 and 5, respectively.

In Study 4, I use regression analyses to look at the relationship between happiness, social trust, and job satisfaction, and their predictors. Regression coefficients, levels of significance, standard errors, and estimates of R-squared are compared between the web, the mail and the telephone samples to assess the differences in results. In addition, I test for differences between the estimated regression coefficients using a Wald test.

Study 5, on the other hand, focused on multi-dimensional constructs of wellbeing and consisted in implementing two multi-group confirmatory factor analyses aimed at assessing the equivalence of multi-item measures across modes. Two models, one for a latent measure of personal subjective wellbeing and another for the latent measure of wellbeing at work, were tested across web, mail and telephone. In order to identify differences in the two models across modes, I tested for the level of measurement invariance. The analyses were implemented with and without socio-demographic controls using a propensity score.

All the analyses were carried out in Stata 12.0 and Stata 15.0 (studies 1, 2, 3 and 4) and in R 3.2.0 (study 5).

3.4. Conclusion

To conclude, this thesis is based on the statistical analysis of data from a methodological experiment implemented in the French-speaking part of Switzerland. By comparing data that come from three response modes –web, paper and telephone – I aimed to respond to the three main research questions formulated in this thesis about the effect of mode in measures of subjective wellbeing, whether such differences are the same for all members of the population, and whether they matter for substantive analyses that involve SWB measures. In the next chapter, I assess the extent of mode effects in mean estimates and distribution of responses about subjective wellbeing measures, while adjusting for differences in selection bias between web, mail and telephone modes.

CHAPTER 4. MODE EFFECTS IN MEASURES OF SUBJECTIVE WELLBEING

4.1. Introduction

Social science researchers use data gathered through survey designs that vary across survey organisations and through surveys conducted by the researchers themselves. Differences can occur due to different aspects of the survey design (the way of asking the questions, using different question wordings or different response alternatives, for example), how survey participants are selected or the mode of data collection, among others. All of these can impact the quality of the data obtained by the survey, and, although mode of data collection is just one aspect, it often has an influence on the decisions researchers take in other aspects of the survey design (Cernat, 2015a).

Nowadays, using data from different modes of data collection is standard (Buelens & Van Den Brakel, 2017). Whether using a mixed-mode survey design or comparing cross-sectional data gathered using different modes of data collection – for example, from different countries or from different time periods – many researchers work with information collected in different ways. This situation has potential advantages - it may help diminish nonresponse bias in mixed-mode designs or allow for comparisons that could not otherwise be done - and also disadvantages related to potential drawbacks when analysing the data by potentially rendering it incomparable across modes (Bowyer & Rogowski, 2017). As already discussed in some detail in Chapter 2, the comparability of data collected via different modes is a classic problem in survey methods literature (de Leeuw, 2005) and has been the subject of much systematic investigation during the last decades. Combining different modes of data collection can mean different responses to the same questions (Nandi & Platt, 2017) due to both differential coverage and nonresponse error (selection effect), and measurement errors across modes (Couper, 2017).

For this reason determining the source of the response difference is a challenge that survey researchers have tried to overcome in order to facilitate the analysis and comparability of data gathered with different modes (Jäckle, Roberts & Lynn, 2010; Kolenikov & Kennedy, 2014; Revilla, 2010; Vannieuwenhuyze, Loosveldt & Molenberghs, 2014). Up to the present survey methodologists have developed different methods to distinguish the sources of mode effects, which I present in this study.

In the field of subjective wellbeing studies, however, there has been a lack of research on how the different aspects of survey design impact the quality of the data. In contrast with the amount of research dedicated to measuring the concept of subjective wellbeing and its different aspects (Diener, 1994; Disabato, Goodman, Kashdan, Short & Jarden, 2016) and its reliability and validity (Lyubomirsky & Lepper, 1999), there has been little research on the effect of other aspects of the survey design, including the effect of mode of data collection (Fleche, Smith & Sorsa, 2012; Pudney, 2010). However, the evidence suggests that inherent characteristics of the different modes could mean different survey estimates of subjective wellbeing across modes of data collection (Springer & Hauser, 2006). The existing research has shown that respondents to telephone survey modes tend to systematically report higher levels of subjective wellbeing than face-to-face and self-completion modes (Dolan & Kavetsos, 2012;

Pudney, 2010; Sarracino et al., 2017). An example of the potential effect that this can have is to look at the league tables that show levels of quality of life (such as life satisfaction) for a wide range of countries may be inaccurate if mode of administration is confounded with country and so lead to the wrong conclusion. A well-known example is the “Better Life Index” (OEDC 2015), based on data from Gallup World Poll conducted by face-to-face or telephone administration depending on the country: out of the top ten countries on levels of life quality, nine used surveys implemented by telephone. These results could indicate that the mode of data collection may have an effect on such comparisons, but the extent of the differences in wellbeing is not clear. While some researchers find no significant differences (Sarracino et al., 2017), others find that differences are considerable (Dolan & Kavetsos, 2012; Pudney, 2010; Springer & Hauser, 2006). These studies, however, differ in a number of ways, including the countries of study, differences in sample sizes, question design, a lack of a mixed-mode methodological experiment, or the restriction to only a few measures of subjective wellbeing (often happiness and life satisfaction) that make it difficult to arrive at definite conclusions about the effect of mode.

In this study, I aim to respond to three research questions that ask about whether there are differences in the means and distributions of measures of subjective wellbeing between mode of data collection. In addition, I investigate whether mode effects due to measurement and/or selection effects and if measurement effects vary as a function of the response format and sensitivity.

The chapter starts with an explanation of mode effects and the methods that have been used to measure their impact on data quality and to separate selection and measurement effects, presenting some results from previous studies on mode effects in SWB measures. The second part discusses the methods by which the chapter’s analyses were conducted. Using data from the methodological experiment introduced in Chapter 3, I compare results from a web, a mail and a telephone survey to show whether there are differences in mean and distribution estimates from a series of subjective wellbeing measures, paying special attention to separating the selection and measurement mode effects, and examining whether such differences may be due response characteristics related to sensitiveness and response format. Finally, I discuss them in the context of previous literature and the limitations of the study.

4.2. Literature review

4.2.1. Differences in survey estimates related to mode of data collection

There are three important aspects that can relate to mode measurement effects (Lugtig, Lensvelt-Mulders, Frerichs, Greven, 2011). First, the presence or absence of an interviewer, who can encourage responses and help clarify the question to the respondent but also make respondents more prone to tailor their answers, influencing the positivity of the responses. Secondly, whether questions are presented aurally or visually can impact response choice as respondents who listen to the questions and response options often memorize and choose the last heard option, whereas respondents reading the response options tend to choose the first appropriate category (Dillman & Christian, 2005). Lastly, response differences can be due to different question formats, such as the presence of a "don't know" option, which may not be available in all modes (Hox, de Leeuw & Zijlmans, 2015). Mode differences may be due to differences in how the questions are asked in different modes, and a unified design may be a way to reduce differences and enhance comparability. Measurement differences caused by different question formats can be avoided to some extent using a unified-mode design (Dillman, Smyth & Christian, 2014), but calculating the difference in measurement error between the different modes of data collection used remains essential before pooling data that come from different modes (Klausch, Schouten & Hox, 2014). However, observed differences in survey estimates across modes, as seen in Chapter 2, are not necessarily due to response differences: it could simply be due to who answers the questions if the kinds of people who respond to the survey differs between modes (Lugtig, Lensvelt-Mulders, Frerichs, Greven, 2011). These two effects of mode are confounded, making it difficult for researchers to adjust for mode effects (de Leeuw, 2005).

There are certain question characteristics and in particular response formats that affect the way in which people respond to survey questions (Böckenholt, 2017), and this effect may vary differently in different modes beyond the control of the researcher. In particular, different question designs can be associated with the tendency of respondents to disproportionately choose some response categories over others, independently of what their "true" answer is. There are different tendencies, which are known as response styles, and they can be exacerbated by some mode effects (Liu, 2014).

Roberts (2016) provides a recent overview of some of the most widely studied types of response styles that can impact answers to response scales. Response styles include *acquiescence response style,* when respondents tend to agree by choosing "only the highest response categories" (Roberts, 2016, p. 581); *mid-point response style*, when respondents tend to select the middle response category; *extreme response style,* when respondents tend to select the highest or lowest options available; and *mild response style*, which refers to the tendency to never select the extreme response options. Response styles are a type of error that can affect univariate distributions, means and variances (Roberts, 2016). Some of these response styles have been previously associated with particular modes of data collection (Dillman & Christian, 2005). For example, Liu, Conrad and Lee (2017) found that acquiescent responding is present in both web and face-to-face. In addition, previous research has found that telephone interviewing is related more to acquiescence compared to collecting information via mail and web (Van Vaerenbergh & Thomas, 2013). Acquiescence and extreme responding appear to be more common in telephone surveys, while mid-point responding is more common in self-completed questionnaires (Roberts, 2016). Telephone respondents are significantly less likely to choose the neutral point in scales (Weijters, Schillewaert & Geuens, 2008). Additionally, the difference between web and mail in tendency to choose the more extreme options has not always been found to be the same: while at times there have been no differences, on other occasions the style has been more common in web than in mail (Van Vaerenbergh & Thomas, 2013; Weijters, Schillewaert, & Geuens, 2008).

4.2.2. Separating selection and measurement mode effects

To make informed decisions about whether to implement a mixed-mode survey or not, survey methodology researchers have put a strong focus on finding a method to disentangle selection and measurement mode effects (Schouten, van den Brakel, Buelens, van der Laan & Klausch, 2013; Tourangeau, 2017). In the following lines, I will introduce the main approaches used in the literature, explaining their advantages and disadvantages. Tourangeau (2017) identifies three ways of disentangling selection and measurement effects: by directly assessing the measurement error, by making statistical adjustment, and by modelling the errors.

Direct assessment of measurement errors

The direct assessment of the measurement error involves using an experimental setting in which data is collected through various modes and the survey estimates and validation data are compared, for example, using a high quality survey ("a gold standard") as a comparison point (Tourangeau, 2017, p. 14). Examples of previous studies that have used this type of approach involving the testing of measurement errors have been able to separate measurement error from selection error because there was information about both respondents and nonrespondents. For example, Kreuter, Presser, and Tourangeau (2008) examined differences between responses to questions about different aspects of academic performance and official university reports. This mixed-mode survey consisted of a parallel mixed-mode design in which respondents were allocated to either telephone, interactive voice recognition or web modes. Their findings showed that responses to the telephone mode were more affected by social desirability, while web responses were the most accurate.

Using a similar approach, Tourangeau, Groves and Redline (2010) compared reports of voting behaviour with information from the sample frame and found selection and measurement mode effects in both the mail survey and the telephone survey, with both types of mode effect being stronger in the telephone mode.

Another strategy of assessing mode effects is to compare responses to similar questions that have been obtained in two different modes at two points in time or during the same single interview (Hewitt, 2002). Klausch, Schouten and Hox (2014) chose this approach using data that came from the Dutch Crime Victimization Survey, for which the information was collected with telephone, face-to-face, mail and web modes. Their analysis revealed no differences in the amount of mode effect when comparing mail to web responses and telephone to face-to-face responses. However, there were strong differences when comparing the self-completion and interviewer-based designs. For this reason the authors warn against combining mixed-mode design that includes these two types of design.

Vannieuwenhuyze and Lynn (2014) describe this model, the instrumental variable model, as based on a binary variable – the instrumental variable – that works as a covariate dividing the sample into two groups according to mode of data collection. Although useful, it cannot be applied in every case as researchers cannot often count on validation data for attitudinal questions, which is the type of question where survey researchers worry the most about mode effects. An advantage of using this model is to be able to calculate the measurement effect and, if needed, correct data in the mixed mode survey on the basis of the single mode. However, it cannot control for selection effect when estimating 'target statistics' (Vannieuwenhuyze and Lynn 2014).

In addition, when implementing this analysis there are some requisites to take into account (Angrist et al., 1996; Heckman, 1996, 1997) that are difficult to verify unless we have an experimental setting. The first of them is that respondents from one of the groups must respond by a single mode; and the measurement error for one of the modes has to be equal in the single-mode and the mixed-mode samples. Secondly, the mixed-mode and single mode samples must represent the same population.

Statistical adjustments

This method consists of comparing survey estimates from at least two single modes of data collection and statistically adjusting for differences in either selection bias or measurement bias. As selection bias can provoke mode effects (Vannieuwenhuyze & Loosveldt, 2013), the first type of statistical adjustment consists of equating the mode groups or, in other words, to make their respondents' characteristics as similar as possible to render them comparable (Vannieuwenhuyze & Loosveldt, 2012). Several studies have attempted to separate selection from measurement effects this way, introducing a series of covariates that control for either measurement or selection effects. The process involves some kind of regression adjustment or post-stratification weighting procedure (D'Orazio, Di Zio & Scanu, 2006; Jäckle et al., 2010), allowing the implementation of other statistical tests such as ordinary, logit or probit regressions or confirmatory factor analysis (Klausch, Hox & Schouten, 2013).

There is also the possibility of adjusting for mode effects through multiple-imputation (Kolenikov & Kennedy 2014). The idea behind the multiple imputation adjustment is that respondents could potentially respond in all modes. As in mixed-mode surveys, respondents are allocated to just one mode, this method consists of imputing the outcomes for the rest of the modes as with missing data, relying on auxiliary data (Hox, de Leeuw & Klausch, 2017.

Vannieuwenhuyze and Lynn (2014) identify two possible ways of doing this: the confounder and the mediator model. The confounder model, also known as back-door method, involves the addition of covariates that explain selection effects in a regression model that has mode of data collection as the predictor variable. This way, the effect of mode on the measure of interest is assumed not to be due to observed differences between respondents but instead to differences in their responses. This method relies on two assumptions: first, the chosen covariates must capture all the selection effect between the mode groups or the confounding problem would remain, and second, mode must not affect the covariates themselves. These conditions mean that the back-door approach has some drawbacks: it is not easy to prove that covariates are not affected by mode unless there is auxiliary data available (such as register data), and when such information exists, it is often not enough to control effectively for selection effects.

Despite this, the model has been extensively applied in previous research (Heerwegh & Loosveldt, 2011; Jäckle, Roberts & Lynn, 2010; Lugtig, Lensvelt-Mulders, Frerichs & Greven, 2011; Roberts & Vandenplas, 2017), often using socio-demographic variables as covariates. For example, Heerwegh and Loosveldt (2011) control for differences between the modes, gender, age, education, job, and urbanization.

In another approach using auxiliary information, Sarracino and his colleagues (2017) used the Blinder-Oaxaca decomposition to find out whether the different modes' samples were comparable. This method was used to explain the difference in the average wellbeing between the web and telephone samples due to selection and measurement differences and it indicated that the difference was due to the way in which respondents respond differently to web and telephone surveys. The analytical approach was completed with coarsened exact matching to also take into account that mode may impact how respondents choose specific categories (Sarracino, Riillo & Mikucka, 2017: 146).

In the literature review, researchers that have used the "back-door" approach have adjusted for selection effects by using different techniques: the auxiliary variables are usually introduced in the statistical models used either in the form of covariates (Jäckle, Roberts, & Lynn, 2010), as a propensity score or using propensity score matching (Lugtig, Lensvelt-Mulders, Frerichs & Greven, 2011), or using coarsened exact matching.

The option of the covariate regression adjustment has been criticised for not being effective enough in rendering the different modes' samples comparable, or "balanced", in terms of sample composition (Lugtig, Lensvelt-Mulders, Frerichs & Greven, 2011; Ross et al., 2015). The two common alternatives are propensity score matching (PSM), which is one of the most used types of matching (King & Nielsen, 2015) and Coarsened Exact matching (CEM), which is an increasingly popular alternative procedure (Sarracino, Riillo, & Mikucka, 2017). Both types of matching are ways of pairing units from two different groups (in this study, the groups are the response modes) “that are similar in terms of their observable characteristics” (Dehejia & Wahba, 2002, p. 151) with the idea of having two populations that are identical “except in their treatment status” (M. E. Ross et al., 2015, p. 990). In the next lines, I will summarize the main differences between the two approaches.

The objective of propensity score matching is to estimate the effect of a treatment (in this case, using a certain mode of data collection compared to another) by accounting for the covariates that predict receiving the treatment or not, which in this study means answering in one mode or another (Rosenbaum & Rubin, 1983). The propensity score can be understood as the probability of an individual being allocated to a certain mode of data collection based on a series of covariates. The idea behind the model is that all the units (or individuals) with the same propensity score value are comparable, meaning that both “treated and untreated units have the same distribution of characteristics” (Pearl, 2009, p. 348) resulting in an approximation of a randomised experiment. While the use of propensity scores for regression adjustment often works correctly, the use of propensity score matching has being discouraged as it has been found to lead to increased imbalance between sample groups (King & Nielsen, 2015), particularly when a large number of covariates are used (Pearl, 2009), rendering it ineffective to adjust for selection differences and even increasing imbalance between the samples. In fact, some authors strongly recommend checking the results with another matching method before deciding to use it (King, Nielsen, Coberley & Pope, 2010).

On the other hand, coarsened exact matching (CEM) is a monotonic imbalance, non-parametric matching method that has the same aim as propensity score matching, and works by temporarily grouping, or coarsening, the covariates into meaningful groups (with respondents that have common characteristics), allowing the easy checking of the balance automatically (Iacus, King & Porro, 2012). The main difference with the use of propensity score matching is that its mechanism is designed to approximate a "fully blocked experimental design" (King & Nielsen, 2015), which as a consequence has a stronger reduction of the imbalance between the examined groups (Iacus, King, & Porro, 2012b, p. 14). In addition, this type of matching is subject to fewer assumptions than other weighting procedures (Blackwell, Iacus, King & Porro, 2009), such as certain modelling assumptions (e.g. normal distribution) which are often impossible to verify (Kantor & Kershaw 2010).

To summarize, even if the PSM and CEM are based on the same idea of grouping individuals with a similar probability of participating in a survey based on a series of covariates, the way of choosing the observations that form each group and pruning the observations that do not fit in any of them is different: the mechanism of the propensity score matching is less precise, possibly matching dissimilar respondents, which can lead to a higher level of imbalance (Burden et al., 2017; King & Nielsen, 2015).

The other alternative is to use the mediator model, by using covariates that explain the measurement effect as an intermediate variable. Such variables could measure respondents' perceptions about surveys such as "response burdens, satisficing, acquiescence, or social desirability" (Vannieuwenhuyze, Loosveldt & Molenberghs, 2014, p. 10). In this case, the remaining amount of mode effect that is not explained by measurement effects is assumed to be due to differences in selection bias. This approach is not as common as the confounder model, but Vannieuwenhuyze and his colleagues tested it in 2014 using a measure of how much respondents liked surveys as the regression covariate. As with the previously described alternatives, its implementation has some requirements that complicate its application: it is necessary that the included covariates capture all the measurement effect between the modes, or otherwise the measurement effect remains. It is also necessary that the covariates used to predict the measurement effect are not affected by selection effects or the confounding would remain. Relying on auxiliary data capable of explaining the selection differences is one of the major drawbacks of this method, and although previous research uses socio-demographic variables as a proxy for unobservable measures, remarkably, studies that use this type of weighting approach have been found to leave uncontrolled between 40 and 70 percent of the selection effect (Tourangeau, 2017).

Results from studies using this approach differ. Jäckle, Roberts and Lynn found mode effects in 16 out of 28 ordinal variables (2010) in a telephone and face-to-face mixed-mode survey. Kolenikov & Kennedy (2014) investigated a survey that involved web and telephone modes of data collection and 19 out of the 297 analysed were sensitive to measurement effects after adjusting for selection differences, possibly also related to a higher level of socially desirable answers in the telephone mode. Schouten and colleagues (2013) showed that weights based on socio-demographic variables, using administrative records, were able to properly control for selection effects, using data from an experiment in the Netherlands.

Modelling measurement error

A third way exists that involves estimating the measurement errors through statistical modelling, using confirmatory factor analysis, latent class modelling, or a regression model. Researchers applying the modelling approach have often used latent class modelling for longitudinal data and cross-sectional data from different modes (Tourangeau, 2017). For example, Biemer (2001) found mode effects in a series of measures after having applied latent class modeling. In particular, seven out of 14 variables were sensitive to mode when analysing face-to-face and telephone, and the measurement error was found to be smaller in the case of the telephone survey. In addition, Revilla (2012) used a split-ballot Multitrait-Multimethod approach in various topics such as political orientation, social trust and political trust to find that there were few differences in half of survey estimates between web and face-to-face. Where there were differences, the findings indicated that the quality of the information was better for the web sample. However, such differences were found to be small in most of the cases. Heerwegh and Loosvelt (2011) tested results from a telephone and mail survey implemented in Belgium. The results from their structural equation models, which account for differences in selection error, found that responses to the telephone survey were more positive than in the mail one. Jäckle and her colleagues (2010), and Klausch, Hox and Schouten (2013) also used this approach, for whom I have already presented some of their results in the previous section.

4.2.3. The impact of mode of data collection in measures of subjective wellbeing

Previous literature has warned about the presence of mode effects in subjective wellbeing measures (Bowling, 2005; OECD, 2013; Sakshaug, Yan & Tourangeau, 2010; Schwarz & Strack, 1991). However, the need for "carefully designed experiments in combination with weighting or regression based inference methods to control for selection effects" (Schouten, van den Brakel, Buelens, van der Laan, & Klausch, 2013, p. 1556) has meant that not many studies have been able to find a method that successfully indicates the separate amount of each type of mode effect in the topic of subjective wellbeing. Such an experiment requires the only difference between the different samples' to be the response mode: if other aspects of the survey design are different, "the effect of mode is confounded with there other differences" (Jäckle, Roberts, & Lynn, 2010, p. 5). For this reason, a unified design that allows comparisons across modes is a requirement to study the extent of the effect of mode. However, it is possible to identify some studies that have focused in this topic and that I present in this section.

Pudney (2010) used data from an experiment implemented by the British Household Panel Survey in which two main modes were used to collect data from the respondents: a face-to-face survey and a telephone survey. In addition, a share of the respondents that were allocated to the face-to-face survey were randomly selected to complete a part of the interview through computer-assisted self-interviewing. One of the challenges the author had to overcome was the fact that face-to-face interviews included the use of showcards, while their use in the telephone survey was not possible. For this reason, some of the questions were slightly different for the two modes. The theme of the survey was satisfaction with different aspects of respondents' lives (health, household income, available leisure time) and also with life overall. For some of the survey designs, response alternatives consisted in 7-response scales, for which 7 was completely satisfied and 1 completely dissatisfied), while for the rest (in all cases telephone respondents), the question was divided into two sections: first asking about satisfaction, neither satisfaction or dissatisfaction, and dissatisfaction, and secondly about its intensity (somewhat, mostly or completely).

Using a nonparametric Kruskal-Wallis test and a Bonferroni adjustment, Pudney found significant differences in the response distributions when comparing self-completion to face-to-face responses, and face-to-face to telephone responses. The proportion of respondents that choose the most extreme satisfaction values was higher in face-to-face than in self-completion; and in telephone than in face-to-face.

Dolan & Kavetsos (2012, 2016) examined mode effects in the Annual Population Survey from the United Kingdom in which respondents answered either in a face-to-face mode or by telephone. However, the allocation to each mode was not random and, as the authors describe it, “not entirely clear” (Dolan & Kavetsos, 2016a, p. 7). In particular, they look at survey items about life satisfaction, feeling worthwhile, happiness and anxiety for which the respondents report their level of happiness using a 11-point scale. Adjusting for socio-demographic differences (age, gender, marital status, education level and ethnicity), they looked at the impact of mode in subjective wellbeing estimates using an ordinary least squares (OLS) regression. They found that that face-to face respondents scored lower on wellbeing than telephone respondents, the difference being bigger for the life satisfaction item, followed by worthwhile, happiness and anxiety.

Hanmer, Hays and Fryback (2007) examine responses about self-reported health that came from a self-completed paper questionnaire and an interviewer-based survey. Results show that interviewer-based modes report higher levels of wellbeing than respondents to the paper-and-pencil questionnaire.

In their comparison of web and telephone survey estimates, Sarracino, Riillo and Mikucka (2017) used data from the Adult Population Survey of the Global Entrepreneurship Monitor, for which the telephone sample was drawn from respondents listed in a fixed telephone line registry and the web sample from a frame of e-mail addresses. In their analysis, they only consider the data from those respondents that have listed fixed telephone numbers. Their items of interest are life satisfaction, whether respondents have obtained the important things they want in life, whether they would change things if they lived again, whether they consider their life conditions to be excellent, and whether they think that their life is close to ideal. Response categories go from 1 to 5, 1 meaning strong disagreement and 5 strong agreement. Their method of studying the mode effects consisted of a back-door approach, using Blinder-Oaxaca decomposition, which gave information about the comparability of the samples and about the role of response mode being the cause of differences in the subjective wellbeing measures (Sarracino et al., 2017, p. 145), and coarsened exact matching to control for differences in selection bias (they adjust for gender, employment status, education level, year of the survey, age and income). Afterwards, a multinomial logit regression estimated the relationship between mode and subjective wellbeing measures. Their findings showed that web respondents reported lower subjective wellbeing levels than telephone respondents, and this difference also remained after adjusting for differences in selection error. The mean of life satisfaction was also significantly different between web, for which the average level score was 4, and telephone, with an average score of 4.15. Results for the other measures of wellbeing showed some different results: web respondents were more likely to choose the most positive response categories than the telephone respondents.

4.2.4. Research questions

R.Q.1: Are there mode differences in estimates of subjective wellbeing?

Based on previous research, I expect to find differences in responses to questions about subjective wellbeing between different modes of data collection, particularly when comparing responses to interviewer-based modes to self-completion modes. As previous research has shown, I expect that survey estimates from interviewer-based modes will show higher levels of subjective wellbeing than survey estimates from self-completed estimates.

R.Q.2: What is the extent of mode differences due to measurement and selection effects?

The review of the literature on mode effects indicated that observed differences were often due to differences between who responds to each mode. I expect that, adjusting for differences in selection, responses to modes in which an interviewer was involved will still be more positive in terms of subjective wellbeing than responses to self-completion modes. Based on previous findings, differences between web and mail are expected to be due to selection errors.

R.Q.3: If there are measurement effects, are they related to the questions' format or question sensitivity?

In the literature review, I have shown that measurement mode effects can be related to the question and response design and how people tend to answer them. Differential response styles and mode of data collection can be interrelated, as some modes may increase the likelihood of respondents choosing certain response categories. While acquiescence and extreme responding might be more common in telephone respondents, I expect web and mail respondents to avoid the extreme response options more. Also, I expect that part of the measurement effect will be related to social desirability, as previous research suggests.

4.3. Methods

In this chapter I compare data from three groups of respondents allocated to web, mail and telephone. In particular, I look at differences in means and response distributions across modes and attempt to model the measurement error by adjusting for differences in selection using three different techniques.

4.3.1. Data

I use the first of the mixed-mode experiments in which all respondents answered by the means of the mode they were allocated to, and exclude the rest of the respondents. This means that I use information from the mail respondents that had been allocated to the mail mode, from the web respondents that had been allocated to the web mode, and the telephone respondents that were allocated to the telephone mode. In the web and telephone, and mail and telephone comparisons, I only use data from respondents that had a listed fixed telephone.

4.3.2. Analytical approach

In order to investigate the effects of mode on measures of subjective wellbeing, I first compared responses given to the question on subjective wellbeing by respondents in each mode of data collection, before trying to model the selection error in order to detect mode effects on measurement. The different analysis steps aim to respond to each of the research questions:

R.Q.1: Are there mode differences in estimates of subjective wellbeing?

To address the first research question, I start by presenting descriptive results comparing means and response distributions for the subjective wellbeing variables. I compare both individual items and composite scores for the items sharing the same response format. I used a series of t-tests to show whether there are significant response differences between web, mail and telephone. In particular, the purpose of this section is to show whether responses to the different modes of data collection appear to be different or not, independently of them being related to selection or measurement effects. As I include both respondents with and without a listed telephone number in this part of the analysis for the comparisons of the web and the mail survey, I used design weights that account for differences in the selection error of the two samples.

Afterwards, I examined the sample compositions of web, mail, and telephone modes (see table 5, below) to investigate whether there were different response biases in each mode of data collection and how they differed. A descriptive analysis of the data, complemented with chi-squared tests showed the differences in the different modes' samples for the socio-demographic variables. The analysis showed significant differences between the web and mail sample in terms of sex, age, marital status, urbanization, and nationality. There was a larger proportion of females, respondents older than 65, married, foreign, and respondents with a listed telephone number in the mail compared to the web group. Moreover, there were differences depending on the language spoken at home with fewer respondents that do not speak French at home in the web group, more respondents with a listed mobile phone in the mail group, and more mail respondents that did not have access or did never use the Internet than web ones. I also found differences between web and paper compared to telephone. In particular, there were statistically significant differences in terms of age (older respondents answered more in telephone than in web, and there were fewer respondents aged 30 to 44 in telephone than in mail), marital status (with fewer single respondents in web than in telephone, and more widowed people in telephone than in web), nationality and country of birth (fewer non-Swiss respondents are found in the telephone mode than the mail and web modes). All telephone respondents had a listed telephone number. In terms of the questionnaire socio-demographic variables, I found differences between the web, mail and telephone groups in the number of foreign respondents who have been living less than 11 years in Switzerland, being better represented in the web and mail surveys than in the telephone survey. Lastly, there are also fewer respondents from French speaking households in the web and telephone than in the paper sample.

Table 5. Composition of the web, mail and telephone samples

Socio-demographic characteristics	CAWI (n= 889)	PAPI (n = 654)	CATI (n = 364)	p
Female	48.4	54.0	53.0	†
Age group				
<30	24.5	20.6	21.2	
30-44	29.5	28.8	18.7	***
45-64	34.5	31.4	37.6	
65+	11.5	19.3	22.5	***
Marital Status				*
Single	36.2	31.4	28.9	
Married	53.0	56.3	58.0	
Divorced	8.0	8.9	6.6	
Widow	2.7	2.9	6.0	
Nationality				
Swiss	76.0	73.1	86.3	***
Bordering country	10.6	8.6	6.6	†
Other	13.4	18.4	7.1	***
Household Size				
1	16.3	17.7	14.6	
2	28.8	30.4	27.8	
3	18.5	20.5	20.6	
4+	36.5	31.3	37.1	
Urban	73.8	68.8	69.5	†
Years in Switzerland				**
1-3	2.9	2.3	0.8	
4-10	5.7	8.3	2.8	
+11	22.1	22.6	20.6	
Swiss	69.3	66.8	75.8	
French spoken at home	95.7	88.4	96.7	***
Listed mobile phone	9.1	12.4	10.2	
No Internet access/use	3.3	14.4	18.2	***

Notes: Pearson's chi-squared test: *** $p<0.001$, **$p<0.01$, *$p<0.05$, †$p<0.10$

R.Q.2: What is the extent of mode differences due to measurement and selection effects?

In this study, I used three different types of regression analysis: ordered logistic regression, partial proportional odds modelling and multinomial logistic regressions. Partial proportional odds regressions were used when the proportional odds assumption was proved false (Williams, 2008). However, although the first two types of analysis have been found to be an appropriate method for studying mode effects in ordinal measures (Jäckle et al., 2010), and all the SWB items examined in this study are ordinal, due to the small number of cases in some of the categories for some of the examined variables, I also implemented multinomial logistic regressions for the web and telephone and mail and telephone comparisons, where the partial proportional odds model did not work properly for every measure.

In the regressions implemented, individual items of subjective wellbeing are the outcome variables, and mode of data collection is the predictor. The comparison of mode of data collection is between web and mail; web and telephone; mail and telephone; and self-completion modes and telephone.

In the literature review, I presented alternative approaches that previous researchers have used to study the extent of measurement and selection bias related to mode of data collection. For this thesis, I decided to model the selection effects to be able to isolate the measurement effect by using the previously described "back-door" approach. As well as being one of the most commonly used techniques to separate selection and measurement effects (Tourangeau, 2017) this approach also requires a less complex experimental design, which may facilitate its use by social science researchers using mixed-mode data or multiple sources of data that come from different modes. In addition, using a golden standard survey (such as the Social European Survey) would have meant restricting the number of the subjective wellbeing measures analysed.

To implement the back-door variable approach, I conducted a series of regression models in which the dependent variable was one of the 27 subjective wellbeing measures, which I will present in the next section, and mode of data collection the predictor. In addition, to adjust for differences in the selection error across modes, I use a series of socio-demographic variables, which I will also present in the following lines.

In order to use the front-door approach, it is essential that the chosen covariates are not affected by measurement effects themselves (Vannieuwenhuyze, Loosveldt &

Molenberghs, 2010) but capture the mode-specific coverage and nonresponse bias. For this reason, socio-demographic variables are often used (Jäckle, Roberts, & Lynn, 2010) even if they do not fully capture the selection differences. However, although the socio-demographic controls are not always effective in controlling for selection effects, in this case they are useful to account for at least some of the differences between the samples. Some socio-demographic characteristics are important because they can be indicators of a way of being in society and their way of answering, even though they are characteristics that are not observed in the questionnaire. The way in which people live and relate to society and social groups, as well as the individual's social status, have been found to be linked to socio-demographic characteristics such as sex (Anderson, John, Keltner & Kring, 2001), ethnicity (Shaked, Williams, Evans & Zonderman, 2016), marital status, or age (Dias de Freitas, Pinheiro Ferrari, Poerschke Vieira, da Silva, Pereira de Carvalho & Cardoso, 2016). In addition, information such as having or not having a fixed telephone number can indicate other respondents' characteristics, as respondents to landline telephone surveys have been found to be more satisfied with their lives (Mohorko, de Leeuw, & Hox, 2013), richer and suffer from less deprivation than face-to-face respondents (Lipps, 2016, p. 23).

Based on the differences in the sample composition of the different modes of data collection presented in table 5, I selected those variables I presumed to not be influenced by measurement mode effects. This is known for those variables for which there was validation register data. In the case of the questionnaire variables, I presume that they are not, following the approach of previous studies (Vannieuwenhuyze, Loosveldt, & Molenberghs, 2010) . To test the assumption that covariates explain the selection effect, I used a series regression analysis with mode of data collection as the dependent variable and the covariates as the predictors, as recommended by Vannieuwenhuyze (2014) and Lugtig (2011).

The auxiliary variables that I considered were sex, age, country of birth or marital status, urbanisation, the use of the Internet and language spoken at the household. In the next section, I present how these variables were coded and the number of categories for each.

To decide which covariates to include for each comparison, I run logistic regressions in which mode of data collection is the dependent variable predicted by the auxiliary variables. This way, I chose those covariates that were significantly related to response mode.

In particular I used area of living, Internet use and language as covariates for the web and mail comparison; age, nationality, marital status and Internet as covariates for the web and telephone comparison; and age, nationality, marital status, area of living and internet use for the mail and telephone comparison. The results for the chosen covariates from these regressions indicated that the covariates explain at least some of the selection effect in the regression models for all mail and web, web and telephone and the mail and telephone (p=0.000).

In this study, I analyse the mode effect using three of the techniques that are widely used to control for selection differences. In particular, I compare whether the regression with covariates, the regression with propensity score and the regressions after implementing Coarsened Exact Matching differ in the extent to which they render the different mode samples' more comparable.

In order to facilitate the interpretation of the results, for this reason I present the estimated coefficients and the odds ratios when they facilitate the interpretation of the results in the comparison of estimates of the composite scores. The odds ratio is a statistic used to assess the association between a factor and a particular outcome (Szumilas, 2010). In our study, wellbeing is the outcome, and mode of data collection the factor. If the odds ratio is 1, there is no association between the response mode and the wellbeing measure. In the ordered logistic regression, odds ratios higher than 1 indicate a positive association (the bigger it is, the stronger the effect mode in the outcome), while below 1 means that there is a negative association between the dependent variable and the predictor.

To create the propensity score, I used a logistic regression analysis predicting mode of data collection and using the previously mentioned socio-demographic variables as predictors. Following this step, the calculated propensity score acts as a control in each regression model or as a weight, for the mean comparisons. In addition, I calculated separately the coarsened exact matching weights using the same variables with the function "cem" in Stata. Afterwards I implemented the same type of analysis as before matching.

The objective of using propensity scores and coarsened exact matching is to make the different modes' samples comparable. This is, to balance by making them more similar in terms of composition with reference to the auxiliary variables used as covariates. To compare how effective they were in comparison to the sample controlled by covariates, I present results from three tests of sample imbalance: the bigger the imbalance, the less similar the sample composition is. First of all, results from the balance check of the covariates showed that the use of both propensity scores and coarsened exact matching using the Stata command "pbalchk". This command shows the standardized difference between the covariates' mean in one mode of data collection, compared to the mean in another mode (see table 6). In addition, the command "pstest" checks whether balance between the samples has been achieved. Results showed that balance was achieved for all mode comparisons independently of the method used.

Table 6. Sample balance test: Standardized mean differences

	Covariates	PS	CEM
Web-mail			
Urbanisation	0.146	0.345	0.000
Internet use	-0.423	0.000	0.000
Language	-0.271	0.000	0.000
Web-telephone			
Age (4 groups)	0.177	-0.019	0.000
Nationality	0.134	-0.008	0.000
Marital Status	0.083	-0.006	0.000
Internet use	-0.37	-0.004	0.000
Mail-telephone			
Age (4 groups)	-0.048	-0.048	0.000
Nationality	0.154	-0.045	0.000
Marital Status	-0.134	-0.023	0.000
Urbanisation	-0.119	-0.028	0.000
Internet use	0.076	0.014	0.000

R.Q.3: If there are measurement effects, are they related to the questions' format or question sensitivity?

In this last part, the same approach used to answer the second research question is followed. However, instead of using individual items, the outcome variables are the composite scores for the items sharing the same response format. This way, I replicate the ordered logistic regressions adjusting for differences in nonresponse and coverage between modes using covariates, propensity score and coarsened exact matching, separately to present some indication on the impact of question design in the mode effect. Lastly, for the different mode comparisons for the 27 measures, I adjust the results from the significant tests using the Bonferroni-Holm method, a sequential version of the Bonferroni correction (Narum, 2016). Ordering the p-values in ascending order, the smallest p-value is multiplied by the number of tested items (27), continuing with the rest of p-values that are each multiplied by m-1, m being the number of items.

4.3.3. Variables

The variables analysed were 27 measures of different aspects of subjective wellbeing, (see table 7, below, and appendix for the French version).

Table 7. Subjective wellbeing measures

Questions on subjective wellbeing * Reverse coded variables for which the highest score is the most positive option	Categories
Taking all things together, how happy would you say you are? (Very unhappy- very happy)	0-10
All things considered, how satisfied are you with life as a whole nowadays? (Very unsatisfied-very satisfied)	0-10
How is your health in general? (Very good- very bad) *	1-5
To what extent do you take the time to the things that you really want to? (Not at all-completely)	0-10
How much time during the past week have you felt depressed?* (None or almost none of the time - all or almost all of the time)	1-4
How much of the time during the past week has your sleep been restless?* (None or almost none of the time - all or almost all of the time)	1-4
In general I feel very positive about myself (Agree strongly – disagree strongly)*	1-5
Most days I feel a sense of accomplishment from what I do (Agree strongly – disagree strongly)*	1-5
I feel free to decide how to live my life (Agree strongly – disagree strongly)*	1-5
How much of the time during the past week you felt anxious?* (None or almost none of the time - all or almost all of the time)	1-4
In the last month, how often have you felt that you were unable to control the important things in your life?* (Never- very often)	1-4
In the last month, how often have you felt confident about your ability to handle your personal problems? (Never- very often)	1-4
In the last month, how often have you felt that things were going your way? (Never- very often)	1-4
In the last month, how often have you felt difficulties were piling up so high that you could not overcome them?* (Never- very often)	1-4
How much of the time during the past week have you felt lonely?* (None or almost none of the time - all or almost all of the time)	1-4
I'm always optimistic about my future (Agree strongly – disagree strongly)*	1-5
How often do you meet socially with friends, relatives or colleagues outside work? (Never - everyday)	1-7
Do you have anyone with whom you can discuss intimate and personal matters? (None - more than 10)	0-6
Comparing to other people who are your age, how often do you take part in social activities? (Much less - much more)	1-5
To what extent do you get support from your close ones if needed? (Not at all - completely)	0-6
To what extent do you give support to your close ones if needed? (Not at all - completely)	0-6
How much of the time do you find your job interesting? (None or almost none of the time - all or almost all of the time)	0-6
How much of the time do you find do you find your job stressful?* (None or almost none of the time - all or almost all of the time)	0-6
How likely would you say it is that you will become unemployed in the next 12 months? (Very likely – not likely at all)	1-5
All things considered, how satisfied are you with your present job? (Very unsatisfied-very satisfied)	0-10

The auxiliary – socio-demographic – variables that I used to adjust for differences in selection error between the different modes' samples come from the questionnaire. This information was checked against the register data to verify that the two sources of data corresponded: sex (0 female, 1 male), age (scale or 4-category variable), nationality (0 non-Swiss, 1 in Swiss), type of residence permit (0 no permit or long-term permit, 1 five year or less permit), marital status (0 not married nor registered partnership, 1 married or registered partnership), urbanization (0 rural, 1 urban), whether respondent has a listed fixed telephone number (0 no, 1 yes), language (0 does not speak French in household, 1 speaks French in household, at least as a third language), internet use (0 uses internet at least sometimes, 1 never uses internet/does not have access at home/work).

In each mode comparison, the models violated the parallel lines assumption were[2]:

Table 8. Items for which the parallel lines assumption is violated

Web- mail (5 of 27)	Web- telephone (14 of 27)	Mail telephone (10 of 27)
Someone to discuss	Life satisfaction	Happiness
Taking control	Happiness	Someone to discuss
Handle problems	Taking time	Giving support
Overcome differences	Meeting close ones	Optimism
Work-life balance	Someone to discuss	Positivity
	Giving support	Freedom
	Optimism	Handle problems
	Positivity	Things going well
	Freedom	Overcome differences
	Taking control	Restless-sleep
	Handle problems	Work-life balance
	Things going well	Expects job loss
	Overcome differences	
	Work-life balance	

[2] Based on the ordered logistic regression without the matching, with the "omodel" test (not available for complex survey design analysis) in Stata, using mode of data collection as a predictor.

4.4. Results

4.4.1. Mode effects in estimates of subjective wellbeing

The purpose of the first empirical question was to investigate whether mode had an impact on the means and distributions of wellbeing. In the table below (9) I compared means from web to mail, web to telephone and mail to telephone. It is possible to observe that, although there were statistically significant differences in all mode comparisons – web and mail, web and telephone, and mail and telephone when we do not control for mode effects on selection – some of these disappear after adjusting the significance test results for multiple testing using the Holm-Bonferroni adjustment. As can be seen from the table, after this adjustment, there were no statistically significant differences when comparing the web and mail samples. In addition, although 22 out of the 27 variables appeared to be affected by mode comparing mail and telephone, after the multiple testing corrections the number was 17 out of 27. The web and telephone comparison shows that there are significant mean differences in a total of 12 of 27 SWB measures. The significance level test shows, therefore, that the stronger differences appear when comparing the self-completion modes (mail and web) to telephone. Closer inspection of the table shows that in all cases, the mean estimates for telephone respondents indicate a more positive level of wellbeing. The differences were also found in all of the different aspects on SWB examined: the general measures, the presence or absence of negative emotions, social wellbeing and wellbeing at work.

Table 9. Mean differences across modes and t-test

SWB measures, by question format	Web (1) (n= 889)			Mail (2) (n= 654)			Telephone (3) (n= 364)		
	Mean	Std. Err.	P>t 1 vs. 2	Mean	Std. Err.	P>t 2 vs. 3	Mean	Std. Err.	P>t 1 vs. 3
11 categories									
Social trust	5.29	0.08		5.39	0.09	(**)	5.88	0.12	***
Life satisfaction	7.56	0.06	(*)	7.32	0.08	***	7.95	0.09	(**)
Happiness	7.67	0.06		7.53	0.07	***	8.14	0.07	***
Take time	6.25	0.07		6.24	0.09	(*)	6.60	0.11	(**)
7 categories									
Meets close ones	4.94	0.04		5.00	0.05	***	5.49	0.06	***
Someone to discuss	2.64	0.05	(*)	2.51	0.06	(*)	2.73	0.08	
Gets support	4.97	0.04	(*)	4.82	0.05	†	4.99	0.07	
Gives support	5.24	0.03	†	5.13	0.04	(**)	5.33	0.05	
5 categories									
Health	4.21	0.03	(*)	4.13	0.03	**	4.26	0.04	
Optimism	3.77	0.03		3.75	0.03	***	4.04	0.05	***
Positivity	3.71	0.03		3.73	0.03	***	3.99	0.05	***
Freedom	4.06	0.03		4.09	0.03	***	4.39	0.04	***
Accomplishment	3.85	0.02		3.90	0.03	***	4.16	0.04	***
Take control	3.78	0.03		3.76	0.04	(*)	3.91	0.06	†
Handle problems	4.04	0.03	(*)	3.92	0.04		3.94	0.06	
Things going well	3.85	0.03		3.78	0.04	**	3.98	0.05	(*)
Overcome diff.	3.98	0.03		3.94	0.04	**	4.12	0.05	(*)
Social activities	2.64	0.04	(*)	2.51	0.04	**	2.76	0.05	(*)
4 categories									
Depression	3.45	0.02		3.42	0.03		3.48	0.03	
Restless sleep	3.25	0.03	†	3.18	0.03	†	3.27	0.04	
Loneliness	3.58	0.02	(*)	3.49	0.03	***	3.69	0.03	**
Anxiety	3.24	0.02		3.23	0.03		3.28	0.04	
SWB at work									
11 categories									
Job satisfaction	7.64	0.08		7.55	0.10	**	8.14	0.11	**
Work-life balance	6.86	0.09		6.74	0.12	**	7.37	0.15	**
6 categories									
Interesting job	4.62	0.05		4.75	0.06	**	5.04	0.07	***
Stressful job	2.41	0.06		2.22	0.07	**	2.68	0.12	
4 categories									
Expects job loss	3.27	0.03		3.16	0.04	**	3.44	0.05	**

*** p<0.001, **p<0.01, *p<0.05, †p<0.10

As I was also interested in the results depending on response format, in table 10 I show the means for the composite score of the previously presented items, demonstrating that the telephone mode does produce self-reported higher levels of subjective wellbeing than mail and web. However, the results for the 4-category questions show very similar means for web and telephone, the mean of web being smaller.

Table 10. Mean differences across modes by response format

SWB measures, by question format	Web (1) (n= 889)			Mail (2) (n= 654)			Telephone (3) (n= 364)		
	Mean	Std. Err.	P>t 1 vs. 2	Mean	Std. Err.	P>t 2 vs. 3	Mean	Std. Err.	P>t 1 vs. 3
11 categories	6.70	0.05		6.61	0.06	***	7.14	0.07	***
7 categories	4.15	0.03		4.12	0.04	***	4.38	0.04	***
5 categories	3.79	0.02		3.75	0.02	***	3.96	0.03	***
4 categories	3.38	0.02	(*)	3.33	0.02	**	3.43	0.02	

*** p<0.001, **p<0.01, *p<0.05, †p<0.10

In the next lines, I illustrate the differences in the distributions across modes of data collection summarized using composite scores in order to ease the interpretation of the results. Figure 4 compares the distribution of scores for the measures happiness and life satisfaction for each mode group. It can be seen from the data in the table that the respondents in the telephone group selected the response options 9 and 10 to a higher extent, and at the same time they chose the categories 0, 1 and 2 less often than mail and web respondents. On the other hand, web and mail respondents chose the values 3, 4, 5, 6 and 7 more than telephone respondents. The response alternative 8 is selected at a similar rate in all three modes (around 30% of the respondents).

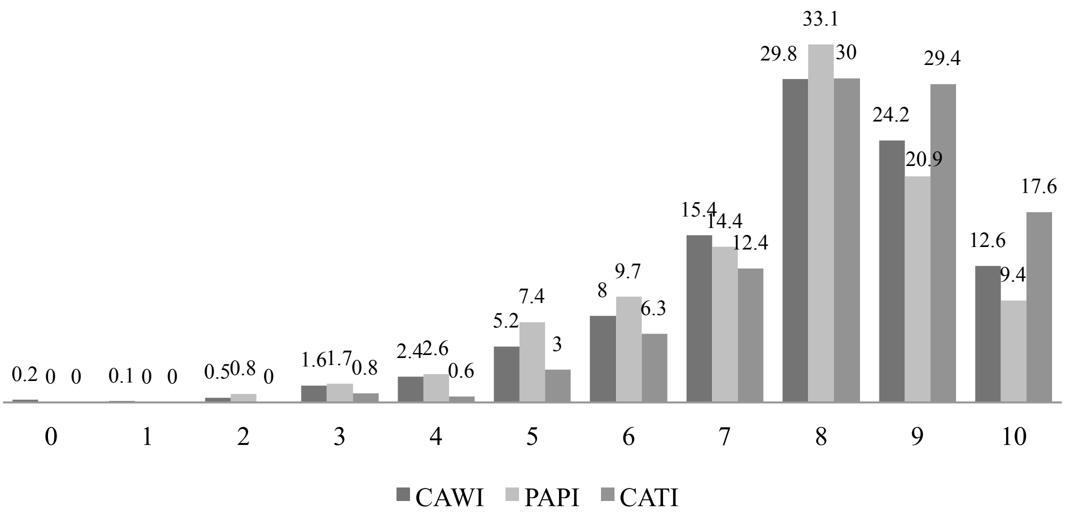


Figure 4. Response distributions by mode, 11-point scale of happiness and life satisfaction (%)

Figure 5 (below) shows the percentage of respondents that chose each response for the combination of all measures with 11 categories. Results are different from those in the first figure in various aspects. First of all, what stands out is that there are many fewer respondents – in all modes – that chose the most extreme categories 0, 1, 2 and 10. Still, it is the telephone group that has the highest percentage of respondents choosing the most extreme and positive options (8, 9 and 10), while middle categories 4, 5 and 6 are preferred by self-completion respondents.

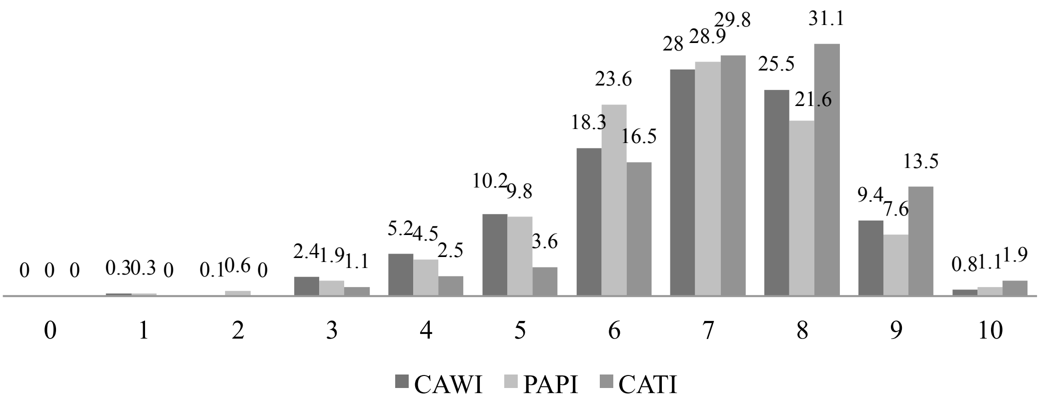


Figure 5. Response distribution, mean score all 11-point scale measures (%)

Figure 6 displays results for the variables with 7 answer categories (all of which are related to the social dimension of subjective wellbeing). Web and mail respondents preferred the middle categories (2, 3 and 4), while a higher proportion of telephone respondents chose the two most extreme categories. Few respondents chose the more negative categories (0 and 1) in any of the modes.

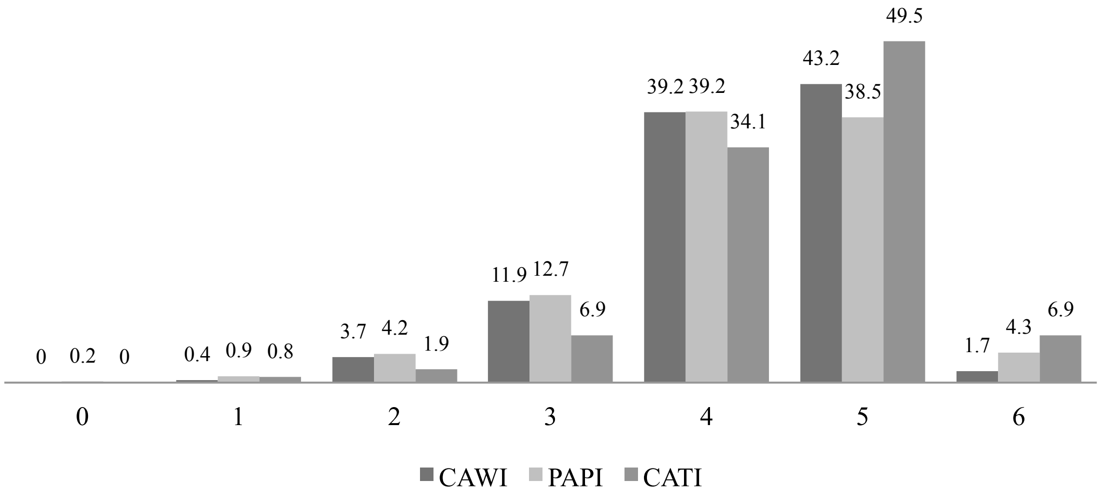


Figure 6. Response distribution, 7 point scale measures (%)

Figure 7 displays the response distributions for the questions that offered 5 response alternatives. In this case, the middle response category (3) was more popular with mail and web respondents than with telephone respondents. Telephone respondents, on the other hand, chose the most positive option (5) more often than self-completion ones. For the rest of the categories, the distribution of responses was similar, although no telephone respondents chose the most negative category.

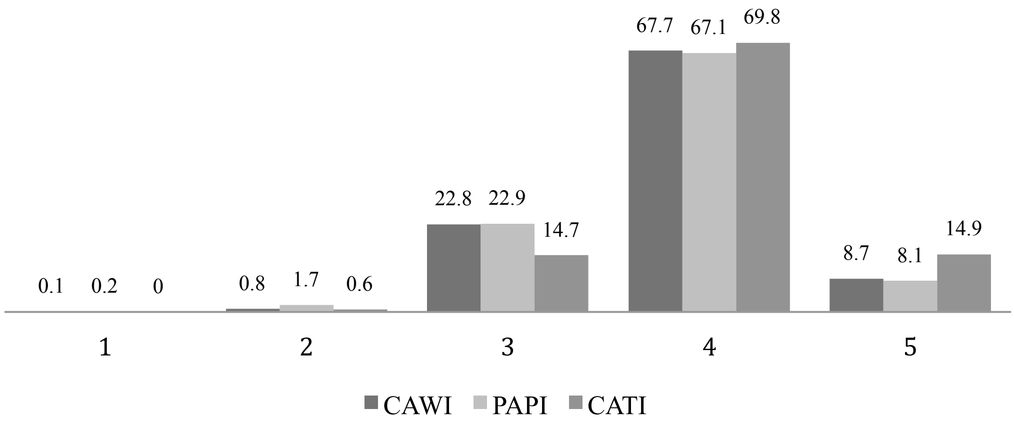


Figure 7. Response distributions, 5-point scale measures (%)

Figure 8 is the final one showing the distribution differences. The figure shows information items with four-category responses. The table shows smaller differences between the percentages. In this case, mail appears to be the mode in which respondents report the two middle categories more frequently, while there are similar numbers of web and telephone respondents choosing the categories 2, 3 and 4.

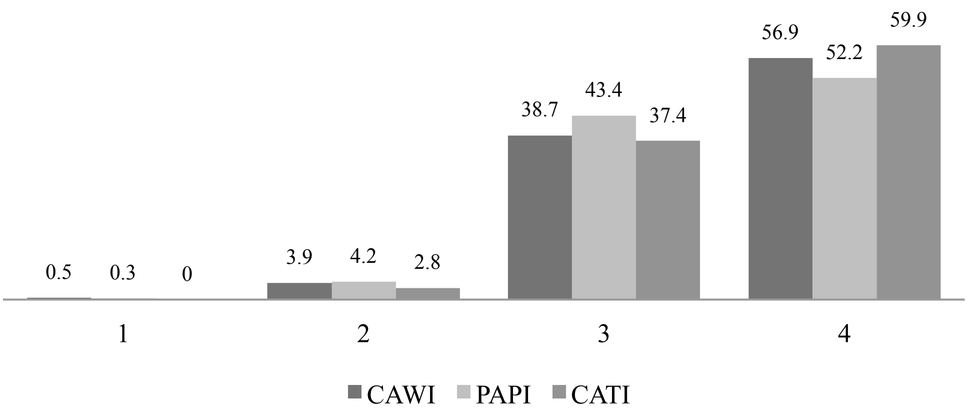


Figure 8. Response distribution, 4-response category measures (%)

The first set of analyses found differences in the SWB mean estimates and distributions and found that the results differ between all modes of data collection, particularly between self-completion and telephone respondents. However, these first analyses did not take into account the differential selection errors between the modes. In the next section, I aim to model the effect of mode of data collection related to measurement effects by adjusting by such selection differences.

4.4.2. The extent of measurement and selection effects in measures of subjective wellbeing

In order to assess the extent of measurement and selection effects, I made three pairs of comparisons in which I examined differences between web and mail, and web and telephone, mail and telephone.

Mode effects in means of subjective wellbeing

The results on the mean comparison showed differences after adjusting for selection effects: a part of the mode effect persists, even if the differences tend to be smaller than in the previous table without the socio-demographic controls.

Results from the mean comparisons after including a propensity score weight shows that there are significant differences for 12 of the examined measures comparing web and telephone results, and 15 for the mail and telephone comparison. After implementing coarsened exact matching, as can be seen in table 9, some statistically significant differences also remain, which can indicate the presence of differences in how respondents answer the questions about subjective wellbeing. The findings reveal that 10 measures show differences in their means when comparing mail and telephone, and 6 when looking at the web and telephone comparison.

Table 11. Means after controlling for selection differences using coarsened exact matching

SWB measures, by question format	Web (1) (n= 889)			Mail (2) (n= 654)			Telephone (3) (n= 364)		
	Mean	Std. Err.	P>t 1 vs. 2	Mean	Std. Err.	P>t 2 vs. 3	Mean	Std. Err.	P>t 1 vs. 3
11 categories									
Social trust	5.35	0.14		5.41	0.16	(*)	5.87	0.12	(**)
Life satisfaction	7.72	0.10		7.52	0.12	(*)	7.94	0.09	†
Happiness	7.89	0.06	(*)	7.59	0.11	***	8.14	0.08	(*)
Take time	6.35	0.11		6.38	0.15		6.62	0.11	†
7 categories									
Meets close ones	5.04	0.07		5.12	0.09	**	5.48	0.06	***
Someone to discuss	2.62	0.08		2.46	0.09	(*)	2.73	0.08	
Gets support	5.06	0.07		4.89	0.10		4.98	0.07	
Gives support	5.25	0.05		5.12	0.09	(*)	5.32	0.05	
5 categories									
Health	4.16	0.04		4.10	0.05	(*)	4.27	0.04	(*)
Optimism	3.76	0.05		3.70	0.06	***	4.04	0.05	**
Positivity	3.67	0.04		3.74	0.05	**	3.99	0.05	***
Freedom	4.05	0.05		4.09	0.06	**	4.39	0.05	***
Accomplishment	3.87	0.04		3.95	0.06	**	4.16	0.04	***
Take control	3.86	0.05		3.82	0.07		3.90	0.06	
Handle problems	4.07	0.04		3.94	0.07		3.93	0.06	†
Things going well	3.90	0.05		3.84	0.07		3.98	0.05	
Overcome diff.	4.02	0.06		3.97	0.07	†	4.14	0.05	
Social activities	2.73	0.05	(**)	2.39	0.07	***	2.77	0.05	
4 categories									
Depression	3.48	0.03		3.42	0.04		3.49	0.03	
Restless sleep	3.25	0.04		3.14	0.05	(*)	3.28	0.04	
Loneliness	3.62	0.04	†	3.51	0.05	**	3.69	0.03	
Anxiety	3.24	0.04		3.22	0.05		3.28	0.04	
SWB at work									
11 categories									
Job satisfaction	7.73	0.12		7.77	0.18	†	8.14	0.11	(*)
Work-life balance	7.00	0.16		7.11	0.18		7.42	0.15	(*)
6 categories									
Interesting job	4.69	0.08		4.87	0.09		5.05	0.07	**
Stressful job	3.45	0.11	(*)	3.85	0.12	**	3.32	0.12	
4 categories									
Expects job loss	3.36	0.05	(*)	3.21	0.05	**	3.44	0.05	

*** p<0.001, **p<0.01, *p<0.05, †p<0.10 Significance results after Wald test

Table 12. Means after CEM

SWB measures, by question format	Web (1) (n= 889)			Mail (2) (n= 654)			Telephone (3) (n= 364)		
	Mean	Std. Err.	P>t 1 vs. 2	Mean	Std. Err.	P>t 2 vs. 3	Mean	Std. Err.	P>t 1 vs. 3
11 categories	6.83	0.08		6.71	0.09	**	7.14	0.07	**
7 categories	4.18	0.05		4.15	0.06	**	4.38	0.04	**
5 categories	3.81	0.03		3.76	0.04	***	3.96	0.03	**
4 categories	3.40	0.03	†	3.32	0.04	†	3.44	0.02	

*** p<0.001, **p<0.01, *p<0.05, †p<0.10

Mode effects in distributions of subjective wellbeing

The main focus of this section, however, is to study the extent of mode effects in the distribution of the subjective wellbeing measures. Table 13 displays the results from the ordered logistic regressions and partial odds models for the comparison of the web and mail samples.

The results of the regression comparing web and paper were very similar independently of which type of control implemented for the differences in sample composition. Even though before adjusting for multiple testing the effect of mode was significant for 4 out of 27 wellbeing variables, once I applied the correction, the p-values for the measures of life satisfaction, feeling able to handle problems, having someone close to discuss, loneliness and the expectation of job loss were greater than 0.05.

Lastly, results after implementing coarsened exact matching showed a different picture before the Holm-Bonferroni correction, as only two measures appeared to be affected by mode. It is possible to summarize that responding in web or e-mail did not affect responses to any of the subjective wellbeing measures.

Table 13. Mode effect coefficients in SWB variables- Web (0) vs mail (1) comparison

SWB measures, by question format		Covariates (n = 1543)			PS (n = 1511)			CEM (n = 1503)		
		Coef.	Std. Err.	P>t	Coef.	Std. Err.	P>t	Coef.	Std. Err.	P>t
11 categories										
Social trust		0.10	0.09	0.266	0.13	0.10	0.191	0.14	0.10	0.178
Life satisfaction		-0.23	0.10	(0.017)	-0.22	0.10	(0.019)	-0.12	0.11	0.239
Happiness		-0.17	0.09	0.068	-0.17	0.10	0.078	-0.06	0.11	0.552
Take time		-0.03	0.10	0.776	-0.03	0.09	0.777	0.04	0.11	0.725
7 categories										
Meets close ones		0.15	0.10	0.152	0.16	0.10	0.128	0.09	0.10	0.333
Someone to discuss	1	-0.29	0.24	0.216	-0.16	0.24	0.501	-0.28	0.28	0.316
	2	0.07	0.13	0.557	0.11	0.13	0.396	0.02	0.14	0.873
	3	0.00	0.11	0.976	0.01	0.11	0.945	-0.01	0.12	0.960
	4	-0.15	0.12	0.222	-0.18	0.12	0.144	-0.22	0.13	0.095
	5	-0.24	0.22	0.276	-0.36	0.24	0.130	-0.52	0.25	(0.035)
	6	0.48	0.35	0.168	0.34	0.37	0.354	0.16	0.42	0.708
Gets support		-0.16	0.10	0.110	-0.16	0.10	0.106	-0.19	0.11	0.084
Gives support		-0.10	0.10	0.308	-0.10	0.10	0.310	-0.14	0.11	0.223
5 categories										
Health		0.01	0.11	0.893	0.02	0.11	0.830	0.05	0.12	0.691
Optimism		0.02	0.10	0.836	0.02	0.10	0.840	0.04	0.11	0.688
Positivity		0.00	0.10	0.975	-0.01	0.10	0.949	-0.04	0.11	0.734
Freedom		0.08	0.10	0.414	0.08	0.10	0.415	0.09	0.11	0.418
Accomplishment		0.17	0.10	0.107	0.17	0.11	0.104	0.15	0.12	0.211
Take control	1	-0.62	0.42	0.143	-0.04	0.10	0.671	-0.04	0.11	0.730
	2	-0.22	0.18	0.238	-	-	-	-	-	-
	3	-0.18	0.11	0.111	-	-	-	-	-	-
	4	0.20	0.13	0.110	-	-	-	-	-	-
Handle problems	1	-0.86	0.40	(0.033)	-0.86	0.40	(0.032)	-0.99	0.44	(0.023)
	2	-0.49	0.21	(0.017)	-0.49	0.21	(0.017)	-0.43	0.23	0.068
	3	-0.34	0.13	(0.008)	-0.34	0.13	(0.008)	-0.28	0.14	(0.048)
	4	0.11	0.12	0.349	0.11	0.12	0.353	0.14	0.13	0.260
Things going well		-0.02	0.10	0.806	-0.03	0.10	0.801	0.05	0.11	0.650
Overcome diff.		0.00	0.10	0.998	0.00	0.10	0.988	0.04	0.11	0.728
Social activities		-0.09	0.10	0.353	-0.09	0.10	0.360	-0.08	0.11	0.472
4 categories										
Depression		-0.07	0.11	0.522	-0.07	0.11	0.508	-0.02	0.12	0.871
Restless sleep		-0.16	0.11	0.134	-0.16	0.11	0.130	-0.05	0.11	0.682
Loneliness		-0.26	0.11	(0.022)	-0.26	0.11	(0.024)	-0.20	0.12	0.112
Anxiety		-0.09	0.11	0.427	-0.09	0.11	0.431	0.04	0.11	0.732

Table 14. Mode effect coefficients in SWB variables- web (0) vs mail (1) comparison

SWB at work, by question format	Covariates (n = 808)			PS (n = 806)			CEM (n = 806)		
	Coef.	Std. Err.	P>t	Coef.	Std. Err.	P>t	Coef.	Std. Err.	P>t
11 categories									
Job satisfaction	-0.03	0.12	0.768	-0.03	0.12	0.783	-0.04	0.12	0.710
Work-life balance	-0.09	0.12	0.454	-0.08	0.12	0.488	-0.10	0.12	0.399
6 categories									
Interesting job	0.15	0.12	0.212	0.16	0.12	0.205	0.04	0.13	0.728
Stressful job	-0.21	0.12	0.080	-0.21	0.12	0.083	-0.21	0.12	0.094
4 categories									
Expects job loss	-0.27	0.13	(0.037)	-0.27	0.13	(0.035)	-0.18	0.14	0.185

*** p<0.001, **p<0.01, *p<0.05, †p<0.10

The following table (15) displays the regression coefficients from the web and telephone comparisons. It is apparent from this comparison that many more measures are affected by mode than when looking at web and mail respondents. It also stands out that the number of significant responses differs in both the amount of significant effects and the categories in which differences between the modes are found. The table shows that there are a total of 14 measures affected by mode independently of the type of correction for selection effects used. The regression with covariates yields 19 measures affected by mode. Although this number is smaller (15) after the controls using coarsened exact matching, it is still a greater number than when using propensity scores as covariates (12).

Some of the differences between the different matching methods could be due to the improvement in terms of standard error in this mode, which may indicate that the coefficients are more reliable. Balancing the samples with this method had particular repercussions for those extreme negative categories (0 and 1) that were not often chosen by respondents in measures with 11 response options. Some of the differences between the CEM and the propensity score and the covariates controls appeared for life satisfaction, happiness, and taking time to do things they enjoy.

Table 15. Mode effects in SWB measures. Web (0) vs telephone (1)

		Covariates (n= 820)			PS (n= 820)			CEM (n= 805)		
11 categories		Coef.	Std. Err.	P>t	Coef.	Std. Err.	P>t	Coef.	Std. Err.	P>t
Social trust		0.37	0.13	(**)	0.38	0.13	(**)	0.49	0.14	***
Life satisfaction	0	16.30	1493.56		15.67	1421.90		13.49	0.61	***
(Base = 10)	1	-18.61	2726.45		-16.25	963.79		-14.46	0.78	***
	2	-1.39	1.18		-1.37	1.21		-1.47	1.17	
	3	-3.23	1.13	(**)	-3.10	1.08	(**)	-2.88	1.06	(**)
	4	0.01	0.62		0.00	0.62		-0.21	0.62	
	5	-0.47	0.35		-0.44	0.34		-0.63	0.39	
	6	-0.84	0.37	(*)	-0.83	0.36	(*)	-0.83	0.37	(*)
	7	-0.56	0.28	(*)	-0.56	0.27	(*)	-0.58	0.29	(*)
	8	-0.63	0.23	(**)	-0.62	0.23	(**)	-0.59	0.24	(*)
	9	-0.67	0.25	(**)	-0.67	0.25	(**)	-0.79	0.32	(*)
Happiness	0	-15.61	2248.90		-14.09	895.03		-13.96	1.07	***
(Base = 10)	1	-16.55	2678.94		-13.95	634.67		-13.96	0.81	***
	2	-17.14	2579.04		-14.15	511.86		-13.96	0.69	***
	3	-1.34	0.76	†	-1.30	0.75	†	-0.81	0.79	
	4	-2.23	0.94	(*)	-2.04	0.89	(*)	-1.74	0.96	†
	5	-0.94	0.41	(*)	-0.91	0.40	(*)	-0.78	0.55	
	6	-0.64	0.35	†	-0.64	0.35	†	-0.52	0.49	
	7	-0.64	0.27	(*)	-0.64	0.27	(*)	-0.41	0.42	
	8	-0.50	0.23	(*)	-0.51	0.23	(*)	-0.13	0.39	
	9	-0.96	0.25	***	-0.97	0.25	***	-0.63	0.40	
Take time	0	-0.17	0.09	†	13.53	607.87		13.13	0.70	***
(Base = 10)	1	15.34	1332.42		-2.76	1.15	(*)	-2.64	1.14	(*)
	2	-2.72	1.12	(*)	-1.52	0.56	(**)	-1.41	0.61	(*)
	3	-1.52	0.57	(**)	-1.40	0.48	(**)	-1.29	0.54	(*)
	4	-1.46	0.49	(**)	-0.68	0.42		-0.95	0.51	†
	5	-0.72	0.43	†	-0.91	0.37	(*)	-0.97	0.45	(*)
	6	-0.97	0.38	(*)	-0.91	0.37	(*)	-0.84	0.45	†
	7	-0.97	0.38	(*)	-1.10	0.36	**	-1.15	0.48	(*)
	8	-1.17	0.37	**	-0.86	0.36	(*)	-0.84	0.44	†
	9	-1.47	0.43	***	-1.39	0.42	***	-1.21	0.50	(*)

*** p<0.001, **p<0.01, *p<0.05, †p<0.10

Table 16. Mode effects in SWB measures. Web (0) vs telephone (1) II

7 categories		Covariates (n= 820)			PS (n= 820)			CEM (n= 805)		
		Coef.	Std. Err.	P>t	Coef.	Std. Err.	P>t	Coef.	Std. Err.	P>t
Meets close ones	1	-16.95	510.26		-17.02	538.19		-17.27	.	.
(Base = 7)	2	-18.09	510.26		-18.11	538.19		-18.33	.	.
	3	-17.15	510.26		-17.21	538.19		-17.45	.	.
	4	-16.99	510.26		-17.06	538.19		-17.36	.	.
	5	-16.42	510.26		-16.49	538.19		-16.95	.	.
	6	-16.64	510.26		-16.72	538.19		-17.25	.	.
Someone to discuss	1	-1.27	0.56	(*)	-1.25	0.55	(*)	-1.58	0.59	(**)
(Base = 7)	2	-2.02	0.50	***	-1.97	0.49	***	-2.04	0.50	***
	3	-1.72	0.49	***	-1.69	0.49	***	-1.80	0.49	***
	4	-1.77	0.49	***	-1.76	0.49	***	-2.09	0.52	***
	5	-1.40	0.49	(**)	-1.40	0.49	(**)	-1.53	0.49	**
	6	-2.44	0.65	***	-2.44	0.65	***	-2.76	0.68	***
Gets support		0.06	0.14		0.07	0.14		-0.03	0.17	
Gives support	1	44.88	3806.68		15.72	1309.07		16.30	0.59	***
(Base = 7)	2	-17.16	6778.16		-15.73	1029.92		-16.39	1.01	***
	3	-19.89	4922.31		-16.20	900.88		-16.39	0.59	***
	4	-0.49	0.40		-0.48	0.39		-0.32	0.38	
	5	-0.20	0.24		-0.21	0.24		-0.14	0.26	
	6	-0.18	0.16		-0.18	0.16		-0.32	0.22	

*** p<0.001, **p<0.01, *p<0.05, †p<0.10

Table 17. Mode effects in SWB measures. Web (0) vs telephone (1) III

5 categories	.	Covariates (n= 820) Coef.	Std. Err.	P>t	PS (n= 820) Coef.	Std. Err.	P>t	CEM (n= 805) Coef.	Std. Err.	P>t
Health		0.44	0.14	**	0.43	0.14	(**)	0.48	0.16	(**)
Optimism	1	-0.30	0.70		-0.23	0.69		0.29	0.72	
(Base = 5)	2	-0.95	0.33	(**)	-0.92	0.32	(**)	-1.11	0.36	**
	3	-1.71	0.25	***	-1.66	0.25	***	-1.67	0.27	***
	4	-0.79	0.19	***	-0.78	0.19	***	-0.79	0.22	***
Positivity	1	-1.26	0.89		-1.25	0.90		-1.43	0.89	
(Base = 5)	2	-1.13	0.31	***	-1.10	0.30	***	-1.03	0.30	***
	3	-1.94	0.27	***	-1.91	0.27	***	-1.94	0.28	***
	4	-0.99	0.21	***	-0.97	0.20	***	-1.09	0.23	***
Freedom	1	0.81	1.16		0.89	1.15		0.86	1.13	
(Base = 5)	2	-0.82	0.36	(*)	-0.81	0.36	(*)	-0.84	0.37	(*)
	3	-1.55	0.28	***	-1.53	0.28	***	-1.68	0.38	***
	4	-1.04	0.17	***	-1.03	0.16	***	-1.05	0.18	***
Accomplishment		0.96	0.15	***	0.96	0.15	***	0.92	0.16	***
Take control	1	0.65	0.62		0.68	0.62		1.09	0.63	†
(Base = 5)	2	-0.37	0.31		-0.37	0.31		0.04	0.35	
	3	-0.60	0.20	(**)	-0.59	0.20	(**)	-0.29	0.27	
	4	-0.73	0.18	***	-0.73	0.18	***	-0.43	0.25	†
Handle problems	1	1.29	0.48	(**)	1.29	0.48	(**)	1.33	0.49	(**)
(Base = 5)	2	0.08	0.34		0.10	0.34		-0.01	0.40	
	3	-0.48	0.24	(*)	-0.46	0.24	†	-0.36	0.27	
	4	-0.74	0.17	***	-0.74	0.17	***	-0.76	0.20	***
Things going well	1	0.58	0.57		0.60	0.57		0.65	0.56	
(Base = 5)	2	-0.11	0.34		-0.11	0.33		-0.42	0.41	
	3	-0.83	0.22	***	-0.84	0.22	***	-0.77	0.24	***
	4	-0.63	0.18	***	-0.63	0.18	***	-0.63	0.21	**
Overcome diff.	1	-0.07	0.67		-0.07	0.66		-0.21	0.69	
(Base = 5)	2	-0.25	0.33		-0.25	0.33		-0.07	0.38	
	3	-0.41	0.20	(*)	-0.40	0.20	(*)	-0.11	0.24	
	4	-0.81	0.18	***	-0.81	0.18	***	-0.63	0.22	(**)
Social activities		0.11	0.13		0.13	0.13		0.04	0.15	
4 categories										
Depression		0.08	0.14		0.08	0.14		0.00	0.17	
Restless sleep		0.17	0.14		0.17	0.14		0.08	0.18	
Loneliness		0.27	0.17		0.26	0.16		0.22	0.19	
Anxiety		0.17	0.14		0.18	0.14		0.22	0.16	

*** p<0.001, **p<0.01, *p<0.05, †p<0.10

Table 18. Mode effects in SWB variables. Web (0) vs telephone (1) IV

SWB at work		Covariates (n= 522)			PS (n= 522)			CEM (n= 504)		
11 categories	.	Coef.	Std. Err.	P>t	Coef.	Std. Err.	P>t	Coef.	Std. Err.	P>t
Job satisfaction		0.36	0.16	(*)	0.35	0.16	(*)	0.33	0.20	†
Work-life balance	0	-1.73	1.17		-1.67	1.15		-1.90	1.17	
	1	-0.72	1.32		-0.93	1.25		-0.93	1.32	
	2	-1.73	0.70	(*)	-1.74	0.70	(*)	-1.81	0.69	(**)
	3	-2.07	0.81	(*)	-2.05	0.81	(*)	-2.02	0.81	(*)
	4	-0.46	0.46		-0.41	0.45		-0.49	0.45	
	5	-0.54	0.36		-0.51	0.36		-0.69	0.37	†
	6	-0.67	0.35	†	-0.65	0.34	†	-0.71	0.35	(*)
	7	-1.05	0.34	**	-1.02	0.34	**	-0.97	0.34	(**)
	8	-1.01	0.30	***	-0.98	0.30	***	-1.03	0.30	***
	9	-1.49	0.38	***	-1.45	0.38	***	-1.85	0.53	***
6 categories										
Interesting job		0.61	0.17	***	0.59	0.17	***	0.65	0.17	***
Stressful job		0.15	0.16		0.16	0.16		0.15	0.18	
4 categories										
Expects job loss		0.36	0.18	(*)	0.35	0.18	(*)	0.27	0.23	

*** p<0.001, **p<0.01, *p<0.05, †p<0.10

Looking at the coefficients in the tables above from the multinomial logistic regression, the way to interpret the coefficients for the multinomial logistic regressions in the tables is to look at the column with the response categories (the second column), whose linked coefficient is estimated in relation to the base outcome. For example, in the case of life satisfaction the effect of mode on category 5 is calculated with relation to the base outcome (10).

Taking happiness as an example, for which the base outcome is 10, it is possible to say that being interview by telephone instead of web, the logit of choosing the response category 2 relative to choosing category 10 decreased by 13.96 log-odd units. For other measures, such as positivity, feeling able to handle problems or feeling that things are going well, all with five category responses, responding by telephone instead of web decreases the likelihood of choosing category 4 relative to choosing category 5. At the same time, responding by telephone increases the log odds of reported a higher level of self-reported health (covariates and CEM), social trust (only in CEM) and accomplishment (independently of the selection differences control method).

The general trend is that respondents in telephone are predicted to choose the most positive and extreme response options more than web respondents, also after adjusting for multiple testing.

Results from the mail and telephone comparison, shown in tables 19 to 21 (below), reveal that there are 16 measures affected by mode, 15 after controlling through a propensity score and 12 after coarsened exact matching. The results present a very similar picture to the previous web and telephone comparison, in which responding by telephone increases the likelihood of giving a more positive and extreme response compared to responding by mail. The mode effect was statistically significant in the cases of life satisfaction and happiness (with telephone respondents less likely to choose categories 5, 4 and 7 relative to category 10), reporting to have close ones to talk to, giving support (only after controlling for CEM), among others. In contrast to the results from the web and paper comparison, mode also affected two of the questions with 4 response alternatives: quality of sleep and loneliness. For the item quality of sleep, responding by telephone decreases the likelihood of choosing category 3, relative to category 4.

Differences between applying covariates, the propensity score, or coarsened exact matching to control for socio-demographic differences between the samples were found in the sense that fewer items appeared to be affected by measurement mode effect after coarsened exact matching, contrary to the results from the previous comparison. In addition, it appeared to work better in the report of differences across modes for the 11-scale measures, for example, the results for happiness when looking at the coefficient for choosing category 2 respective to category 10- effect that would have been missed looking at the results from the propensity score and covariates adjustment.

Table 19. Mode effects in SWB measures. Mail (0) vs telephone (1) comparison I

			Covariates (n = 715)			PS (n = 702)		CEM (n = 638)		
		Coef.	Std. Err.	P>t	Coef.	Std. Err.	P>t	Coef.	Std. Err.	P>t
11 categories										
Social trust		0.25	0.14	†	0.26	0.14	†	0.26	0.15	†
Life satisfaction		0.47	0.14	***	0.48	0.14	***	0.50	0.15	***
Happiness	0	-	-	-	-	-	-	-	-	-
	1	-	-	-	-	-	-	-	-	-
	2	-16.97	1747.92		- 15.28	786.16		-13.96	0.75	***
	3	-1.84	0.76	(*)	-1.69	0.74	(*)	-1.40	0.79	†
	4	-2.65	0.83	***	-2.48	0.82	(**)	-2.30	0.85	(**)
	5	-1.69	0.41	***	-1.64	0.41	***	-1.32	0.44	(**)
	6	-0.46	0.42		-0.41	0.41		-0.40	0.45	
	7	-0.95	0.29	***	-0.93	0.28	***	-1.10	0.31	***
	8	-0.78	0.26	(**)	-0.77	0.25	(**)	-0.83	0.28	(**)
	9	-0.83	0.29	(**)	-0.82	0.28	(**)	-0.83	0.30	(**)
Take time		0.25	0.14	†	0.23	0.14	†	0.24	0.15	
7 categories										
Meets close ones		0.47	0.14	***	0.46	0.14	***	0.39	0.16	(*)
Someone to discuss	1	-0.99	0.55	†	-0.99	0.55	†	-1.24	0.62	†
	2	-1.24	0.50	(*)	-1.25	0.49	(*)	-1.53	0.57	(**)
	3	-1.37	0.49	(**)	-1.35	0.49	(**)	-1.81	0.55	***
	4	-1.67	0.49	***	-1.65	0.49	***	-2.03	0.55	***
	5	-1.12	0.49	(*)	-1.12	0.49	(*)	-1.46	0.55	(**)
	6	-1.86	0.69	(**)	-1.78	0.68	(**)	-1.89	0.74	(*)
Gets support		0.23	0.14		0.24	0.14	†	0.27	0.16	†
Gives support	1	0.29	1.07		0.36	0.96		-1.27	1.26	
	2	-16.11	1507.44		15.50	1123.80		-14.99	0.61	***
	3	-15.81	1048.59		15.08	815.57		-14.99	0.60	***
	4	-0.79	0.40	(*)	-0.78	0.39	(*)	-0.94	0.45	(*)
	5	-0.14	0.26		-0.15	0.26		-0.01	0.29	
	6	-0.35	0.18	(*)	-0.35	0.17	(*)	-0.47	0.19	(*)

*** p<0.001, **p<0.01, *p<0.05, †p<0.10

Table 20. Mode effects in measures of subjective wellbeing . Mail (0) vs telephone (1) II

		Covariates (n = 715)			Propensity Score (n = 702)			CEM (n = 638)		
5 categories		Coef.	Std. Err.	P>t	Coef.	Std. Err.	P>t	Coef.	Std. Err.	P>t
Health		0.32	0.15	(*)	0.29	0.15	†	0.34	0.17	(*)
Optimism	1	0.54	1.12		0.77	1.11		0.00	1.10	
	2	-1.12	0.34	***	-1.06	0.33	***	-0.89	0.39	(*)
	3	-1.74	0.27	***	-1.70	0.27	***	-1.86	0.29	***
	4	-0.98	0.21	***	-0.97	0.21	***	-0.93	0.23	***
Positivity	1	-0.41	1.27		-0.30	1.25		-1.25	1.43	
	2	-0.74	0.33	(*)	-0.74	0.32	(*)	-0.51	0.36	
	3	-1.66	0.28	***	-1.64	0.27	***	-1.57	0.31	***
	4	-0.77	0.21	***	-0.77	0.21	***	-0.72	0.24	(**)
Freedom	1	14.88	1103.10		13.50	553.02		14.68	0.51	***
	2	-1.03	0.37	(**)	-1.01	0.37	(**)	-1.04	0.39	(**)
	3	-1.17	0.29	***	-1.17	0.29	***	-1.10	0.31	***
	4	-0.87	0.17	***	-0.85	0.17	***	-0.84	0.20	***
Accomplishment		0.74	0.15	***	0.74	0.15	***	0.73	0.18	***
Take control		0.27	0.14	†	0.25	0.14	†	0.32	0.16	(*)
Handle problems	1	0.49	0.39		0.49	0.39		0.64	0.42	
	2	-0.11	0.34		-0.13	0.34		0.15	0.36	
	3	-0.52	0.24	(*)	-0.50	0.24	(*)	-0.51	0.28	†
	4	-0.58	0.18	**	-0.58	0.18	**	-0.53	0.21	(**)
Things going well	1	-0.05	0.51		-0.05	0.51		0.05	0.53	
	2	-0.29	0.34		-0.28	0.34		-0.17	0.38	
	3	-0.80	0.24	***	-0.79	0.23	***	-0.69	0.26	(**)
	4	-0.60	0.19	**	-0.59	0.19	**	-0.61	0.22	(**)
Overcome diff.	1	0.08	0.78		0.08	0.76		0.01	0.89	
	2	-0.83	0.31	(**)	-0.82	0.31	(**)	-0.63	0.34	†
	3	-0.33	0.21		-0.31	0.21		-0.29	0.24	
	4	-0.84	0.19	***	-0.82	0.19	***	-0.74	0.22	***
Social activities	1	-0.73	0.45		-0.70	0.45		-0.64	0.50	
	2	0.04	0.42		0.04	0.42		0.21	0.47	
	3	0.19	0.41		0.17	0.41		0.30	0.46	
	4	0.48	0.45		0.46	0.45		0.80	0.50	
4 categories										
Depression		0.28	0.15	†	0.28	0.15	†	0.21	0.17	
Restless sleep	1	-0.37	0.40		-0.37	0.40		-0.49	0.43	
	2	-0.26	0.28		-0.26	0.27		-0.33	0.31	
	3	-0.55	0.17	***	-0.54	0.17	***	-0.54	0.19	(**)
Loneliness		0.75	0.17	***	0.72	0.17	***	0.63	0.19	***
Anxiety		0.26	0.15	†	0.26	0.15	†	0.24	0.17	

*** p<0.001, **p<0.01, *p<0.05, †p<0.10

Table 21. Measurement effects in measures of subjective wellbeing. Mail (0) vs telephone (1)

SWB at work **11 categories**		Covariates (n = 422) Coef.	Std. Err.	P>t	PS (n=421) Coef.	Std. Err.	P>t	CEM (n = 370) Coef.	Std. Err.	P>t
Job satisfaction		0.30	0.18	†	0.33	0.18	†	0.26	0.20	
Work-life balance	0	-0.44	1.51		-0.38	1.53		-0.12	1.45	
	1	-1.82	1.21		-2.02	1.20	†	-1.65	1.25	
	2	-0.87	0.97		-0.57	0.89		-0.97	0.89	
	3	-3.00	0.83	***	-2.80	0.80	***	-2.93	0.84	***
	4	-0.17	0.57		-0.17	0.56		0.24	0.62	
	5	-0.41	0.42		-0.40	0.41		-0.32	0.46	
	6	-0.68	0.39	†	-0.67	0.38	†	-0.78	0.42	†
	7	-1.17	0.38	**	-1.14	0.36	**	-0.97	0.41	(*)
	8	-0.75	0.34	(*)	-0.79	0.33	(*)	-0.71	0.38	†
	9	-1.40	0.41	***	-1.39	0.40	***	-1.26	0.46	(**)
6 categories										
Interesting job		0.31	0.19		0.32	0.19	†	0.38	0.20	†
Stressful job		0.43	0.18	(*)	0.46	0.18	(**)	0.53	0.19	(**)
4 categories										
Expects job loss	1	-0.17	0.69		-0.12	0.69		-0.12	0.71	
	2	-0.66	0.39	†	-0.69	0.39	†	-0.52	0.42	
	3	-1.20	0.23	***	-1.17	0.22	***	-1.11	0.25	***

*** p<0.001, **p<0.01, *p<0.05, †p<0.10

Lastly, table 22 illustrates the effect of mode in the distributions of the composite scores based in the question format. Results show significant differences across all types of question, although the odds ratios values are 1.50 or smaller (1.49) in the case of the questions with 11 response alternatives, and less than 1.30 for the rest of them, indicating very small mode effects.

Table 22. Mode effects by type of question format, odds ratios

SWB measures, by question format	Self-completion (0) and telephone (1) (n= 1907) Covariates OR	Std. Err.	P>t	PS OR	Std. Err.	P>t	CEM OR	Std. Err.	P>t
11 categories	1.50	0.13	***	1.50	0.13	***	1.49	0.14	***
7 categories	1.23	0.07	***	1.25	0.07	***	1.24	0.07	***
5 categories	1.22	0.04	***	1.22	0.04	***	1.20	0.04	***
4 categories	1.10	0.03	**	1.10	0.04	**	1.07	0.04	*

*** p<0.001, **p<0.01, *p<0.05, †p<0.10

4.5. Discussion and conclusion

Prior studies have noted the importance of studying the impact of mode of data collection in measures of subjective wellbeing (Dolan & Kavetsos, 2016; Pudney, 2010; Sarracino et al., 2017). In a context in which many studies draw on mixed-mode survey designs or use different sources of data – such as in country or year comparisons – it is particularly relevant to study the comparability of survey estimates across modes, and this requires an assessment of the extent of mode effects that could compromise comparability. As mentioned in the literature review, mode effects can be due to both differential respondent characteristics and the way in which they respond to the different modes. However, while researchers using mixed mode surveys often seek differences in who responds to each mode, differences in the way responses are measured can hinder the comparability of the data across modes.

An initial objective of the chapter was to identify distributional and mean differences between modes in measures of subjective wellbeing. The findings indicated that telephone interviewing produced systematically higher mean estimates of wellbeing than mail and web questionnaires. With respect to the distributional differences, it was found that telephone respondents were more likely to choose the most positive response alternatives, and mail and web the middle-options, although there were small variations depending on the format of the question and the measure examined. However, these results did not take into account that the differences in nonresponse and coverage for the different modes, even though a comparison of the respondents' socio-demographic characteristics had shown differences in nonresponse across the modes. Such results are in accordance with previous findings on the relationship between mode effects and response styles (Dolan & Kavetsos, 2016a; Pudney, 2010).

The second question in this study sought to examine how much of the mode effects on measurement persist once mode effects on selection were controlled for. After having applied three different ways of adjusting for differences in selection between the modes, I found that some measures of subjective wellbeing remain sensitive to mode after controlling for mode effects on selection through different methods of matching and weighting the data. Differences were found between telephone and self-completion, but not between the mail and web modes. In the case of the differences in the distribution of the responses, they were not statistically significant after using coarsened exact matching.

What was surprising is that the results after the different controlling stages were not consistent for all the SWB variables. For example, responses about happiness, which were not affected by mode in the regression with covariates and in the regression with a propensity score, appeared to be mode sensitive after coarsened exact matching. The results showed that fewer measures were sensitive after applying either the propensity score or coarsened exact matching adjustment, compared to the covariates as controls option, and both ways of controlling for sample differences improved the balance between the samples. However, the overall results show that coarsened exact matching was able to control for sample imbalance to a higher extent than propensity score matching, particularly when looking at the mail and telephone comparison. The different ways in which the survey participants are paired to balance the different modes' samples (King, Nielsen, Coberley, & Pope, 2010) may be related to the few discrepancies that were found, which was to be expected.

Furthermore, although both CEM and propensity scores improved the balance of the samples, regression standard errors after coarsened matching were smaller for those measures in which only a small number of respondents had chosen the most negative response options in the 11-category questions (particularly for life satisfaction and happiness).

With respect to the third research question in which I aimed to gain a better understanding of the source of the measurement effect, I found that mode can have an impact on responses depending on the number of response categories a question has, and this can affect both means and distributional differences. Comparing means across modes of data collection, I found significant differences in all cases except in responses to questions with 4 response alternatives. In addition, mode of data collection had the strongest effect in the responses to 11 category questions, as results showed they had the higher value of odds ratios. However, these differences were small and possibly due to both question format and the level of sensitivity as perceived by the respondent.

In this investigation, there is one main source of uncertainty. In the back-door approach selected, it is particularly important to include all the variables linked to the selection effect, but at the same time it is essential not to include variables that may have been affected by measurement effects (Vannieuwenhuyze et al., 2010). By including socio-demographic controls, I only control for the correlations of the characteristics that predict differential response to the different modes (Lipps, 2016), and therefore a part of the selection effect might have been missed in the analyses implemented.

To conclude, I found that there are some measures of subjective wellbeing that are sensitive to mode. In particular, I found that answering by telephone makes respondents more likely to choose the most positive response alternatives compared to responding by mail and web. The findings presented in this chapter support the notion that implementing a telephone survey can increase the likelihood that a person reports a higher level of wellbeing compared to web and paper mode. This is, however, not true for every wellbeing measure, and could be related to the response format and the sensitiveness of the subjective wellbeing measure of interest. The results might further indicate that mixed-mode surveys consisting of mail and web can be adequate to measure wellbeing, although it is important to note that systematic differences were found when comparing the self-completion to the telephone responses about wellbeing. However, the extent of the mode effects depends on the type of adjustment made, which means that this could make a difference in whether reports of widely used measures such as happiness or life satisfaction suffer from mode effects or not.

There are still many unanswered questions about the impact of mode of data collection in the subjective wellbeing results: the findings from this study showed that the mode effects, when found, are overall small and therefore may not have a significant impact in the types of analyses used by social science researchers. However, they do indicate that mean comparisons pose a bigger problem, even when looking at modes of data collection that share characteristics such not involving an interviewer in the data collection process. Although differences in the estimates may still be due to selection differences, adjusting for socio-demographics does not appear to be enough to allow for such comparisons, which could have implications for cross-country comparisons. Furthermore, the impact of mode could also be important in additional ways, for example, it may also have an effect on responses to open-ended questions, it may not affect all respondents equally, and it is not clear whether mode effects matter in substantive research on wellbeing. In the following chapters I develop these ideas in order to better understand the mechanism underlying the impact of mode on wellbeing studies. In particular, in the next study I will examine the impact that mode effects have in responses to open-ended questions on the topic of life events.

CHAPTER 5. MODE EFFECTS IN ANSWERS TO SENSITIVE OPEN-ENDED QUESTIONS

5.1. Introduction

In social science surveys, closed-ended questions are the most common way of obtaining information from the population of study (Krosnick, 1999). Closed-ended questions are often regarded as easy to respond to – the respondent does not have to phrase the answer – and as uncomplicated tools for social science researchers due to the ease for coding and comparing responses to the same question (Reja, Manfreda, Hlebec & Vehovar, 2003). In spite of this, closed-ended questions are not adequate for every situation. Open-ended questions may require more effort, as the respondent has to answer in his or her own words, but they are sometimes chosen over closed-ended questions for their capacity of getting a deeper understanding of respondents' attitudes, behaviours and/or opinions. For this reason, they are popular in surveys measuring social phenomena (Revilla & Ochoa, 2016).

Despite having some advantages over closed questions, concerns about the extent to which mode of data collection determines survey quality are common to both types of questions. But in contrast to mode effects on closed-ended questions, there is much less information about the effects that mode has on open-ended questions, particularly so when questions ask about sensitive and personal issues.

Moreover, there is little published research on the comparison between self-completion and interviewer-based modes for this type of question. Some research has been carried out on the topic (Börkan, 2010; Denscombe, 2008), but few studies have addressed the problem of confounding selection and measurement differences (see Chapter 2 for an explanation).

The existing literature offers contradictory findings about the effect of mode on open-ended questions, which appear to depend greatly on the topic of study as well as the population researched. Nevertheless, recent literature has emerged (Butz, Waiters, Deatrick & Usher, 2013; Kwak & Radler, 2002; Smyth, Dillman, Christian & McBride, 2009) that observed that different modes might cause different levels of item-nonresponse, different levels of response detail, and different levels of disclosure. Such results still depend on the conditions of the studies. For example, researchers are often constrained to studying small populations such as students or teachers. As a result, to date there has been little agreement on the role of mode of data collection in the quality of responses to open-ended questions.

In the case of wellbeing and vulnerability studies, open-ended questions are important for gathering information that is not easily obtained using closed-ended questions. For example, questions about critical events that have marked respondents' lives and studying both positive and negative life events are key measures in life course and vulnerability research, as having gone through critical life events can be related to " stress or chronic stress" (Spini, Bernardi & Oris, 2017), or illness (Lydeard & Jones, 1989).

In this chapter, I present new evidence for understanding how mode of data collection influences responses to sensitive open-ended questions about events that have marked respondents' lives. I set out to gain further understanding of mode effects on measures of subjective wellbeing. The key research question is: To what extent do responses to open-ended survey questions vary as a function of the mode of data collection? To answer the question, I employed a mode comparison approach, using data from the mixed-mode experiment obtained through web, mail and paper to explore different survey quality indicators. In particular, I examine differences in item nonresponse, response length, theme of the response, and positivity.

This chapter begins by introducing previous literature relevant to the use of open-ended questions in social science research. It will then go on to show how mode of data collection has been found to affect responses to open-ended questions. The second part describes the analytical approach utilised to explore the different elements of survey quality for open-ended questions. Afterwards, I present the results and include a discussion of the implication of the findings for future research on mode effects in responses to sensitive open-ended questions.

5.2. Literature Review

Open-ended questions are popular in social-science research – even if not as widespread as questions with an already defined set of response alternatives – because they have a series of advantages over closed-ended questions (Emde, 2014).
Emde (2014) explains how one of the main advantages is that they allow for more detailed answers which can be both quantified or qualified by the researcher at a later stage and tailored depending on the research question and the responses obtained. Also, by not offering a set of answer categories, the influence of the research tool in the answers can be smaller, by not suggesting what the “right” answers are. Researchers can also ask questions that require narrative answers, which can be as long or short as the respondent wants. In addition, and most importantly, the

advantage is not only how much detailed information the researcher requires but also the fact that closed-ended questions can fail to provide an adequate set of response alternatives, not allowing the respondent to choose their desired answer. This can be the case when measuring subjective feelings, opinions, or behaviours. Finally, when dealing with sensitive topics, or topics that have to be answers with a numerical quantity, open-ended questions have on some occasions been found to provide better information than close ended ones. For example, Bradburn and his colleagues (1979) found that people report higher levels of alcohol consumption when answering open-ended questions compared to closed-ended ones. Even including a final open-ended question such as “Do you have any other comments?” can be useful to get important insights on the research topic that would have been otherwise left out (Mclauchlan & Schonlau, 2016). An example of this is provided by Daniels, King, Smith, and Shneerson (2001) in their assessment of life quality of people that suffer from narcolepsy. In this study, this final open-ended question provided useful, additional information about the aspects that determined their subjective wellbeing, such as the fact that they were not able to have a bath if alone at home.

For these reasons, Tourangeau and Smith (1996) argue that open-ended questions can increase the accuracy of the information provided by the respondent. In other less flexible types of questions, respondents may choose response alternatives that – even though “acceptable” – are only an approximation of what they would have said in an open-ended question.

Open-ended questions can also pose a series of problems to researchers working with them. Although there are disagreements on the matter, researchers such as Reja and his colleagues (2003) and Denscombe (2008) found that non-completion rates in open-ended questions are particularly high when compared to closed-ended questions. Some other concerns often involve the capacity of respondents to provide elaborate information, the fact that the extra effort they involve can affect response rates, and the excessive coding work needed to study such responses. Moreover, there is limited support for the claim that closed-ended questions only provide “acceptable” response alternatives, as long as the question is well designed (Burton & Blair, 1991) and, together with a lack of research on the study of measurement effects in open-ended questions (Saris & Gallhofer, 2007), has made researchers reconsider the convenience of open-ended questions.

5.2.1. Survey quality and open-ended questions

The quality of an open-ended question is often associated with the length of the response, because a lengthy answer can be related to more detailed and thorough information. In fact, one of the assumptions often made by researchers that study open-ended questions is that response length is an indicator of its usefulness (Walsh & Brinker, 2016). This is not necessarily the case: there are still doubts about whether less wordy answers are less informative than longer ones (Walsh & Brinker, 2016). In addition, there is the possibility that, depending on whether it is respondents who write the responses or the interviewers, responses are recorded superficially in comparison to what someone may write by his or herself.

Item non-response is another concern of this chapter. While nonresponse for closed-ended questions is a popular research subject, not many studies look at open-ended question nonresponse. Some of these studies show that open-ended questions produce higher item nonresponse rates than other question styles, regardless of mode of data collection (Börkan, 2010). However, other researchers argue that item nonresponse in open-ended questions is related to a lack of interest in the topic of the questionnaire or an inability to answer the question, not just because the question is open-ended (Denscombe 2008). Their argument is that the reasons for item nonresponse apply equally to closed-ended questions: not being able to give a response, for example, because respondents cannot understand what the question asks for (Beatty & Herrmann, 2002) and a lack of motivation to respond the question (Scholz & Zuell, 2012).

5.2.2. Mode effects on open-ended questions

Item non-response and mode of data collection

Research on item-nonresponse offers contradictory findings. In some occasions, when responses to open-ended questions were compared between different modes of data collection, no statistical significant result was found. However, this has not been the case in all implemented research. For example, in a comparison of responses to e-mail and mail questionnaires, Schaefer and Dillman (1998) showed that open-ended question achieved a 12% higher completion rate for the e-mail version than for the paper version when examining responses given by faculty members of the Washington State University. Denscombe (2008) examined the length of answers to mail and web questionnaires controlling for socio-demographics (sex, age and educational aspirations) and found no significant differences between the amount of information given. In the study, he looked at teenager students, which means that perhaps the most serious limitation of both studies is that they apply to specific sub-groups rather than the general population.

Börkan (2010) obtained the same result when he examined responses between mail and web given by teachers about their level of job satisfaction. As was the case in the research by Denscombe, it may not be possible to extrapolate the results to the whole population due to the sample being exclusively composed of teachers. In addition, he adjusted the differences in selection error using socio-demographic information about age, sex, ethnicity, years of experience, level of education and household income.

Mode comparisons that include an interviewer-based mode as a way of collecting data have revealed differences in response length between interviewer-based modes and self-completed responses. The study implemented by Fricker, Galesic, Tourangeau and Yan (2005), in which some open-ended questions were included, found a higher number of respondents that did not answer certain questions in telephone than in web modes. They argue that mode effects can be due to interviewers recording responses superficially, and therefore insufficiently, in comparison to what someone may write by his or herself. In spite of this, interviewers have also been found to have the opposite effect: they can clarify and read silences, knowing when to incentivise more answers (Dillman & Christian, 2005), although many times interviewers do not make any difference in the response process: Dillman and Christian (2005) show that differences in item nonresponse between face-to-face and web are less than 1%. They argue, however, that there is a lack of conclusive results and that there is a need for a reliable method that provides consistent results. A possible explanation for if web item-nonresponse is high could be that respondents may be more insecure about the anonymity of their answers than respondents to other modes of data collection, especially if they think their information is linked to an e-mail address. Some research has shown that this does not pose a problem and should not be a subject of concern (Couper, Blair & Triplett 1997), although Dillman and Christian suggest that this may depend on the sensitivity of the topic of study.

Response length

Previous research has found evidence that the amount of information respondents give can depend on the mode of data collection (Schaefer & Dillman, 1998; Smyth et al., 2009). When comparing modes, it is possible to argue that if there are length differences across the modes, the information content obtained through modes associated with lengthier responses is more informative than that obtained by modes that are associated with shorter responses. When the length of the answers is similar, the content is assumed to be of similar quality as well. However, it has been found that, although responses to modes that involve oral responding are often longer, they also contain less useful information than written responses (Ravid & Berman, 2006).

When focusing on question length, researchers show that responses to web surveys tend to be longer than responses to the mail ones (Schaefer & Dillman, 1998). Smyth and colleagues (2009) corroborated this situation: they found that responses to an Internet survey had an average of 40 words, compared to the less lengthy mailed responses, whose average length was 10 words. Responses in the web mode were not only lengthier but also more detailed (Smyth, Dillman, Christian & McBride, 2009). This can be due to respondents finding that writing their answers in paper and pencil requires too much effort compared to how easy it can be for some to type answers into a computer. The authors indicate, though, that other type of mode effect could not be detected because both of the examined modes are self-completed (Schaefer & Dillman, 1998).

Previous studies reported mixed conclusions (Denscombe, 2008; Dybdahl, Shaw & Blahous, 1997) about differences of response length. In particular, Denscombe's web and mail comparison did not show a marked difference in terms of response length for the different types of questionnaire. The average number of words for the paper mode was 13.6 compared with 13 for the online questionnaire, and therefore not statistically significant. Several studies have investigated the aforementioned different (Christian, Dillman & Smyth, 2007; Couper, Conrad & Tourangeau, 2007; Couper, Traugott & Lamias, 2001) and found that the design and size of the response space can influence the way in which the respondent chooses his or her answer. This way, larger response fields tend to get longer answers than smaller ones. Additionally, mail and web surveys tend to have different designs. For example, the mail questionnaire's response field can be a line while the web one is a square, which may give the impression of allowing more space to write the response. Although it is not clear what exactly influences the difference in length, it appears that both shape and size of response field affect response length and quality (Fuchs, 2009). Denscombe (2008), and Stern, Dillman and Smyth (2007) found that different age, sex and education had a mediator effect in the responses given to open-ended questions.

In his work, Börkan (2010) examined the amount of useful information provided by respondents and compares how it differs between web and mail modes, concluding that there is no significant evidence that mode has an effect on this aspect.

To summarize, previous research shows that open-ended questions using the Internet suffer from less item nonresponse and tend to be more complete than mail ones. In fact, some of the authors discussed in this section discourage the use of open-ended questions in mail surveys because of their tendency to suffer from reduced data quality.

Response content and mode

In the previous chapters, I showed how respondents sometimes tailor their answers depending on the mode of data collection and the perceived level of privacy, particularly when giving information about sensitive topics (Kreuter, Presser & Tourangeau, 2008). The general trend is that respondents are less likely to report embarrassing or non-socially desirable behaviours, opinions or attitudes when an interviewer is involved in the data collection process than in self-completed questionnaires. This is the case for both closed-ended and open-ended question types (Tourangeau, Rips & Rasinski, 2000).

5.2.3. Life events and open-ended questions

The types of meaningful life events present in the previous literature depend on the measurement instrument that has been used to raise the information. It can either be a closed-ended question in which the respondent has to “check” all events that apply, it can be a general question on whether they went through any negative or positive events that affected their life, or it can be open-ended questions or semi-structured interviews that ask the respondent to retrieve the information and express it in their own words.

Such events considered in the literature can be illnesses (Bevan, Gomez & Sparks, 2014), death of a close ones, rape, having had depression, being a parent of children with some type of illness or having survived an accident or catastrophe (Park, 2010). Live events are not necessarily negative, however: for example, the life events that Trappman and colleagues (2015) are able to study are divorce, which can be seen as positive, childbirth and having a new job. Other potentially positive events are moving home (Cleland, Kearns, Tannahill & Ellaway, 2015). Cleland and colleagues look at the impact that serious health episodes, being victim of a crime, getting a new job and separating from a partner have on health. Getting a job was found to be the most positive effect on health. Other events considered important could be getting married, having a child, falling in love, death of a parent and starting school (Dickinson et al., 2016).

Studies involving questions on life events often focus on threatening events, that may provoke depression, depending on the type of the event and the characteristics of the person that suffers the event (Brown, Bifulco & Harris, 1987). For example, Garnefski, Kraaij and Spinhoven (2001) investigate the coping strategies that people adopt after a negative life event, which are important mediators between the negative event and suffering from depression and anxiety. They studied negative or unpleasant events without identifying the particular event, instead asking a general question about whether respondents had suffered from an important negative event.

Spence and his colleagues (2015) argue that life events are not always measured in an effective way with respect to context and timing. Indeed, when analysing negative meaningful events, which are often related to suffering from some kind of loss, danger or humiliation, checklists are not seen as a good approach to ask about life events. It is often preferable to have information on the context in which they took place so that the researcher can decide how meaningful an event could have been for a particular life course. This is to say, when studying life events it is important to take into account the context in which the event happened. For this reason, semi-structured interviews are popular in the field (Spence et al., 2015), as are life history calendars (Morselli et al. 2016).

Measuring such events and the meaning for the respondents has been the object of discussion, as biases – both related to nonresponse and to measurement – can appear if someone that suffers or has suffered depression is asked about the meaning of the event. Trappman, Gramlich and Mosthaf (2015) study the impact that previous major life events have on attrition in panel surveys, expecting them to have a strong impact on whether survey participants respond or not. This effect can be explained by socio-demographic and psychosocial characteristics (Lemay, 2009). The socio-demographic effect is based on the idea that the status of the respondent has changed: for example, they live in a different home, or they are now divorced and cannot be reached any longer; the psychosocial element is linked to the person being in "shock", which makes them more reluctant to respond.

5.2.4. Research questions

The previous section has shown that there is still mixed-mode evidence on whether mode of data collection has an impact on responses to open-ended questions, and specifically when the questions ask survey participants about important life events. Evidence for the effect of mode on open-ended questions has been mixed, and although there is a base of research that covers the topic, the influence of mode on sensitive and personal open-ended questions is not fully understood. The aim of this chapter is to develop an understanding of the extent to which measurement related mode effects impact on responses to questions on life events. In particular, this research seeks to address the following questions:

R.Q.1: Is the level of item-nonresponse to open-ended questions on life events significantly different for the web, mail and telephone samples?

Based on the findings of previous research, I expect that item nonresponse will be higher for mail than for web, although there is the possibility of no significant difference between responses to the web and mail modes. In addition, having an interviewer asking the questions may encourage respondents to give an answer.

R.Q.2: Is response length to open-ended questions on life events significantly different for the web, mail and telephone samples?

Some of the previous research showed that responses to the web mode are longer and more elaborate than to the mail mode, and at the same time responses given orally can also be still longer, although this may depend on how the interviewer records the answers.

R.Q.3: Is response content to open-ended questions on life events significantly different for the web, mail and telephone samples?

I expect that respondents to the telephone mode will mention positive events to a higher extent than respondents to the self-completion modes (web and paper). At the same time, I expect respondents to report fewer socially desirable answers (more negative events) in the self-completion modes. There is no previous evidence on differences in the types of life events that respondents mention in different modes of data collection.

R.Q.4: Does controlling for mode related selection errors on socio-demographic variables reduce or eliminate mode related measurement errors?

Previous studies that control for sample differences between the modes found fewer differences on level of item-nonresponse and response length. I expect that controlling for respondents' characteristics such as sex, age or nationality will have an impact on the results to the previous questions.

5.3. Methods

A number of techniques have been developed to analyse responses to open-ended questions, which often consist of narrative text. The coding and analysis of responses to open-ended questions are often regarded as more labour intensive for responses to questions with a pre-established set of response alternatives. In spite of this, there is a variety of statistical software that facilitates the analysis of raw text (Schonlau & Guenther, 2016), providing tools that match keywords present in the raw text to variables and categories defined by the researcher.

5.3.1. Data

Data used in this section is the same as for the previous thesis chapter. The population of study is the same as in the previous chapter and includes those survey participants that answered in their allocated mode of data collection.

5.3.2. Variables

The outcome variables in the analysis were three open-ended questions included in the mixed-mode experiment that asked about important life events which happened at some point in respondents' lives included in the mixed-mode experiment. The questions were the final ones in the subjective wellbeing section and were situated before the questions on socio-demographic information. They were introduced by a short introduction that read: '*First of all, we would like to know about the life events that have marked your life, afterwards we would like to know whether their impact has been positive or negative. Which were the three events which had the biggest impact in your life?*'[3] Following the introductory text, the respondent has to phrase his or her responses about the three events, either orally if responding to the telephone survey – interviewers were required to transcribe the answers – or writing in the

[3] French version is: '*Tout d'abord, nous aimerions aborder les évènements qui ont marqué votre vie, puis nous aimerions savoir si son impact a été positif ou négatif.* Quels ont été les 3 événements les plus marquants de votre vie ?'

response box provided in both web and paper survey. Respondents were asked to give as much detail as possible.

Afterwards, they were asked whether the event had a positive or a negative impact. There was one question per event which mentions the event number (1, 2 or 3) and with the following five response alternatives: very negative, mostly negative, as negative as positive, mostly positive or very positive.

This chapter consists of two main parts, corresponding to the descriptive analysis and the analytical approach. These two sections are at the same time divided depending on the type of statistical technique used and research question investigated. For this reason, when explaining the analysis that I used, I will briefly mention which is the corresponding research question.

A series of new variables were created in order to measure different aspects of survey quality. Three variables measure item nonresponse for each life event measure. Each variable on item nonresponse measures whether the respondent answered (1) or did not answer (0). The questionnaire did not provide response options for those respondents that did not report any event. Another three variables measure response length: they are numerical variables created using the "wordcount" function in Stata, that automatically calculated the number of words given by the respondent if they answered (the minimum value is 1). Additional variables were created to measure the themes/life events mentioned by the respondents. As there are three questions about life events, I created three variables about the theme reported. I did this using automated text mining with Stata (*ngram* and *screening*). These functions allow one to visualise the most common words mentioned by respondents and to create a new variable in which is respondent gets the value 1 is certain theme is reported, and 0 if it is not. The analysis of narrative text is not as straightforward as the analysis of responses to closed-ended questions that are often associated to a numerical value. In spite of this, there is adequate and easy to use software that aides that analysis of raw text, providing tools that match keywords present in the responses to variables defined by the researcher. Completely automated text mining is often less accurate than manual categorization (Schonlau & Couper, 2016). For this reason, the approach I followed is semi-automatic. Before converting the narrative responses into numeric variables, I examined the information given by the respondents in order to find patterns and common subjects. This way, using a qualitative method of content analysis, text segments that made reference to different types of life events were identified by a keyword. Once I identified the commonly mentioned events (for example, marriage), it was possible to create categories that comprise keywords related to the same event (for example, having had children, having given birth). To create the categories, I started by looking for the keywords previously identified in the theoretical framework section. I examine the number of respondents that mention keywords which identify each major life event and automatically create a new variable that takes the value 1 if the word is present in the response, and 0 if it is not. In the following lines, I show the new variables created, which are three for each life event measure, together with the keywords that were used to identify the event. The entry of data onto the computer is different for each mode of data collection, and for this reason the phrasing of the keywords is not always the same: sometimes there are spelling problems due to computerisation, sometimes because they were written in English or Italian. Although marginal, this situation could have an impact in the number of words computed, which I do not control for in this study. Using this program, the words were identified and taken into account for the analysis. For this reason, the *screening* tool is able to select words that have 4 or 5 letters in common, thanks to which I was able to identify the most common mistakes in terms of accents and spelling. The events used to create the new variables that describe whether a respondent mentions each event are:

Table 23. Life events and identified keywords

Event	Keywords
Marriage	Marié mariage, matrimonio
Death of someone close	Deces, dècës, mort, deuil, suicide, perte de, perte d’
Having studied and gone through exams	Ètudes, examen, scolaire, diplùme
Having had an accident to oneself or someone close	Accident
Having moved to Switzerland	En Suisse
Having had children	Naissance, enfant
Having suffered an aggression, been abused/raped	Abus, agression, viol
Illness and other health related issues	Maladie, santè, hùpital, hospital
Divorce and separation, relationship break-up	Divorce, separation, sèparation, rupture
Parents’ divorce:	Divorce de
Having met loved and important people in live	Rencontre
Profession	Profession, stage, job, travail
Success	Rèussite
Job loss	Chùmage, licencement, perte travail
Moving home	Demenagement, déménagement, dèmènagement
Travelling	Voyage

For example, for the first open-ended question, 1,267 of respondents mention one of the identified events (N= 1,907, 175 nonrespondents, the rest mentioned another non-identified event). Widely mentioned events are marriage (284), death of someone close (245), the suicide or attempted suicide of someone close (13), having studied and taken exams (32), having had an accident affecting oneself or someone close (31), having moved to Switzerland (45), having had children (340), having suffered an aggression, being abused or raped (10), illness and other health related issues (51), divorce and separation (45), parents’ divorce (41), finding loved and important people in life (73), having done an internship, moving home (22), travelling outside of Switzerland (27), unemployment (2). The distribution for the other two life event questions is similar (see annexe for more detail on the distribution of identified life events for the second and third life event questions).

In addition to the indicators of response quality for open-ended questions and mode of data collection, I used the same socio-demographic variables as in the previous chapter to adjust for differences in selection error. These variables are sex, age, nationality, whether the respondent has a listed fixed telephone number, whether they have a partner, living area and use of the Internet.

5.3.3. Analytical approach

The first step in this analytical process is to give descriptive information about the level of item non-response for each life event question; the mean of words for each life event question; the commonly mentioned key words per life event question and the amount of negative and positive themes. In particular, I show results on:

1) The level of item non-response for each life event question, for each mode

I calculated the number of respondents that answered the questionnaire but decided not to give information on life events.

2) The mean of words for each life event question, for each mode

In order to see if the different modes of data collection have similar minimum, maximum and average number of words.

3) The most mentioned life events question, per mode

After having identified the common life event themes, to compare the popularity of each event by mode of data collection.

4) Amount of negative and positive themes, per mode

The follow-up questions about the events' positivity make it possible to know whether respondents tend to give information that is more or less positive in different modes. Complementing the previous section on life event themes, this positivity/negativity variable is also important because the same kind event can have different repercussions depending on respondents (for example, a marriage, a divorce or giving birth can be seen as either a positive or negative event by respondents).

To answer the previous questions, I compare the distribution of the different modes' samples for each variable and implement chi-square tests that indicate whether differences are significantly different for the different modes. In addition, for the variable that measures the number of words per question, I implement an analysis of variance and covariance (ANOVA), which shows whether the average length is different between mail, web and paper modes.

In addition to the descriptive analysis, and to establish the extent of mode effects in the responses to life event questions, I implement a series of regression analyses in which the dependent variables are again the level of item non-response for each life event question; the mean of words for each life event question; the commonly mentioned key words per life event question and the quantity of negative and positive themes.

To avoid confounding effects as much as possible, I controlled for differences in sample composition using socio-demographic information on age, sex, nationality, partner status, area of living, whether the respondent has a fixed-phone number and use of the Internet. To control for selection bias, I used the coarsened exact matching presented in Chapter 4 for adjusting for selection error differences between the mail, web and telephone samples. Accounting for the sample compositions makes it possible to establish whether there are measurement differences between mail, the Internet and telephone modes while controlling for differences on observed variables in the composition of samples responding in different modes.

To test whether there are measurement effects, the instruments chosen are logistic regression for addressing the question on whether mode has an impact on level of item nonresponse, a negative binomial regression for the variable that measures the length of responses (a count variable), logistic regression to test whether respondents to different modes systematically mention different life events, and finally ordered logistic regression to learn whether respondents tend to give information about events of different positivity depending on mode of data collection. For all the regression analyses, the dependent variable is the one that makes reference to the outcome of the question on life events, and mode of data collection is the independent variable. To better understand the role of selection effects, I implement the regression analyses twice. The first step consists of mode of data collection as the only independent variable, and the second step, in order to establish the extent of measurement differences, includes the propensity score that is used to control for the socio-demographic differences. The hypotheses being tested are related to whether differences between the modes are significant of not, with the null hypothesis meaning that responses to the web, mail and telephone are not significantly different. Together with the regression coefficients, I present the p-values for the regression that indicates the strength of the relationship between mode and response. A p-value of less than 0.05 indicates a significant difference, and the smaller the p-value is, the stronger the relationship. In addition, I correct the p-values using the Holm-Bonferroni approach and I implemented the Pearson goodness-of-fit tests that indicated a good fit for the logistic model (Prob>F=1.000).

5.4. Results

5.4.1. Findings on item nonresponse

The first question aimed at finding out whether there were differences in the level of item-nonresponse to questions on important life events that were asked in three different modes of data collection: web (CAWI), mail (PAPI), and telephone (CATI). In order to make this comparison, I calculated the number of respondents that answered the questionnaire but decided not to give information about the life events that marked their lives for each mode of data collection. Table 17 (see below) displays the results obtained, showing that item-nonresponse levels varied depending on mode, and particularly when comparing the self-completion modes to the interviewer-based mode. There were also differences between the questions being answered. For instance, when looking at the results for the first question, 5% of the telephone respondents decided not to answer the question, which is half the amount of web and paper. The second question on life events had a higher level of nonresponse in all modes, although the increase was bigger in telephone – with a 4 point-difference – than in web and mail (1 and 3 point-differences respectively).

Table 24. Item nonresponse for each life event (%)

	Life event 1 (%)	Life event 2 (%)	Life event 3 (%)	Overall (%)
Telephone (n = 654)	4.95	9.34	20.6	11.6
Web (n = 889)	10.24	11.59	14.74	12.2
Mail (n = 364)	10.09	13.15	14.74	12.7
All	9.18	11.69	16.94	12.6

In the previous table (see table 24), differences on item nonresponse could be due to differences between the web, mail and telephone respondents due to selection. For this reason, it is necessary to test whether differences remain after controlling for such differences between samples.

Table 25. Relationship between item nonresponse and mode of data collection

	Web (0) vs. Mail (1) (n = 1432)			Web (0) vs. Telephone (1) (n = 653)			Mail (0) vs. Telephone (1) (n = 1018)		
	Odds ratio	Std. Err.	Sig.	Odds ratio	Std. Err.	Sig.	Odds ratio	Std. Err.	Sig.
Life event 1	1.50	0.32	†	0.35	0.12	**	0.42	0.16	**
Life event 2	1.24	0.23		0.60	0.16		0.68	0.20	†
Life event 3	1.00	0.16		1.38	0.24		1.26	0.28	

Note- Results after coarsened exact matching accounting for selection differences.

*** p<0.001, **p<0.01, *p<0.05, †p<0.10

As can be seen from the table 25 (above), there were no statistically significant differences on item nonresponse between the different groups after controlling for selection errors in the comparison of the mail and web samples. However, the results obtained by the same regression analysis showed significant differences in item-nonresponse levels between telephone and web for life event 1 (OR = 0.35, p < 0.02) and between mail and telephone, also for life event 1 (OR = 0.42, p < 0.01). These results indicate that responding using the web and mail modes increases the likelihood of not responding to the first open-ended question. There were no significant effects for the other life event questions nor for the other mode comparisons.

5.4.2. Findings on response length

The second set of analyses examined the impact of mode on the number of words that respondents gave to open-ended questions on life events. Table 26 (below) compares the average number of words given to each life event question. In addition, the table includes the minimum, maximum and average number of words per mode of data collection.

Table 26. Mean, standard deviation and maximum number of words

	Life event 1			Life event 2			Life event 3			Overall
	Mean	Std. Dev.	Max	Mean	Std. Dev.	Max	Mean	Std. Dev.	Max	Mean
Telephone (n = 654)	4.67	4.86	42	4.57	4.57	46	5.40	4.43	44	4.9
Web (n = 889)	10.59	11.20	57	10.42	11.13	51	10.33	10.52	51	10.4
Mail (n = 364)	9.56	9.08	48	9.04	8.57	47	9.95	9.34	51	9.5

The mean number of words varied depending on mode of data collection. For the first question on life events, respondents to the telephone interview gave on average less detailed responses, with a mean of 4.7 words, compared to 10.6 and 9.6 average response length to the web and paper surveys. This is also reflected on the maximum number of words for each mode, which was smaller for the telephone mode. Results were similar for the other two questions, with average length falling for all the modes. In this case, the maximum number of words given to the telephone interview increased to 46 words. Web and paper respondents' length diminished to 51 and 47 respectively. The third question obtained an average of 5.4 words in telephone, 10.3 in web and 9.9 in paper. The difference in response length also occurred in this question, with an average and maximum length increase for telephone mail respondents, while the average number of words stays stable for web ones.

Table 27. Regression coefficients for the effect of mode on the response length

	Web (0) vs. Mail (1) (n = 1432)			Web (0) vs. Telephone (1) (n = 653)			Mail (0) vs. Telephone (1) (n = 1018)		
	Coef.	Std. Err.	Sig.	Coef.	Std. Err.	Sig.	Coef.	Std. Err.	Sig.
Life event 1	0.11	0.06	†	-0.84	0.09	***	-0.67	0.09	***
Life event 2	0.12	0.07	†	-0.81	0.09	***	-0.67	0.09	***
Life event 3	-0.016	0.04		-0.69	0.07	***	-0.64	0.09	***

Note- Results after coarsened exact matching accounting for selection differences

*** p<0.001, **p<0.01, *p<0.05, †p<0.10

From the data in table 27 (above), it is apparent that the main difference in response length related to mode was between self-completion and telephone modes. There was an increase in response length associated with respondents answering the self-completion modes compared to the telephone mode in all life event questions and with the same strength. The negative binomial regression results show that telephone respondents give fewer words than mail (coeff. = - 0.67) and web (coeff. = - 0.84) respondents, particularly in the responses to the first question in the web and telephone comparison. For the other questions, there was also an increase of words associated with answering by mail or web.

The table also showed differences in the number of words between web and mail, but there were not significant differences between the two samples.

5.4.3. Findings on response themes

The table below (28) illustrates which were the most mentioned life events by respondents to the different modes. The themes identified were common to respondents in all modes, and particularly present in responses to the mail mode.

Although the identified themes were common to the different modes, it is possible to observe that they were not always mentioned with the same frequency in the different modes. In fact, for the first question about important life events, six out of the seventeen identified themes that were reported to a different extent in all modes. In particular, differences were found in the reporting of the themes about marriage, moving to Switzerland from a different country, giving birth, illness, separation and meeting someone important for the respondent's life.

Table 28. Percent respondents by mode mentioning most frequently reported life events (%)

	Life event 1				Life event 2			Life event 3			
	Web (889)	Mail (364)	Tel. (654)		Web (889)	Mail (364)	Tel. (654)	Web (889)	Mail (364)	Tel. (654)	
Getting married	13.50	18.04	15.93	**	9.34	10.4	12.64	3.94	4.43	3.02	
Close loss	11.81	14.83	11.54		8.66	10.55	9.89	9.00	10.86	12.64	
Studies	3.49	2.14	1.65		2.81	3.06	2.75	2.25	2.91	1.37	
Accident	1.24	2.45	1.10		1.69	1.83	1.37	1.69	1.99	0.55	
Moved to CH	2.14	2.75	2.20	(**)	2.81	1.83	1.10	2.47	3.21	0.82	(*)
Giving birth	20.36	22.17	28.02	**	23.96	27.37	25.00	14.85	16.36	12.64	
Aggression	0.56	0.46	0.55		0.34	0.46	0.00	0.79	0.92	0.55	
Illness	3.82	5.35	2.47	(*)	4.27	3.06	4.12	4.84	5.20	5.22	
Separate	5.74	5.35	2.75	(*)	5.17	5.96	4.67	5.96	5.81	3.85	
Parent separation	2.59	2.14	1.10		0.56	0.46	0.27	0.45	0.15	0.27	
Meet someone	4.61	4.28	1.10	***	4.84	3.52	3.30	5.29	3.36	1.92	**
Job	4.27	4.89	2.47		5.06	6.12	4.95	9.34	7.80	9.62	
Success	1.12	1.53	2.20		1.69	1.99	0.82	1.46	2.14	1.65	
Job loss	0.11	0.15	0.00		0.34	0.00	0.27	0.45	0.15	0.27	
Moving house	2.14	0.61	1.10		0.90	0.15	0.27	1.35	1.22	0.82	
Travels	2.70	1.53	1.37		3.49	2.29	1.92	2.81	2.14	2.20	
Drug problem	0.11	0.15	0.00		0.45	0.15	0.00	0.34	0.31	0.00	
TOTAL	80.31	88.82	75.55		76.38	79.2	73.34	67.28	68.96	57.41	

*** p<0.001, **p<0.01, *p<0.05, †p<0.10

The theme of marriage was more common in the mail survey (a little more than 18% of respondents mention it) than in telephone (nearly 16%) and web (13.5%).

Having moved from another country to Switzerland was also more common in the paper survey. Giving birth and having children were more often mentioned on telephone by more than 28% of the respondents, compared to 20 and 22% of the web and mail ones. Mentioning illness is less common in telephone (2.5%) than in web (3.8%), and especially less than in mail (5.4%). The impact of separation is also frequent, but less so in telephone, mentioned by just under 3% of respondents, and more in web and mail, being indicated by over 5.3 % of respondents in both modes. Having met someone important in their lives and separation were mentioned significantly less frequently in telephone than in mail and web: only 1% of respondents discussed it compared to 4.6 and 4.3% in web and mail.

Some differences in how frequently events were mentioned are likely to be explained by sample composition differences. For example, in the previous chapter I had identified that there were more foreign respondents in the mail survey than in web and telephone, which may explain why more mail respondents mentioned having moved to Switzerland. The difference in illness as an important event in the mail survey might as well be related to sample difference such as age.

The results obtained from the logistic regression analyses controlling for sample differences are presented in table 29 (below). This table shows results for the first question on life events. There were no significant differences when comparing web than in mail. The rest of the mode comparison results show that after the multiple test adjustments there is only one event for which there were mode effects: having met someone important for respondents' lives in the mail and telephone comparison (OR = 0.16, p = 0.004).

Table 29. Relationship between life event (1) theme and mode of data collection

	Web (0) vs. Mail (1) (n = 1432)			Web (0) vs. Telephone (1) (n = 653)			Mail (0) vs. Telephone (1) (n = 1018)		
	OR	Std. Err.	Sig.	OR	Std. Err.	Sig.	OR	Std. Err.	Sig.
Marriage	0.82	0.13		1.17	0.22		0.67	0.14	†
Loss	0.72	0.17		0.60	0.23		0.74	0.29	
Studies	0.86	0.53		0.74	0.62		1.14	1.24	
Moved to CH	1.02	0.42		1.30	0.76		1.29	0.90	
Children	0.77	0.11	†	1.54	0.31	(*)	1.29	0.26	
Illness	0.64	0.20		0.33	0.19	†	0.30	0.16	(*)
Separation	0.83	0.23		0.80	0.35		0.69	0.32	
Parent separation	0.82	0.30		0.56	0.32		0.48	0.30	
Meet so	0.86	0.27		0.23	0.15	(*)	0.16	0.10	**
Profession	0. 78	0.21		0.50	0.23		0.53	0.24	
Travelling	2.14	0.90		0.78	0.48		1.39	0.94	

Note- Results after coarsened exact matching accounting for selection differences.
*** p<0.001, **p<0.01, *p<0.05, †p<0.10

The following table contains information on whether or not the respondent mentioned each event in any of the three questions. In this case, there were no differences between the different samples' results. Not controlling for sample composition differences meant mode had an impact stronger impact, on the life events signalled by the respondents: talking about traveling, having had children, getting married, parents splitting up or separating. Whether respondents mentioned it or not appeared to depend on mode of data collection. However, the effect disappears after applying coarsened exact matching and the Holm-Bonferroni correction for multiple testing.

Table 30. Relationship between life event theme and mode of data collection, overall

	Web (0) vs. Mail (1) (n = 1432)			Web (0) vs. Telephone (1) (n = 653)			Mail (0) vs. Telephone (1) (n = 1018)		
	Odds Ratio	Std. Err.	Sig.	Odds Ratio	Std. Err.	Sig.	Odds Ratio	Std. Err.	Sig.
Marriage	0.82	0.11		1.02	0.19		0.92	0.17	
Loss	0.78	0.13		1.10	0.26		1.08	0.26	
Studies	0.80	0.29		0.47	0.27		0.45	0.28	
Moved to CH	1.18	0.21		0.76	0.28		0.65	0.27	
Children	0.81	0.09	†	1.12	0.18		0.86	0.15	
Illness	0.82	0.16		0.81	0.24		0.80	0.23	
Separation	0.89	0.16		0.78	0.22		0.78	0.22	†
Parent separation	0.92	0.29		0.62	0.31		0.58	0.30	
Meet so	1.14	0.20		0.52	0.14	(*)	0.53	0.16	(*)
Profession	0.85	0.14		0.95	0.20		0.69	0.16	
Travelling	1.64	0.37	(*)	0.71	0.23		1. 20	0.42	

Note- Results after coarsened exact matching accounting for selection differences.
*** p<0.001, **p<0.01, *p<0.05, †p<0.10

5.4.4. Findings on positivity of life events

The same type of event can be associated with a positive, negative, or neutral outcome by different respondents. This means that even though mode did not have an impact on responses for most themes, their positivity or negativity could be interpreted in a different way. Table 31 (below) illustrates the level of positivity that respondents associated to their different life events, showing that positive events are more common in the telephone mode (77.6%) than in the web (66.71%) and mail (65.98%) modes. Such differences appeared between the self-completion modes and the telephone mode, situated in the neutral responses, which are a much more common choice in the self-completion mode.

Table 31. Negativity and positivity of responses by mode

	Life event 1			Life event 2			Life event 3		
	Negative	Neutral	Positive	Negative	Neutral	Positive	Negative	Neutral	Positive
Telephone (n = 654)	18.08	4.37	77.55	22.02	6.42	71.56	28.72	5.32	65.96
Web (n = 889)	23.12	10.18	66.71	21.27	9.94	68.79	23.15	13.1	63.76
Mail (n = 364)	24.4	9.62	65.98	21.79	7.5	70.71	27.89	11.57	60.53
Sig.		0.001			0.326			0.003	

The results identified in the next step of the analysis, shown in table 32, are quite revealing in several ways. Firstly, what stands out is that I found no significant effect of mode for the second and third questions on life events –after Holm-Bonferroni adjustments between mail and web, and mail and telephone. For the first question, though, mode appeared to be a predictor of the positivity outcome when comparing web and mail to telephone. In both cases, responding by telephone was associated with a higher level of positivity towards the mentioned event. The relationships between mode and positivity were medium for the web and telephone comparison (odds ratio = 1.72) and for the mail and telephone comparison (odds ratio = 1.65).

Table 32. Relationship between response positivity and mode of data collection

	Web (0) vs. Mail (1) (n = 1432)			Web (0) vs. Telephone (1) (n = 653)			Mail (0) vs. Telephone (1) (n = 1018)		
	OR	Std. Err.	Sig.	OR	Std. Err.	Sig.	OR	Std. Err.	Sig.
Life event 1	1.07	0.13		1.72	0.28	**	1.65	0.28	**
Life event 2	1.02	0.12		0.92	0.16		1.05	0.18	
Life event 3	1.30	0.16	(*)	1.14	0.19		1.44	0.25	(*)

Note- Results after coarsened exact matching accounting for selection differences.
*** p<0.001, **p<0.01, *p<0.05, †p<0.10

Not controlling for sample differences meant that mode had an impact on the reported positivity of the event in all of the three questions, also when comparing web to mail responses.

Because different respondents may interpret the same type of event as positive or negative, I tested the given positivity or negativity depending on mode for a series of identified life events. In the table below (33) it is possible to find information about whether respondents evaluated events in the same way independently of mode.

For responses to the live event 1, the results show significant differences for the event having had children. From this data, we can see that there is a difference in the way in which such events are seen (p = 0.002). While most telephone respondents evaluate this event as positive, and only 2 % of respondents report it being negative or neutral, the extent of web and mail respondents evaluating it as neutral and negative is significantly bigger (approximately 20 percent).

Table 33. Positivity of events, by event and mode of data collection

Events	Evaluation	Life event 1 Self-completion (n = 654)	Telephone (n = 364)	Sig.
Getting married (n = 284)	Negative	6.34	0.00	
	Neutral	5.72	1.75	
	Positive	87.95	98.25	
Close loss (n = 120)	Negative	56.55	66.67	
	Neutral	36.07	16.67	
	Positive	7.37	16.67	
Giving birth (n = 428)	Negative	8.70	1.04	**
	Neutral	13.92	1.04	
	Positive	77.38	97.92	
Separate (n = 78)	Negative	64.21	50.00	
	Neutral	27.99	25.00	
	Positive	7.80	25.00	
Illness (n = 70)	Negative	38.27	50.00	
	Neutral	51.77	16.67	
	Positive	9.96	33.33	

Note- Results from Pearson Chi squared test, after coarsened exact matching accounting for selection differences.
*** p<0.001, **p<0.01, *p<0.05, †p<0.10

The evaluation of illness related events was significantly different between the modes for respondents to the third life event question (p = 0.0072): 44 % of the telephone respondents evaluated it as positive, and 56% as negative, while none of them saw it as neutral. On the other hand, 47% of the respondents that filled the self-completion questionnaires reported it being negative, 40% as neutral, and 13% as positive.

5.5. Discussion and conclusion

The goal of this chapter was to investigate the extent of mode effects in responses to open-ended questions on important life events, particularly measurement-related effects. Open-ended questions are popular in social science research as a way of overcoming the shortcomings of the – often more popular – closed-ended questions. They allow survey participants to respond in their own words, avoiding the influence that response categories may have, and providing deeper insights that are often useful for researchers. However, they also require more effort from the respondent and from the researcher, which can lead to lower quality compared to closed-ended questions. In addition, previous research – although not conclusive – has found some indications that mode effects can affect them, further complicating their analysis, and increasingly so when the topic of study is regarded as sensitive by the respondents. For this reason, this study set out to assess the effects of mode in open-ended questions about important life events by comparing responses to three different modes of data collection: web, mail and telephone.

In order to address my research questions, I reported results from a mode comparison between telephone, mail, and web-based surveys and examined different indicators of data quality for open-ended questions across modes. Replicating prior research, I compared the level of item nonresponse and the length of responses between web, mail and telephone. In addition, I completed the analysis by comparing the level positivity of responses and the themes indicated by respondents across modes.

The first question in this study sought to determine whether mode impacts the level of item nonresponse. Previous research offered a variety of results, but the overall tendency was to find a higher level of item nonresponse in mail surveys compared to web. Results from the current chapter show significant differences between the modes, even after adjusting for selection effects, when comparing results from the self-completion and telephone samples. However, this appears to depend on the number of open-ended questions to complete: while item nonresponse is lower for telephone respondents when looking at the first open-ended question, the item-nonresponse rate for telephone goes up to 21% in the third question about life events. However, this increase might be due to the way in which the interviewers reacted to obtain responses to this question.

It was also hypothesised that mode would have an impact on response length. The findings from this thesis did not show mode effects in the length of responses to the web and mail questionnaires, and although answers to the web questions were found to be slightly longer than to the paper mode, there is no significant difference. The difference between web and mail compared to telephone indicated the average number of words for the life event questions was close to 10 for both self-completion modes while the telephone average is between 4 and 5 for the three open-ended questions. However, such differences could be related to the way in which the interviewer recorded the information and not due to response differences.

The third question in this chapter was about the response content. In order to know the extent of the impact of mode on response content, I tested mode differences in the responses' level of positivity and in the reporting of different types of life event. Due to the effect the presence of an interviewer can have – by increasing the likelihood of respondents giving socially desirable responses – I expected telephone respondents to report positive events to a higher extent than self-completion ones, whilst avoiding referring to events that had a negative impact in their lives. First, I checked whether mode had an impact on the type of event, and afterwards I examined whether the levels of the events' positivity varied or were the same independently of mode. For most of the examined life events, I found no association with mode of data collection. One interesting finding is that these differences were not only found between the self-completion modes and telephone, but also between mail and web (in the case of the marriage event). This may indicate that differences are not so much due to social desirability, for example, but potentially to selection differences that I was not able to control for.

The hypothesis of the impact of mode on the positivity of responses was supported by the results for only the first question. As expected, being interviewed by telephone had a small to medium positive impact on the reporting of events that had a positive impact on the life of respondents. However, for the next two questions, the observed differences between telephone and self-completion in this study were not significant.

In addition, I took into account whether certain events were perceived as positive, neutral or negative, finding significant differences in the interpretation of the modes and their positivity in two cases: giving birth and illness, which were reported as positive more often in the telephone interview than in the self-completed questionnaires.

Finally, a number of limitations need to be considered. With regard to the research topic of the open-ended questions – major life events – cross-sectional data presents a disadvantage over longitudinal data. The recentness of an event is an important aspect of life-course studies, because the effect of many events tends to be important only temporarily, whether they are good or bad. Not knowing how recent events are for respondents can hinder conclusions, especially when taking into account how positive or negative an effect was in respondents' lives and how this perception may have changed over time. In addition, measuring the detail and quality of the information provided by using response length is not necessarily the best way of doing it. Especially when responses are going to be recoded into wider categories that miss the higher level of detail. This can be the case for life events: while respondents to the self-completion questionnaires may give unnecessary details for the researchers such as the dates of birth of their children, the date of their wedding anniversary or describe the details of how their partner was unfaithful to them. In this case, response length may not be an important factor to take into account when evaluating the quality of responses and it would be necessary to develop a better way of evaluating usefulness. For this reason, depending on the researchers' objective and the type of analysis that is going to be implemented, the choice of one mode over another could make a difference or not. If the objective is just to get information about an event that is related to marriage, birth, or loss of someone close it only makes a small difference if self-completion modes or interviewer-based modes are used. Lastly, it is difficult to know how much of the impact of mode in the positivity-negativity level of responses is related to being measured through a closed-ended question: differences can be due to the question formatting.

To conclude, it is possible to say that mode of data collection affects responses about life events, both in terms or item nonresponse, response length, and evaluation of the positivity and negativity of such events: telephone responses display higher levels of positivity. This finding supports the idea that respondents want to present themselves in the most positive way, particularly when responses are given orally to an interviewer instead of being self-completed. In spite of this, the theoretical implications are unclear, as this effect is only observed for the first open-ended question and disappears afterwards. In addition, results also suggest that there is no great difference in the content of responses with respect to reported life events. The results from the chapter suggest that not controlling for different sample compositions in the appropriate way could mean obtaining different results, not only when comparing self-completion modes, but also when testing differences between web and paper.

More research is required to understand the role that respondent characteristics have in mode effects for open-ended questions. Previous research noted that age, sex and education can have a mediator effect in the responses given in open-ended questions to different modes of data collection (Stern, Dillman, & Smyth, 2007). For this reason, in the chapter that follows I will move on to consider how respondents' characteristics can interact with mode of data collection and affect results in studies on life events.

CHAPTER 6. DOES MODE AFFECT ALL RESPONDENTS EQUALLY?

6.1. Introduction

The mode of data collection has an impact on subjective wellbeing measures. In the previous chapter, I found that not only do different modes attract different types of respondents, but also that similar people answer self-completion questionnaires differently compared to telephone ones. In particular, comparisons of web and telephone, and mail and telephone showed that mode affected 21 of the 27 examined measures in a significant way. Of those, 11 were related to individual levels of subjective wellbeing: responses about social trust, life satisfaction, happiness, taking time to do the things they enjoy, self-rated health, optimism, positivity, freedom, feeling of accomplishment, feeling that things are going well, overcoming differences quality of sleep and loneliness, 19 from the comparison of web and telephone comparison, and 16 from the mail and telephone one.

When comparing respondents that were as similar as possible – in socio-demographic terms – by using coarsened exact matching, the number of affected measures was 12 in the mail and telephone comparison, and 15 introducing the propensity score. In the case of the web and telephone comparison, there are 15 measures affected by mode if using coarsened exact matching. When the propensity score method of controlling for mode effects on selection is used, the number of observed differences in measurement between the modes is 12.

As a summary, after coarsened exact matching, it is possible to say that mode had an impact on responses in a total of 18 measures.

In addition to studying the impact of mode in the general population, it is also important to investigate whether everyone is affected in the same way. Especially because the increasing interest in the population's wellbeing level has also been related to the research objective of comparing different population subgroups. Predictors of wellbeing are often related to socio-demographic characteristics that define subgroups such as being an immigrant vs. a native of a country, having a higher income vs. living a low income, a higher education vs. little formal education, being married vs. being single, divorced or widowed, or being old vs. young. In vulnerability research, examining the situation of minorities and other sub-groups at risk is essential, and it is often those vulnerable people who are more difficult to reach and less willing to provide information about their lives (Rothenbühler & Voorpostel, 2016). Additionally, the process of responding to a questionnaire varies according to personal characteristics (for example, age or education level). This poses a problem if the already mentioned sub-groups differ from the rest of the population on the subject of interest, for example in their wellbeing level: obtaining biased information on them could lead to the wrong conclusions (Conti & Pudney, 2011).

When mixed-mode designs are involved in comparing sub-populations, arriving at accurate conclusions can be challenging (Hox, de Leeuw & Klausch, 2017). At the moment, this is not stopping researchers carrying out their surveys and investigation using different modes of data collection, and for this reason it is important to address the question of whether mode effects interact with respondents' characteristics, and whether this could affect comparisons across subgroups.

To address the questions raised above, in this chapter, I investigate whether mode effects affect all parts of the population in the same way. Until now, most studies of mode effects on measurement quality have tended to focus on broad-brush comparisons across respondents answering in different modes. In contrast, relatively few studies have addressed the question of whether all respondents are equally affected by mode, or whether the effect of mode varies by population subgroup.

The first section of this chapter introduces existing work that has been done on the topic, even if scarce. I discuss the results from studies that have focused on explaining how respondents' characteristics influence the responding process and how this is related to survey design, and more specifically, to mode of data collection and formulating the research questions. In the following section, I explain the analytical decisions and procedures related to the comparisons of responses on measures of SWB across population sub-groups and modes, testing for differential response distributions and means and differences in the indicators of quality in open-ended questions. Finally, I present the findings discussing how mode effects that affect different types of respondent differently impact current research that involve group-comparison based on mixed-mode data. To conclude, this study presents new evidence about the role that respondent characteristics play in the conclusions obtained from different modes on vulnerability and subjective wellbeing studies, for which group comparisons are essential.

6.2. Literature review

Survey researchers have investigated many causes that can trigger response bias. Two of particular importance are the mode of data collection and the disposition that the respondent has to answer the questions, which may vary across socio-demographic characteristics (Kieruj & Moors, 2013). Although both have been documented, few studies have attempted to investigate the relationship between respondents' characteristics and mode effects (Revilla, 2012). Most studies that examine mode effects in survey estimates assume that the impact of mode in all respondents is the same (Saris & Gallhofer, 2007), but the different question characteristics can have a differential effect in how different respondents answer, particularly for sensitive or complex questions (Revilla, 2012). Previous research warns about the possibility that such an interaction is likely to have negative consequences if the quality of responses is higher for respondents, but it has rarely been studied systematically.

The interaction between survey error and respondents' characteristics is particularly important for life quality research. Comparisons across socio-demographic groups are common, and there is a perception that the datasets being used are not able to fully measure those who are the most vulnerable (Rothenbühler & Voorpostel, 2016).

6.2.1. The link between the response process and the mode of data collection

Respondents answering survey questions go through different stages that involve understanding what is being asked, finding the required information from their memory, deciding which answer to give, and finally to formulate it (Jobe, 2003). Survey design interacts with this response process, for example, by influencing the way in which the concept being measured is understood, and the response error (Robling et al., 2010).

Mode of data collection plays a particularly important part in the response process: the presence or absence of an interviewer, together with the other survey characteristics, such as the sensitivity of a question, can modify someone's response behaviour, as I presented in Chapter 2. For this reason, self-completed questionnaires are often better for motivating sincere and accurate responses to sensitive questions (e.g., health) than interviewer-based modes (Bowling, 2005).

In addition to social desirability, it is possible to identify a series of response patterns that researchers should be aware of when choosing a mode of data collection and analysing the data. For example, the tendency to choose extreme response options on telephone surveys and continuously agreeing regardless of the question content, or the tendency to give milder responses or even not having an opinion when the questionnaire is self-completed. These response styles vary according to respondent characteristics (Pickery & Loosveldt, 1998).

Overall, studies suggest that differences in response styles due to respondents' characteristics can affect substantive conclusions drawn from surveys, and that if these differential response effects are not taken into account they may lead to biased survey estimates (Tutz & Berger, 2016). What, however, is the interaction between response styles, mode of data collection and response characteristics?

6.2.2. Response styles and respondents' characteristics

There are a few important reasons why people may adopt different response strategies. Previous research has focused on two main aspects: personality and socio-demographic characteristics. Even though personality has been found to be an important factor in explaining response style (Kieruj & Moors, 2013), it is difficult to disentangle psychological traits from response styles in psychological tests. For this reason, and due to its availability, most studies look at certain socio-demographic characteristics to explain observed differences. Moreover, variables such as sex or race have been found to work well as a proxy for sub-group cultural differences (Holbrook, Cho & Johnson, 2006).

There is evidence about the relationship between response patterns and socio-demographic characteristics such as age, sex, education, income and ethnic origin. Up to the present, the extent to which such respondent characteristics' moderate responses remains unclear, but in the following lines I offer an overview of the impact each respondent characteristic may have. A recent study implemented by Kieruj and Moors' (2013) illustrates how low education and older respondents may be less likely to ask for clarification or to manifest problems during the survey completion process in cognitive survey test. For this reason, such tests fail to identify the full extent of the relationship between respondent characteristics and response styles.

1) Sex

Respondent's sex is also correlated with responding styles. In fact, women tend to report higher levels of happiness and life satisfaction then men (Pudney, 2010; Wood, Rhodes & Whelan, 1989). Gove and his colleagues (Gove, Hughes & Style, 1983) assessed the extent of response bias associated with sex by estimating men and women's tendency to agree and disagree and their need for social approval. They found differences between men and women, and controlling for biasing factors did not reduce the observed sex differences. However, the findings about the differential effect of sex were heterogeneous. Some studies show that women tend to acquiesce more than men (Weiters et al., 2010), while others found no effect (Marin, Gamba & Marin, 1992), or even that women are less likely to agree with the responses, such as the study by Ross and Mirowsky (1984, p.193) which found that women appear to be less prone to acquiescence than men. However, Jia He and her colleagues (2015), in a cross-country comparison, found that women have the tendency to exaggerate their positive traits or behaviors, while men were found to be more likely to underreport negative aspects.

2) Age

Studies that have looked at the effect of age in responding styles have found that acquiescence and extreme responding increase with age (Kieruj & Moors, 2013; Meisenberg & Williams, 2008; Schneider, 2016). This could be related to the fact that cognitive resources and memory decrease with age, although these studies have also shown that the lack of cognitive skills does not suppose a bigger problem for older people than for youngsters. Older respondents might have more trouble, however, when they are requested to retrieve information from the distant past. There is, therefore, mixed evidence on what the repercussions would be: age has been found to interact with the response process when the questions are complex and presented quickly by the interviewer (Holbrook et al., 2006), which does not happen in self-administered survey questions.

3) Education

Similarly to age, education is also related to the cognitive resources of the respondent. The lower the cognitive ability, the strongest the response effects

(Holbrook et al., 2006; Kieruj & Moors, 2013; Krosnick, Narayan & Smith, 1996). Consistent with this, studies have found that a lower level of education is often associated with a higher influence of the questionnaire design on responses, especially extreme responding and acquiescence in rating scales. This is related to satisficing, as respondents with lower cognitive resources might find it harder to answer the questionnaire than higher educated one, either as a result of a lack of motivation or due to ability. A higher level of education has also been linked to lower levels of social-desirability bias, potentially related to a 'liberalizing' effect on respondents' tolerance level (Heerwig & McCabe, 2009). In addition, Alwin and Krosnick (1991) argued that the level of formal education is relevant because accessing the highest levels of education requires higher cognitive skills, although at the same time they clarify that such cognitive skills can be developed. Students are also accustomed to tests and exams, among them multiple-choice ones that can facilitate and normalize questionnaire completion.

4) Language and nationality

In the United States, race and ethnicity are widely used as a proxy to the way in which respondents answer related to the mother-language and the way of interpreting and understanding life (Kieruj & Moors, 2013). Non-white respondents have more difficulties comprehending survey questions that are often tailored for a white-dominant cultural group. Cultural background is correlated with the way of responding (Harkness, Vijver & Mohler, 2003) and belonging to a minority group can lead to differential response processes.

5) Income

Socio-economic background is an important factor when studying survey errors. There are mixed findings on how it related to survey response, but so far, having a low income has been associated with a higher level of extreme responding and dis-acquiescence, which is the tendency to disagree rather than agree with the question statement (Białowolski, 2016). However, not all studies find a strong link between income and response style. For example, the correlation between income and extreme scale responding and acquiescence was not found to be significant in the Netherlands (Kieruj & Moors, 2013).

6) The role of respondent's motivation

Whether someone is motivated to answer a questionnaire, or reluctant, is an important characteristic that can influence the response process. The idea behind this interaction is that those respondents that are less likely to respond are also the ones that give information of questionable quality (Olson & Kennedy, 2006). Some experiments have been implemented to explore the interaction, but often only on specific sub-populations such as students and with no conclusive results (Sakshaug, Yan & Tourangeau, 2010). For example, Olson and Kennedy ran an experiment in which they predicted that less successful alumni – for which register data were available – would be less likely to participate and, at the same time, more likely to give lower quality data. Interestingly, results on the relation between motivation and data quality was not consistent across all items and for some cases the relationships were not significant. The authors point out that the level of effort needed to get responses appears to have an important role moderating the relationship between the errors.

The interaction between response motivation and measurement error is particularly important for vulnerability studies. Rothenbühler & Voorpostel (2016) explain how certain vulnerable people (Spini, Bernardi & Oris, 2017) may not have the resources needed to respond to surveys in a pleasurable way. As mentioned, vulnerable people may be among the most reluctant respondents, whose situation is related to some of the characteristics described above: different ethnic origin, lower income, etc. The majority, on the other hand, people who are better integrated socially, are often more motivated to participate, collaborate and may be more interested by the topics covered in surveys. In their work, Rothenbühler & Voorpostel focus on attrition in longitudinal surveys and on how vulnerable respondents drop-out more often than the rest of interviewees, their analysis is based on the whole population but they warn that sub-group comparisons may point towards important differences.

In their study, some of the characteristics that can be identified in reluctant respondents and non-respondents in Switzerland are age (higher age is identified with an increased likelihood of responding, up to 55 years old, when the tendency changes), education (the higher the level of education, the more likely a person is to respond), nationality, working status, income and health condition (Rothenbühler & Voorpostel, 2016).

6.2.3. The interrelation between mode effects and respondents characteristics

In a large longitudinal study, Pudney (2010) and Holford and Pudney (2015) investigated the incidence of respondents' characteristics on mode effects. Using data from the British Household Panel they investigate the mechanisms by which men and women's responses were systematically different depending on survey design and demonstrated how different survey designs can affect population sub-groups differently. To determine the extent to which response characteristics interact with survey design effects, they compared computer assisted interviewing (CASI), face-to-face (F2F) telephone (CATI) and paper self-completion (PSC). The measures that they examined are seven category scales on health, satisfaction with income, leisure time and general life satisfaction. Pudney and Holford report differential response effects depending on whether respondents are men or women, for example on the importance they give to aspects of their lives such as work-life balance. However, differences were not consistent across measures – apart from extreme responding, which was found for all satisfaction measures – suggesting that being interviewed by telephone is more likely to bias the distribution of respondents' answers.

Their findings demonstrated that the effects of mode of data collection are significantly different for different types of respondent when looking at response distributions. In this case, the effect of mode on women's responses is stronger than for men when one of the interview modes is telephone. An important conclusion is that these differential mode effects can easily influence substantive research into wellbeing (Conti & Pudney, 2011).

Additional survey characteristics that may interact with respondent and mode characteristics include the type of question being asked. A study carried out by (Kieruj & Moors, 2013) compared different scales but found no evidence of differential response effects according to the type of response scale, while Pudney and Holford (2015) only found weak differences.

6.2.4. Research questions

Previous research has found that there are differences in how people with shared characteristics respond to surveys, in terms of their tendency to respond in particular ways to certain types of question. Different response tendencies have been observed for both different types of respondents, and for different modes of data collection. In this chapter, I address the following research questions:

R.Q.1: Does mode affect all respondents in the same way when answering questions on subjective wellbeing?

Based on previous findings on differences in the way different types of people respond to sensitive questions, the fact that subjective wellbeing is a sensitive topic, and that responses can be affected by social desirability bias – particularly in interviewer-based modes – I expect that there will be differences in the size of mode effects for different modes of data collection. In particular, my hypothesis is that there will be differences between the oldest respondents and the rest, with responses from the oldest respondents being more affected by mode. I also expect there to be differences between men and women, particularly when looking at the measures of happiness and life satisfaction, for which women were found to be more affected by the survey design's characteristics than their counterparts. People with a lower level of education are also expected to be more affected by survey design, as they are less familiarised with tests and rating scales than higher educated respondents and have been found to have less cognitive resources: extreme responding, acquiescence and social desirability bias are more frequently found in the responses of people with a low education. Previous literature has found that being a native speaker of the language in which the survey is conducted can be related to differences in the interpretation and understanding of the questions. For this reason I expect that there may be some differences between native and non-native speakers.

Finally, being a motivated or a reluctant respondent also influences the response process. In particular, reluctant respondents often give worse quality answers, which may indicate that such responses may be more mode sensitive than motivated respondents' reports.

R.Q.2: If there is an interaction between respondents' characteristics and mode of data collection, is it consistent across the different question formats?

To expand on the research from the first study in which I investigate whether mode effects are due to the number of response categories, I also look at whether there are disparities in how different groups of people respond to the various question types in different modes. For example, respondents that have spent many years in education were found to be more accustomed to scales and multiple-choice format exams than respondents with no formal education (Alwin & Krosnick, 1991). It is therefore possible that the responses of the last group are affected by mode when there is no interviewer to potentially clarify how to respond.

6.3. Methods

The aim of this chapter is to look at the possible interaction between respondents' characteristics and mode effects. In order to do this, I selected a series of characteristics that identify subgroups of the population that had been previously found to respond in different ways to survey questions. To implement the study, I implemented a series of ordered logistic regressions that show the effect of mode in the response outcome to questions about subjective wellbeing depending on respondents' characteristics. To date, various methods have been developed and introduced to measure mode effects (see chapter 3), however, there are few precedents on how to measure their relation to respondents' characteristics. There are different ways to study the relationship between variables that allow us to both describe differences between groups and to test whether they are statistically significant or not. In this section, I describe the statistical techniques used to answer the research questions.

6.3.1. Data

The study's analysis builds on the previous chapters. I look at respondents that answered through the mode they were assigned to, and with listed telephone numbers.

6.3.2. Variables

The subjective wellbeing measures analysed in this chapter are related to the different aspects of subjective wellbeing analysed in the previous two empirical chapters. To complement the analysis of chapter 5 on open-ended questions, I also look at the variables related to life events. In particular, I examine the effect of the interaction of respondents' characteristics and mode of data collection on the differences in response length, item-nonresponse, and positivity of reported life events.

The second research question focuses on testing the influence of question format in mode effects, and for this reason the variables are classified by number of response categories. Eleven category measures are on social trust, happiness, life satisfaction, and being able to take time to do the things they enjoy. Five category measures are self-rated health, optimism, positivity, freedom, feeling of accomplishment, feeling able to take control of their lives, handling of problems, doing well, overcome difficulties, and social activities. Finally, the four-category questions are on depression, restless sleep, loneliness and anxiety.

Finally, mode of data collection is the key element of the analysis. I examine two types of mode of data collection, self-completion and interviewer based ways of gathering data. Self-completion involves paper and web on the one hand, as no measurement effects were found between them in the previous study, and telephone on the other.

There were some subjective wellbeing items for which the parallel lines assumption (see chapter 4 for an explanation) was violated. This was the case for the following items: Happiness, life satisfaction, having someone to discuss, feeling able to take control, optimism, positivity, freedom, feeling able to handle problems, feeling that things are going well, and feeling able to overcome differences.

6.3.3. Sub-groups analysed

The population sub-groups based on respondent characteristics were chosen based on the literature review, if there was enough information in the database to identify them. This way, I selected different types of respondent based on the socio-demographic characteristics sex (man (1) or woman (0)), age (younger than 65 (0), or 65 and older (1)), whether the respondent lives in an urban or rural area, and nationality (identifying cultural background and level familiarity with the French language: for this reason this is a dummy variable in which one group is composed by Swiss and French respondents (0), and another group by the rest of respondents (1), see figure below for more detail). In addition, educational level can also interfere with the answering process (using a dummy variable in which there is a group of respondents with compulsory education or less (1), and another group with a higher level of education (0)). Finally, I distinguished between differential motivation to respond by dividing the sample between those that respond during earlier stages of the survey fieldwork (0) and those that needed more time and effort to answer (1). The objective of this measure was to create one variable with two categories that measures willingness to respond. As people respond differently to different modes, I created this variable separately for web, mail and telephone.

Table 34. Language spoken depending on nationality (%)

Nationality	Language French	Language Other	Total
CH/FR	90.08	21.05	86.71
Other	9.92	78.95	13.29

Motivated telephone sample respondents are those that answered after receiving up to 4 calls, which corresponds to the time in which participants with higher response propensities stop participating. Reluctant respondents are defined as those that answered after the 5th call. The survey participants that were allocated to the web and paper samples and who completed the survey before December 2012 are considered to be motivated, which corresponds to the first phase of the data collection (see chapter 2 and Roberts, Joye, & Ernst Stähli, 2016).

Table 35. Types of respondent by mode of data collection

	WEB % (n = 457)	MAIL % (n = 351)	TELEPHONE% (n = 364)
Reluctant	45.11	59.17	42.03
Motivated	54.89	40.83	57.97
Men	51.63	46.02	46.98
Women	48.37	53.98	53.02
Urban	73.79	68.81	69.51
Rural	26.21	31.19	30.49
<65	88.53	80.73	77.47
65+	11.47	19.27	22.53
30+	75.48	79.36	78.85
<30	24.52	20.64	21.15
Higher ed.	85.84	79.68	80.00
Low ed.	14.16	20.32	20.00
Swiss/border	86.61	81.65	92.86
Foreign	13.39	18.35	7.14

6.3.4. Analytical approach

In chapter 4, I showed that being interviewed by telephone increases the likelihood of reporting a high level of subjective wellbeing. In this chapter, the first research question focuses on whether there are significant differences with the response distributions of a series of SWB measures for different population sub-groups and across modes of data collection. In previous literature, there is a strong focus on the type of response style that is being analysed. In this study I look at the impact of question format and question content.

The measures of subjective wellbeing analysed here are ordinal, and for this reason looking at the response distributions across the ordered categories is particularly important. The analysis is a comparison of response distributions between the self-completion and the telephone mode, as we are particularly interested in studying the differences between an interviewer based mode and an self-completed questionnaire.

Following Pudney's approach (2010), I start by presenting the differences in the overall response distribution differences for the chosen SWB measures across the two types of mode and socio-demographic groups. For the response distributions, I assess statistical significance through the non-parametric Kruskal-Wallis test, which gives information on the equality of response distributions for telephone compared to self-completion modes and is not based on the mean value (Pudney, 2010). This approach works well for group comparisons because it can test for differences between groups of different sample size, even if they are small (there should be at least 5 respondents in each group for it to work). Like in the previous chapters, the significance levels presented are corrected using the Holm-Bonferroni method. Results from this analytical step show the measures for which there are significant differences on their response distribution between different types of people, but in order to better understand the differences.

The main objective of this chapter is to identify differential mode effects depending on the respondents' characteristics. In order to do this, I implemented a series of ordered logistic regressions and multinomial logistic regressions (for those measures for which the parallel lines assumption was violated) in which the mode of data collection (self-completion or telephone) is the predictor of the subjective wellbeing measure. In addition, I included the interaction terms between mode of data collection and respondents' characteristics, indicating whether there is a combined effect that explains response differences. By including the interaction effects of mode with sex, age, education, nationality, and response motivation, the main effect of each characteristic is also reported. For this reason, in order to control for differences mode selection effects, I used coarsened exact matching approach, using only the variable "use of the Internet" to balance the sample compositions of the modes.

In order to interpret the results from the analysis of the effect of the interaction of mode and respondents' characteristics, I do not present the results from the multinomial logistic regression or the ordered logistic regression themselves. Instead I use the Stata command "margins" that estimates the marginal effect of the mode effect (StataCorp 2013) and the differences in probabilities for each pair of respondents' characteristics with respect to mode of data collection, with self-completion being the reference category. To ease the interpretation of these results, I also present the results from the contrast of margins, indicating the difference in the marginal effect and the significance level, for each comparison of respondents' characteristics (for example, men compared to women) and response category, and the significance level for the whole model (independently of the category examined).

Because the coefficients for each interaction term would make the understanding of the results difficult, due to the name of items and interactions, in the results section I only show those cases in which there was a statistically significant interaction effect for the measures of personal and social subjective wellbeing (before and after the Holm-Bonferroni adjustment). Whether significant or not, the last part of the chapter consists of the analysis of the same interaction effects with the open-ended questions about life events examined in the previous chapter.

In order to show some examples of whether the effect of mode was different depending on population subgroup, I present the odds ratios of the effect of mode for each type of respondent, for the composite scores. This can facilitate the understanding of the strength of the influence that the variable has in explaining the overall variation in the odds of reporting high subjective wellbeing, and provide information on whether this is different for different population groups. When examining the odds ratios, they should be interpreted following the indication that odds ratios higher than 1 indicate a positive association; from 1.5 means there is a small mode effect; and below 1 mean that there is a negative association between the dependent variable and the predictor.

As was the case in the first part of the analysis, the presentation of the odds ratios involves estimating the parameters of ordered logistic regression equations and multinomial logistic regressions that are implemented for each population sub-group separately. For each regression model the wellbeing measure is the dependent variable, while mode of data collection is the predictor. The mode effect sizes can be calculated and compared between groups. To illustrate the analysis, it is useful to think about an example with one of the outcome measures: how often did the respondent feel lonely last week, with the response options never or almost never, from time to time, most of the time, and always or almost always. The regression results will show what relationship exists with mode of data collection, while adjusting for selection differences.

6.4. Results

To answer the first research question, I examined whether the differences in the different SWB measures' distributions are equal across types of respondent. I present results for 22 measures related to different aspects of personal subjective wellbeing. The first part for the section is based on a series of Kruskal-Wallis tests, which were implemented for each personal subjective wellbeing measure. Having tested the null hypothesis that the distributions were identical, results show that mode of data collection impacts in a statistically significant way, at the 5% level, responses on personal subjective wellbeing measures (see table 36). There is also some evidence that mode does not impact everybody in the same way. In fact, there are distributional differences across all pairs of respondents' characteristics comparisons. The differences are not consistent but vary greatly depending on which variable is being inspected.

Before adjusting with the Bonferroni method, there were differences on 18 variables between men and women, and afterwards there were six; for example, for the distribution of responses about whether respondents' feel they have of being doing well in their lives. The differences between responses in self-completion and telephone modes are significantly different for men (p=0.000), but not for women (p=0.108). Results are different for the distribution of responses on social trust, for which responses are mode dependent for women (p=0.000) but not for men (p=0.114).

Table 36. Kruskal-Wallis test for equality comparing response distributions from telephone (364) and self-completion (n = 808) modes

SWB measures	Sex		Age		Education		Nationality		Response	
11 categories										
Social trust	**Men**	(*)	**<65**	***	**Higher ed.**	***	**Swiss/border**	***	**Motivated**	***
	Women	***	**65+**		**Low ed.**	(*)	**Foreign**		**Reluctant**	
Life satisfaction	**Men**	(**)	**<65**	***	**Higher ed.**	***	**Swiss/border**	***	Motivated	***
	Women	***	**65+**		**Low ed.**	(*)	**Foreign**	†	Reluctant	(**)
Happiness	**Men**	***	**<65**	***	**Higher ed.**	***	**Swiss/border**	***	Motivated	***
	Women	(**)	**65+**	(*)	**Low ed.**		**Foreign**		Reluctant	(**)
Take time to enjoy	**Men**	**(**)**	<65	†	Higher ed.	(*)	Swiss/border	(*)	Motivated	(*)
	Women		65+		Low ed.		Foreign		Reluctant	(*)
5 categories										
Health	Men		<65	**(*)**	Higher ed.		Swiss/border		**Motivated**	
	Women		65+		Low ed.	†	Foreign		**Reluctant**	(*)
Optimism	Men	***	**<65**	***	**Higher ed.**	***	**Swiss/border**	***	Motivated	***
	Women	***	**65+**		**Low ed.**	(**)	**Foreign**	(*)	Reluctant	(**)
Positivity	**Men**	***	**<65**	***	**Higher ed.**	***	**Swiss/border**	***	Motivated	***
	Women	**)	**65+**		**Low ed.**		**Foreign**	(**)	Reluctant	(*)
Freedom	Men	***	<65	***	Higher ed.	***	**Swiss/border**	***	Motivated	***
	Women	***	65+	***	Low ed.	***	**Foreign**	†	Reluctant	(**)
Accomplishment	Men	***	<65	***	**Higher ed.**	***	**Swiss/border**	***	Motivated	***
	Women	***	65+	(*)	**Low ed.**	†	**Foreign**		Reluctant	(**)
Take control	Men	(**)	<65	**)	Higher ed.	**(**)**	**Swiss/border**	***	**Motivated**	***
	Women	(*)	65+		Low ed.		**Foreign**		**Reluctant**	
Handle problems	Men		<65		Higher ed.		Swiss/border		Motivated	
	Women		65+		Low ed.		Foreign	(*)	Reluctant	
Doing well	**Men**	***	**<65**	***	**Higher ed.**	***	**Swiss/border**	***	Motivated	(**)
	Women		**65+**		**Low ed.**		**Foreign**		Reluctant	(**)
Overcome difficulties	Men	(**)	<65	(*)	Higher ed.	**(**)**	**Swiss/border**	***	Motivated	(**)
	Women	(*)	65+	(*)	Low ed.	†	**Foreign**		Reluctant	(*)
Social activities	Men	(*)	<65	(*)	Higher ed.	(*)	**Swiss/border**	(**)	Motivated	(*)
	Women	†	65+	(**)	Low ed.	(*)	**Foreign**		Reluctant	(*)
4 categories										
Depression	Men		<65		Higher ed.		Swiss/border		**Motivated**	
	Women		65+		Low ed.		Foreign		**Reluctant**	(*)
Restless sleep	Men		<65		Higher ed.		Swiss/border		Motivated	
	Women		65+		Low ed.		Foreign		Reluctant	
Loneliness	Men		**<65**	***	**Higher ed.**	(*)	Swiss/border	(**)	Motivated	(**)
	Women	(**)	**65+**		**Low ed.**	†	Foreign		Reluctant	(*)
Anxiety	Men	†	<65		Higher ed.		Swiss/border	†	Motivated	
	Women		65+		Low ed.		Foreign		Reluctant	

*** p<0.001, **p<0.01, *p<0.05, †p<0.10. Respondents' characteristics in bold show items where results were different for each type of respondent.

What stands out in the table are the distribution differences between Swiss people and French speakers and the rest of respondents. There were differences for 11 out of the 18 tested measures, in which responses given by respondents from other countries appear to differ more depending on mode than for Swiss and French-speaking respondents. The table also shows many differences in terms of age and education (in the case of 8 items), and less for the comparison between reluctant and motivated respondents (in 7 of the measures). Responses from higher educated people appear to be more affected by mode than responses from lower educated respondents. As are responses from younger respondents compared to older ones; and from motivated respondents compared to reluctant ones.

Table 37, below, displays the sizes of odds ratios showing the effect that the mode of data collection has in responses to SWB questions for different groups of the population adjusting for differences in selection bias. Looking at the table, it is apparent that the effect of mode in responses varies depending on the population subgroups.

There were differences in the effect of mode between men and women when looking at the combined measure for the 11-category scale items, the effect of responding by telephone has a positive effect in choosing greater categories of wellbeing, while the effect is different for women. On the other hand, there were differences for almost every summed score between older and younger respondents, these last ones responding more positively in telephone than in self-completion compared to older ones. Answers from reluctant respondents also appear to be more affected by mode than motivated respondents' reports, particularly for the 4-categories variable.

Table 37. Mode effect (OR) by subgroup and type of question

	Number of categories											
		11			7			5			4	
Subgroup	OR	Std. Err.	Sig	OR	Std. Err.	Sig	OR	Std. Err.	Sig	OR	Std. Err.	Sig
Female (n = 616)	0.53	0.18	***	1.27	0.09	**	1.18	0.06	**	1.09	0.05	*
Male (n = 556)	1.42	0.18	**	1.25	0.1	**	1.38	0.17	*	1.20	0.08	**
Older than 65 (n = 237)	1.37	0.25		1.21	0.15		1.18	0.09	(*)	1.09	0.08	
Younger than 65 (n = 935)	1.48	0.14	***	1.28	0.75	***	1.21	0.44	***	1.08	0.04	(*)
Low education (n = 192)	1.37	0.31		1.38	0.17	*	1.18	0.1	(*)	1.13	0.08	
Higher education (n = 927)	1.53	0.14	***	1.20	0.08	**	1.21	0.04	***	1.07	0.04	†
Foreign (n = 155)	1.41	0.37		1.48	0.13	***	1.58	0.19	***	1.31	0.15	*
CH/FR (n = 1,017)	1.19	0.19		1.28	0.07	***	1.26	0.09	**	1.24	0.1	**
Reluctant (n = 562)	1.19	0.12	†	1.20	0.04	***	1.21	0.05	***	1.19	0.06	***
Motivated (n = 610)	1.06	0.09		1.09	0.04	**	1.12	0.05	*	1.05	0.05	

*** p<0.001, **p<0.01, *p<0.05, †p<0.10.

Results show similar mode effects between respondents by nationality. For French and Swiss respondents, and for respondents from other places, the effect of mode was significant in 4 cases (all apart from those for the 11-category score). However, looking at the odds ratios it is possible to see that they are quite similar in all cases, and that, although the odds ratios are slightly bigger for “other country” respondents, they point in the same direction.

When examining the mode effects in responses of people with different levels of education, there were differences between the low educated respondents and the rest for the responses with 11, 7 and 5 response options. In every case, the effect of telephone appears to be stronger for respondents with the higher level of education. However, the strongest effect was found when looking at the responses to the 11 scale score in both groups, even though the effect was bigger for the more educated respondents.

The differences between more motivated and reluctant respondents appear in the final two columns of the table. They show different sizes of mode effects for some of the composite scores related to the SWB measures. For those measures that are sensitive to mode in both groups, the effect sizes were bigger in the reluctant group in all cases.

To better assess the interplay of the effects of mode and respondents' characteristics I assessed whether the effect that mode has on the subjective wellbeing outcome differs depending on the respondents' characteristics. Table 38 presents the results from the analyses that were implemented adding interaction effects between the two variables, providing information about the combined effect of mode and respondents' characteristics. Of all the significant effects reported in table 38, only two remain significant after adjusting for multiple testing: the interaction effects of mode and age on self-reported health, and the interaction of mode and motivation to respond on life satisfaction. However, the effect does not happen for the overall variable, but is identified in just one category for each item.

Table 38. Interaction effects between mode effects and respondents' characteristics

Self-completion (n = 634) and telephone (n = 808) comparison
Statistically significant interaction effects by response category, adjusted using Holm-Bonferroni correction.

Self-completion mode is the base outcome

Interaction effects	Categories		Marginal effects		Contrast	Std.Err.	Sig.
Mode & sex			Female	Male			
Positivity	(1-5)	5	0.09	0.22	-0.12	0.05	(0.025)
Accomplishment	(1-5)	4	0.01	-0.07	0.07	0.03	(0.007)
Take control	(1-4)	4	0.03	-0.14	0.16	0.06	(0.010)
Anxiety	(1-4)	3	0.00	-0.12	0.13	0.07	(0.051)
		4	0.00	0.12	-0.14	0.06	(0.033)
Mode & age			Old	Young			
Things going well	(1-4)	3	0.06	-0.10	0.17	0.07	(0.013)
Social activities	(1-5)	3	0.11	0.02	0.07	0.03	(0.040)
Positivity	(1-5)	5	-0.01	0.20	-0.19	0.07	(0.005)
Health	(1-5)	2	-0.05	0.00	-0.04	0.01	(0.018)
		3	-0.13	-0.02	-0.09	0.04	(0.036)
		4	0.08	-0.05	0.11	0.03	0.001
		Total					(0.025)
Mode & education			Low education	Higher ed.			
Happiness	(0-10)	3	0.00	0.00	0.01	0.00	(0.015)
Meets close ones	(1-7)	5	-0.02	-0.04	-0.05	0.03	(0.044)
Handle problems	(1-4)	1	0.11	0.02	0.11	0.05	(0.028)
Social activities	(1-5)	3	0.11	0.02	0.08	0.04	(0.033)
Health	(1-5)	2	-0.05	-0.01	-0.03	0.01	(0.015)
		3	-0.13	-0.03	-0.08	0.04	(0.024)
Mode & nationality			Other country	CH/FR			
Happiness	(0-10)	3	0.00	-0.01	0.01	0.00	(0.014)
Handle problems	(1-4)	1	-0.02	0.05	-0.09	0.04	(0.036)
		5	0.33	0.08	0.23	0.10	(0.019)
		Total					(0.020)
Positivity	(1-5)	4	-0.24	0.04	-0.28	0.10	(0.004)
		5	0.33	0.12	0.18	0.10	(0.054)
							(0.010)
Restless sleep	(1-4)	1	-0.03	0.01	-0.04	0.01	(0.004)
Mode & reluctance			Reluctant	Motivated			
Life satisfaction	(0-10)	4	-0.05	0.00	-0.04	0.01	0.001
Freedom	(1-5)	4	-0.06	-0.21	0.15	0.07	(0.023)
Loneliness	(1-4)	1	-0.04	0.01	-0.03	0.01	(0.004)
		Total					(0.032)

In the first case, for the item self-reported health, the results showed that the additive effect of mode on responding by telephone on the category 4 is an addition of 0.08 for respondents older than 65, and a reduction of 0.05 for those younger than 65, with the difference between the two being statistically significant (p = 0.001). The second case, on the life satisfaction measure, the marginal effects results showed that the additive effect of responding by telephone on the category 4 results is a reduction of 0.05 for reluctant respondents, and 0.003 for motivated ones. This means that while telephone respondents are on average less likely to choose the category 4 compared to mail ones, there are significant differences in the average marginal effects (p = 0.001).

The table below shows that there were no cases in which there was a statistically significant combined effect of mode of data collection and respondents characteristics, once the p-values were adjusted

Table 39. Interaction between mode effects and respondents' characteristics
Regression coefficients (n = 1,172)

SWB measures	Mode & Sex		Mode & Age		Mode & Education		Mode & Nationality		Mode & Reluctance	
Mean scores	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
11 categories	0.03	0.17	-0.17	0.22	-0.02	0.24	-0.07	0.29	0.21	0.17
7 categories	-0.02	0.11	-0.19	0.15	0.31	0.15 (*)	-0.17	0.19	0.06	0.11
5 categories	-0.07	0.07	-0.01	0.09	0.05	0.10	0.04	0.11	0.02	0.07
4 categories	-0.05	0.06	-0.01	0.09	0.08	0.09	0.00	0.10	0.09	0.07

*** p<0.001, **p<0.01, *p<0.05, †p<0.10

Lastly, I present the results from the interaction of the different indicators used in the empirical study 2 with the different respondents' characteristics. The table below shows that there were no significant interaction effects between mode and any of the respondents' characteristics examined on nonresponse rates, neither on the positivity of the themes reporter, nor on the themes of the negative life events.

However, there was a significant interaction effect between response length and age. The coefficient for the interaction is 3.50, indicating that older respondents in telephone give longer responses than younger respondents in the same mode.

Table 40. Interaction between mode effects and respondents' characteristics, open-ended questions (n = 1,172)

	Mode & Sex (Female = 1)		Mode & Age (Older than 65 = 1)		Mode & Education (Low education = 1)		Mode & Nationality (No CH/FR = 1)		Mode & Reluctance (Reluctant = 1)	
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Nonresponse 1	0.03	0.03	-0.04	0.04	0.01	0.05	0.08	0.07	-0.06	0.03(*)
Nonresponse 2	0.00	0.04	0.00	0.05	0.01	0.07	0.14	0.09	-0.06	0.04
Nonresponse 3	-0.04	0.05	-0.09	0.06	0.02	0.07	0.13	0.09	-0.06	0.05
Length all	-0.40	0.82	3.50	0.96***	2.11	1.05 (*)	1.67	1.31	-0.18	0.81
Positivity	-0.17	0.13	-0.07	0.18	0.11	0.21	-0.32	0.21	0.00	0.13
Themes										
Marriage	-0.01	0.06	0.07	0.08	-0.01	0.08	-0.11	0.09	0.06	0.06
Loss	0.11	0.05(*)	0.03	0.06(*)	0.13	0.06	-0.13	0.05 (**)	-0.03	0.05
Children	0.05	0.06	-0.10	0.08	-0.12	0.08	-0.12	0.10	-0.01	0.06
Separation	0.02	0.04	-0.01	0.04	-0.05	0.04	0.00	0.04	-0.01	0.04

*** p<0.001, **p<0.01, *p<0.05, †p<0.10

6.5. Discussion and conclusion

In this investigation, the aim was to assess the differential impact of mode of data collection for different subgroups of the population. Survey methodology studies pointed towards the idea that some respondents' characteristics may be related to the quality of the information given (Collins, 2003), while at the same time there has been a close collaboration between survey methods and cognitive psychology that has provided the field with information on how to understand the response process (Belli, Conrad & Wright, 2007). Ross and Mirowsky (1984) warned that response tendencies – whether to agree or to report socially desirable information – are more common in some population groups than in others. These authors found that people in vulnerable situations, those not well integrated in the rest of society, such as some foreigners that may not speak the native language, can be particularly prone to try to give a good impression when answering questionnaires. In addition, other types of respondents such as older ones, which are in more dependent positions and may seek others' approval, are more likely to give less truthful answers

Having implemented a comparison of the mode effect in responses of both closed-ended and open-ended questions about wellbeing and vulnerability, I found that there is an interrelation between mode of data collection, response process and respondents' characteristics. This is, I found that the effect of mode of data collection varies depending on type of respondent.

Differences between the subgroups, however, were not very common, such as the unique significant interaction effects between respondents' characteristics and mode of data collection was between response length and age. However, based on the literature review, it is possible that older respondents, that may have more trouble writing or typing their answers, are the ones that give longer responses in telephone – while younger respondents may be more used to complete forms or exams and feel more at ease to express themselves.

The effect of the interaction between subgroups and mode in closed-ended questions, on the other hand, also showed few significant cases. In particular, the results showed that older telephone respondents tend to choose the category 4 (the "good" option) to a higher extent than younger than 65 respondents also responding by telephone.

Another important finding was that there were differences in the effect of responding by telephone between motivated and reluctant respondents in the extent to which they choose category 4 out of the 11 response alternatives to the question on life satisfaction, which reluctant respondents are less likely to choose compared to motivated ones.

These findings partially support the expectations about the mode affecting people in different ways, and could be interpreted as differences potentially due to social interaction with the interviewer. Speculating, a potential reason for older respondents to choose more the "good" option is to be agreeable during the social interaction, and understanding the health state in a more relative way than younger respondents. At the same time, the fact that responses of reluctant respondents were less likely to choose a not-so-good category of subjective wellbeing in telephone than in mail, could indicate a confirmation that less motivated respondents can be more affected by mode and question design (as it is a middle category, instead of choosing one of the most extreme responses). However, this is outside of the scope of this thesis.

At the same time, one of our main challenges was related to the fact that different modes of data collection attract different types of people, people that not only differ in socio-demographic characteristics but also in subjective wellbeing and personality characteristics. This complicates the objective of identifying differences exclusively due to response styles: it is possible that the sample composition for the different modes of data collection is different too.

In spite of the shortcomings, results from this chapter allowed looking at the impact of mode and the respondents' characteristics that are so often studied in vulnerability studies. It is possible to conclude that, although few, the type of respondent can have an interaction with mode of data collection, in widely used variables such as life satisfaction and self-reported health.

In the next chapter, I will continue investigating the repercussion that mode effects can have on substantive data by looking at three regression models that look at the relationship between subjective wellbeing measures and their predictors.

CHAPTER 7. DO MODE EFFECTS MATTER? IMPLICATIONS FOR THE ANALYSIS OF WELLBEING AND ITS PREDICTORS

7.1. Introduction

Subjective wellbeing is composed of different aspects such as social activities, wellbeing at work, or individual indicators of quality of life including levels of depression and anxiety. These measures are used in different types of analysis, in which researchers can investigate the predictors of wellbeing (Dolan, Peasgood & White, 2008) or look at the relationship between different wellbeing indicators (Helliwell & Putnam, 2004). They can also be used as a predictor of measures such as health, income and social relationships (de Neve, Diener, Tay & Xuereb, 2013; Diener et al., 2017). Such analyses of interest to social science researchers are often based on statistical analyses using data that comes from multiple modes of data collection, either due to country comparisons, the combination of multiple-data sets from the same country, or mixed-mode survey designs, as they have become increasingly popular (Dillman et al., 2014). These situation can complicate the way in which survey data is used (Martin & Lynn, 2011) and the conclusions that researchers draw from their analyses.

Survey researchers have worked on identifying the sources of error that are associated to a greater or lesser extent with the different modes, and have put considerable effort into investigating the effects of using mixed-mode designs on item nonresponse, response validity, social desirability and self-disclosure to help the many social researchers that have to deal with mixed-mode data (Maggiori, Rossier, Krings, Johnston & Massoudi, 2014; Martínez-Martí & Ruch, 2016). However, relatively few studies have attempted to tackle the more pertinent question of whether mode effects actually matter for the kinds of analyses social scientists typically conduct using survey data.

If means and distributions across wellbeing variables are affected by mode – whether via social desirability bias or via response effects – then it is important to understand the extent to which the relationships between variables might be affected. Yet there is still a lack of studies looking at the impact of mode of data collection on statistical analyses such as correlations or regression coefficients. First studies show that mode effects can impact composite scores (Revilla, 2013) and regression coefficients (Dolan & Kavetsos, 2012). In the field of subjective wellbeing, Dolan and Kavetsos (2012) found that the correlation between life circumstances and wellbeing differed between a telephone and a face-to-face survey; and Conti and Pudney (2011) and Jäckle and Pudney (in McFall, 2012) found that mode effects had an impact on predictors of wellbeing at work. Although in some cases the regression coefficients varied across modes, mode of data collection was not always found to affect the coefficients' sign and significance (Sarracino, Riillo & Mikucka, 2017). These findings suggest it is possible that mode effects, although affecting some univariate survey estimates, do not provoke differences in the correlational and relationship analysis of variables of interest. However, the extent to which mode effects contribute to differences in substantive research on wellbeing remains unclear.

In the previous empirical chapters, I showed that the way in which data are produced had an impact on how people responded to survey questions about specific variables on the topic of subjective wellbeing. However, mode effects were not the same for all the measures tested and the effect of mode was medium or small. For this reason, it is necessary to find out whether response differences (mode related measurement effects) do actually impact the analysis of social researchers working on the topic of wellbeing. The aim of this chapter is, therefore, to contribute to the understanding of whether mode effects in individual items of wellbeing have repercussions in the analysis on the data. In order to do this, I use data from the LIVES methodological mixed-mode experiment to replicate regression analyses that are widely used by researchers to predict happiness, satisfaction at work and social trust for each mode of data collection, with and without controlling for selection effects. Specifically, the aim is to investigate whether the question-answer process – which differs for each mode of data collection – influences the way in which the respondent interprets the question and the repercussions this has on the responses given (Hox et al., 2015). In addition, I look at whether conclusions drawn from comparisons across population subgroups could also be affected by mode effects.

In the following lines, I present an overview of research that has already been done conducted on this topic. Then I present the conceptual models that I compare across modes, which are based on widely replicated analyses in wellbeing research, and explain the methodological approach I use to implement the analyses and compare the regression coefficients across modes. After presenting the results, I discuss whether mode plays an important role in the conclusions obtained in cross-sectional research on wellbeing when comparing web, mail and telephone survey data.

7.2. Literature review

7.2.1. The impact of mode in substantive research

Research on mode effects has generally focused on estimating the magnitude of differential errors across modes and exploring ways to minimize them. Specifically, researchers such as Lugtig (2011), Revilla (2013), Vannieuwenhuyze and Loosveldt (2013), and Cernat (2015) among others, have explored ways of measuring the extent of the bias introduced due to mixing modes of data collection in order to reach conclusions about whether it is worth mixing different modes of data collection. While some of these studies show evidence for the fact that using mixed modes has an impact on the comparability of results, comparisons of point estimates and population subgroups can be complicated (de Leeuw, ESRA 2017). Previous studies showed that mode effects can impact composite scores (Revilla, 2013), multivariate analysis (de Leeuw, Mellenbergh & Hox, 1996), correlations (Vannieuwenhuyze, 2015) and regression coefficients (Jäckle, Roberts & Lynn, 2010; Dolan & Kavetsos, 2012). In the present study, I focus on the results from studies looking at the effect of mode in regression coefficients and correlations, while in the next chapter I will summarize the results from the studies that have looked at the effect of mode in multidimensional measures.

To determine the impact of mode in regression results, Jäckle, Roberts and Lynn (2010) used data from a mode experiment conducted alongside the European Social Survey from 1) Hungary collected using face-to-face, telephone, mail and web modes, and from 2) Hungary and Portugal collected with face-to-face and telephone interviews. Their study consisted of OLS models predicting attitudes towards immigration by mode, a series of socio-demographic variables and additional items on voting behaviour, political interest, gender-role attitudes, social trust, life satisfaction and religiosity. Results only showed mode differences for the relationship between voting behaviour and immigration attitudes, which was not significant in the telephone mode.

Martin and Lynn (2011) looked at the differences in the estimates of regression coefficients across a single mode face-to-face survey and two mixed-mode survey designs that involved face-to-face, telephone and web modes from the European Social Survey carried out in the Netherlands between 2008 and 2009. They found some differences when comparing results from the different survey designs and argued that they were due to measurement effects and not so much due to differences in the types of respondent answering each mode. Their results showed that economic egalitarianism was positively related to interest in politics in the mixed-mode survey design (using telephone, web and face-to-face) but the direction of the relationship was the contrary for the face-to-face results (independently of controls for differences in age, sex and education). For the rest of their analyses, there were no significant differences between the two survey designs.

To examine the question of the relationship between variables differing across modes, Vannieuwenhuyze (2015) carried out an analysis of data from another European Social Survey mode experiment conducted in Estonia, the data from which had been collected using computer-assisted personal interviews and self-completed web questionnaires. He used a multi-trait-multi-mode approach that did not show a strong effect of mode in the results from the correlation between the items on the topics of democracy, depression and feeling of engagement with everyday life. However, he found small to moderate measurement and selection effects when using a simplified version of the previous MTMM model. He estimated the size of measurement effects and found that the estimates for correlations were not consistent across the variables that measure democracy, but they were in the case of variables measuring depression and life engagement: the correlations between the variables were larger in the case of responses given by telephone respondents than for web respondents.

In the field of quality of life studies, not much attention has been paid to how survey design influences data quality and its statistical analysis (Pudney, 2010) and what little research there is has mainly been related to operationalization issues, without a clear focus on mode of data collection (Fleche, Smith & Sorsa 2011). In fact, Pudney (2010) argues that for certain topics – such as the case of surveys that measure wellbeing – there are reasons to think that survey estimates may be affected by survey design, due to both differential coverage of the population of study by the different modes of data collection and to measurement effects. Not only is impact of mode on survey results confined to unequal distributions and averages of subjective wellbeing measures by mode of questionnaire administration, it also has an influence on its relationship with other variables – whether predictors or correlates (Dolan & Kavetsos, 2012) – to the extent that we would obtain different conclusions on which situations or characteristics are the drivers of happiness or life satisfaction.

Dolan and Kavetsos (2012) showed how four measures of SWB (life satisfaction, happiness, feeling worthwhile, and anxiety) are sensitive to mode when comparing telephone and face-to-face. Specifically, they found that the level of wellbeing was higher for the telephone mode, and that telephone respondents were more likely to choose the categories 9 and 10 for life satisfaction, happiness and anxiety and feeling worthwhile than face-to-face respondents. In addition, wellbeing predictors varied significantly depending on the mode of administration and might bias the way the data is analysed: socio-demographic variables that have a negative effect on wellbeing when the interview mode is face-to-face were not important predictors of the levels of wellbeing when the interview is implemented by telephone. To illustrate this, if sex and marital status seem to play an important role when predicting life satisfaction, the effect disappears when the interview was implemented over the telephone.

For this reason, differences in the distribution of measures such as self-reported health (Schwartz et al., 1991; Bowling 2005; Skashaug et at., 2010), financial difficulties (Breuning & Mckibbin, 2011), depression (Li et al., 2012); and for wellbeing in particular for measures of happiness, life satisfaction and optimism (Dolan & Kavetsos, 2012). The objective is to investigate the relationship between the variables, to find out whether conclusions may be biased due to the effect of mode of data collection, because the situation can worsen when one of the objectives is to be able to raise information from all population sectors, including difficult-to-reach groups in general, for example, are often mentioned as groups that are likely to be underrepresented in telephone surveys (Laganà et al., 2011).

Pudney studied the effect of mode in the relationship between wellbeing measures and its predictors in several studies (Conti & Pudney, 2011; Holford & Pudney, 2015; in McFall, 2012; Pudney, 2010). In the first of these, he found that the impact of health, income, gender, family size and hours of work varied for different survey design features. Using data from a methodological experiment implemented within the UK Household Longitudinal Survey ('Understanding Society') and looking at data collected by telephone, face-to-face and computer-assisted self-interviewing, he examined responses to questions on a general life satisfaction question, and on satisfaction with different aspects of people's lives: health, household income, and available amount of leisure time. Due to the experimental design, the question format varied the number of response options (including a scale with seven categories, or a two-stage branching question providing three options for each question, tailored based on whether they had chosen from the options "dissatisfied", "neither satisfied nor dissatisfied" or "satisfied"). Their analysis separated results for men and women, based on previous work by Conti and Pudney (2009), in which they did not find the same gender differences in the relationship between satisfaction and work, hours of work and work-life balance. In his study from the year 2010, Pudney looked at the effect of gender, income, working hours, and number of children in the satisfaction measures previously mentioned, finding that responding in telephone mode appeared to increase the difference between men and women's job satisfaction, and the difference in the interaction between gender and working hours. In addition, mode effects varied for respondents with different levels of income and family composition. The same differences between the mode effect in the interaction of gender with hours of work predicting job satisfaction were found by Jäckle and Pudney (McFall, 2012).

Holford and Pudney (2015) investigated the way in which different survey designs influence substantive research on the topic of subjective wellbeing. In their study, they compared ACASI to face-to-face estimates on the one hand and telephone to face-to-face estimates on the other. Looking at the same dependent variables as Pudney's previous work in 2010, they examined the relationship of wellbeing with gender, income, and their interaction; and found that the use of face-to-face interacts with income and gender when predicting satisfaction with income and satisfaction with health: the effect shows women reporting higher satisfaction levels and underreporting the relationship between income and income satisfaction.

Saris and Revilla (2016) devoted some attention to the impact that measurement effects may have in analyses of the relationships between different variables. They illustrated their conclusions giving the example of the relationship between job satisfaction and life satisfaction, two items affected by measurement errors, which made them concerned about the interpretation of the results. For example, a researcher interested in estimating the effect of age on both of the variables of interest, could arrive to the wrong conclusion when looking at the correlation of age with the two measures.

More recently, Sarracino and his colleagues (2017) look at mode effects in estimates of regression coefficients in a telephone and web comparison in which they look at five measures of subjective wellbeing: life satisfaction, having obtained important things the respondents wanted in their lives, whether they would change anything if they were born again, whether their life conditions were excellent, and whether their life was close to their ideal. Respondents could choose between 5 response categories, ranging from strongly agree (1) to strongly disagree (5). Their results showed that there were significant differences in the relationship between variables when comparing the telephone and web samples, although the sign was the same for the two modes. For example, the strength of the relationship between life satisfaction and income (which was found to be stronger in web than in telephone), and the item "If I could live my life again I would not change anything" and age differed.

7.2.2. Research questions

In the previous chapters, I tested the extent of selection and measurement related mode effects using data from a mixed mode experiment and found significant differences in terms of who answers and how they answer. In this chapter, I focus on examining the effect that mode of data collection may have in commonly implemented substantive research on the topic of wellbeing. Based on the findings from the literature review, specifically, I aim to answer two research questions:

R.Q.1: Do mode effects impact conclusions about wellbeing and its predictors in commonly used regression analyses?

Answering this question, I aim to show whether different social researchers studying wellbeing and using the same statistical models, whose data was collected using different modes of data collection, would arrive at the same conclusions. In previous research, such as that of Dolan and Kavetsos (2012) and Holford and Pudney's work (2015), individual items affected by mode were found to vary in how they were associated to other items on the topic of subjective wellbeing, particularly in the topic of subjective wellbeing. However, no previous research has been able to compare whether this may be the case in Switzerland.

RQ2: Are there significant differences in the interaction between respondents' characteristics and predictors of subjective wellbeing depending on mode?

Research shows that different types of respondent may respond differently, not only due to mode of data collection, but also due to their cognitive level and their position in society. Although in the previous chapter I did not find many significant differences in mode effects across subgroups, I expect there may be some differences in the interaction between the predictor variables and some socio-demographic characteristics depending on which type of respondent is answering, for example, comparing men and women.

7.3. Methods

To address the research questions, I chose four widely used regression models predicting happiness, social trust, and job satisfaction. I replicate the models across modes to find whether researchers using the same analytical approach but different modes of data collection would arrive to the same conclusions.

7.3.1. Data

The analysis involves respondents to the main survey questionnaire, who answered through the mode they were asked to in the first place, as was the case for the previous chapters, and only respondents with a listed telephone number in order to study samples with similar sample compositions. This decision was taken to avoid additional confounding errors that could impede interpretation of the results. The sample sizes by mode are as follows: there were 457 web respondents, 351 paper respondents, and 364 telephone respondents.

The last part of the analysis only includes respondents that have a paid job, and as a consequence, there are 308 web, 208 mail and 214 telephone respondents.

7.3.2. Analytical approach

The analysis consists of four ordered logistic regression analyses that have three different subjective wellbeing indicators as dependent variables, and a series of predictors based on previous substantive research on the topic. The models are predictions of happiness, social trust and job satisfaction. Measures of subjective wellbeing are ordinal, but previous research has treated them as cardinal without the results being altered (Dolan & Kavetsos, 2012; Ferrer-i-Carbonell & Frijters, 2004). To help choose the type of regression to be used, I compared the BIC, Bayesian Information Criterion, to test the overall fit of the models and help decide whether ordered logistic regressions or OLS fitted the data best (the modes are introduced in the next section). Smaller BIC coefficients indicate better fit. If the difference in BIC between two models is 10 or more, there is strong evidence that the model with the lowest BIC is a better option.

Table 41. BIC comparison across types of regression

		Web	Mail	Telephone
Model 1	OLS	1483.745	1078.677	1001.967
	Ordered logit	1427.726	1047.466	969.9
Model 2	OLS	1580.157	928.718	986.214
	Ordered logit	1524.088	888.625	962.987
Model 3	OLS	1826.785	927.759	1218.66
	Ordered logit	1802.884	932.737	1210.015
Model 4	OLS	809.855	477.093	545.811
	Ordered logit	782.471	451.467	522.497

Once the ordered logistic regressions was chosen, I implemented and approximate likelihood-ratio test of proportionality of odds across response categories for each model, the results showing that it was appropriate to implement in the four cases.

7.3.3. Predicting happiness, social trust and job satisfaction

Model 1: Predictors of happiness I: life events

The **happiness** model is based on the relationship between the accumulation of stressors and happiness, and it is related to the ability to cope with such stressors during the life-course.

The model is based on similar analyses reported by e.g. Norris and Murrell (1987), Ballas and Dorling (2007), and Madero-Cabib (2015). Although the impact of negative life experiences is stronger when recent while the memory of positive events remains for a longer time, and older respondents have been found to report fewer vulnerable moments than younger respondents (Dasoki, Morselli & Spini, 2015), I have included this measure to compare the effect between the different modes of data collection. In addition, it illustrates the effect that mode may have in the reporting of events and its interaction with age. Because the effect of suffering from negative events has been found to be stronger for poorer respondents, I also looked at the interaction of having a low income with having gone through negative events.

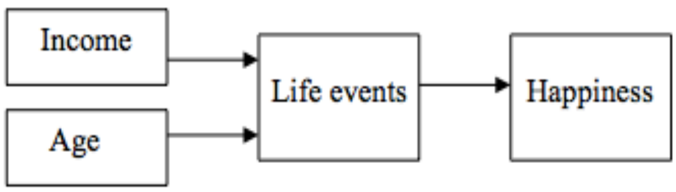


Figure 9. The relationship between happiness and important life events

Model 2: Predictors of happiness II: income

An additional prominent topic in the analysis of happiness and life satisfaction is their relationship with household income, which has been found to be significant in over 50 countries across the world (Diener & Biswas-Diener, 2002; Layard, 2005; Pouwels, Siegers & Vlasblom, 2008; Stevenson & Wolfers, 2013). Research on the topic often found that the relationship between the two measures is positive although moderate (Dolan & Metcalf, 2011). In these studies, it has been found that the risk of unhappiness is much higher for poorer respondents, which have also been found to be more likely to experience more stressful life events (Diener & Biswas-Diener, 2002), although this effect is smaller in rich countries such as Switzerland (Frey & Stutzer, 2000). For this reason, I use both variables of household income and difficulty of living with income. It is also important to note that this relationship has been found to be different for men and women, in that the correlation between happiness and income is not significant for women. However, single women with low incomes were more likely to suffer from depression, which was not the case for married women (Diener & Biswas-Diener, 2002). Another commonly studied gender difference is that the effect of being unemployed on subjective wellbeing is stronger for men than for women (van der Meer, 2014).

Other researchers, however, have identified that income was a stronger predictor of the absence of negative emotions, such as sadness, restlessness, hopelessness, worthlessness, nervousness, and feeling that everything was an effort, than of a high level of happiness (Clingingsmith, 2016; Kushlev, Dunn & Lucas, 2015). Lastly, Weiting and Diener found that satisfaction with income was a particularly strong predictor of subjective wellbeing (2014).

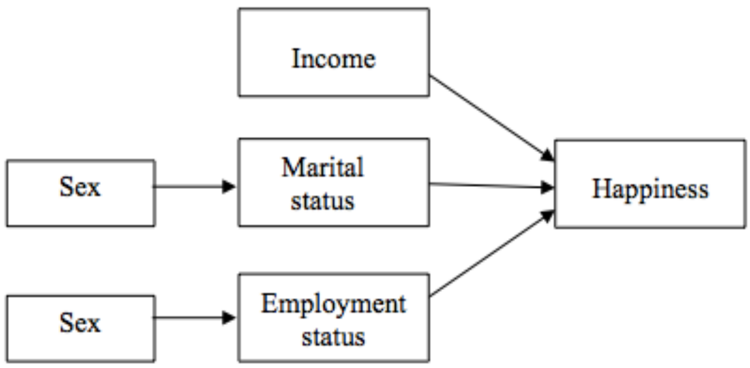


Figure 10. The relationship between happiness and its predictors income, marital status and job status

Model 3: Predictors of social trust

It is possible to find two main theories that aim to explain **social trust** and that can be combined: the voluntary organization theory, which concerns the involvement of individuals in different organizations and/or associations, and the success and wellbeing theory (Newton, 2013). The model replicated here is proposed by the education resource within the European Social Survey, 'ESS EduNet', which is a widely consulted source to guide the research process for the study of various topics that are covered by the European Social Survey. The measures related to the success wellbeing theory were found to be stronger predictors than participating in organizations and associations (Newton, 2013, see chapter 3).

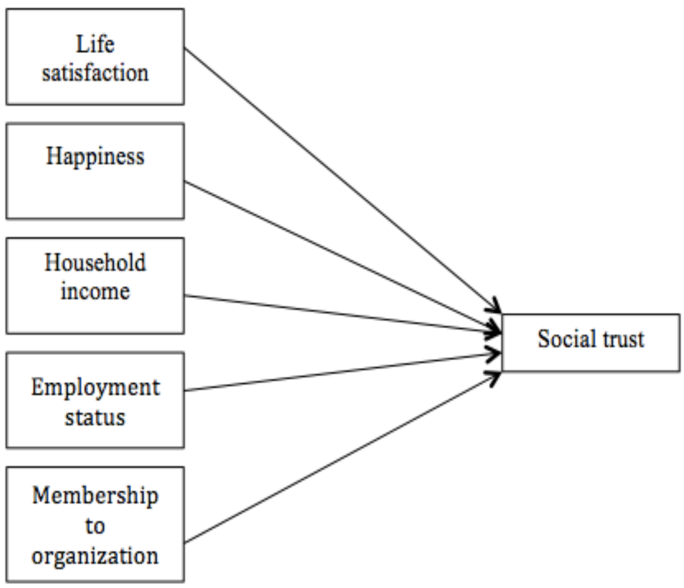


Figure 11. Social trust and its predictors

Model 4: Predictors of job satisfaction

The factors that determine **job satisfaction** are related to different job characteristics, previous employment experiences and social characteristics. Hulin and Judge (2003), as well as Judge and Kingler (2007) identify multiple aspects related to the evaluation of the job, such as the emotional and behavioural response. Figure 12 shows the different predictors that have been previously used to predict satisfaction at work (Lockwood, 2003; Ahn & Garcia, 2004; Clark, 2005; Davoine, 2005). In addition, the level of education is often included as one of the factors that affects satisfaction at work. Previous findings show respondents with secondary and university education are more likely to be satisfied with their jobs than those respondents with primary and no education levels (Millán, Hessels, Thurik & Aguado, 2013). However, this also depends on whether people's work corresponds to their level of education (Allen & van der Velden, 2001). Being male or female also makes a difference in the level of satisfaction at work (Singhapakdi et al., 2014) and importance of job security, these items being more important for women than for men. Millán and colleagues (2013) also found the importance of job security was not very important for middle-aged respondents.

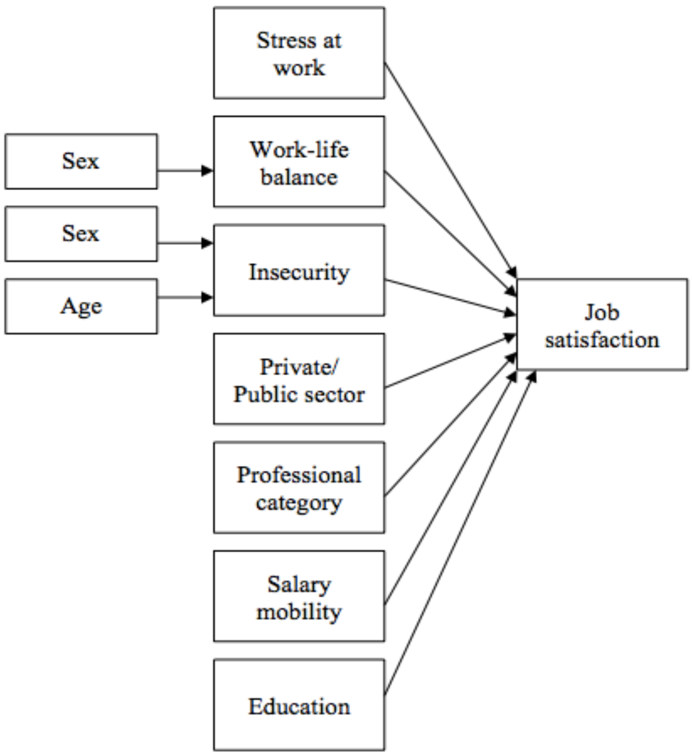


Figure 12. Satisfaction at work and its predictors

The results from the regressions I implemented based on these models predicting happiness, social trust and life satisfaction are presented with and without socio-demographic controls using a propensity score matching, which I will explain in the final part of the methods section. To be able to examine differences in the models across the modes, I examine the value of the McFadden R-squared, and indicator of the fit of the model, but that cannot be interpreted as R-squared in an OLS model (Williams, 2017) and its level of significance, in addition to the coefficients indicating the strengths of the relationship between the dependent variable and the predictors. In order to compare the fit of the models, I also report the fit statistic BIC, which was additionally used to decide whether an OLS model was a better choice than the ordered logistic regression model.

In addition, I used the Variance Inflation Factor (VIF) to check that multicollinearity between the variables household income and difficulty to live with income was not an issue estimating the model. If the VIF is smaller than 10, it is not a problem. In this case it was between 1.11 and 1.44 for all samples.

In a second step, to answer the second research question, I include an interaction effect between mode of data collection and population subgroups, to assess whether there are any differential mode effects depending on respondents' characteristics in the models. In particular, I look at the interaction between mode and age, sex and education, as they have been the main object of attention in both subjective wellbeing research and survey methodology looking at differential mode effects across subgroups of the population.

7.3.4. Variables

In particular, I look at three different situations for which the dependent variables are happiness, social trust and job satisfaction (see table 42).

Table 42. Outcome variables: quality of life measures

Question	Source	Categories
Taking all things together, how happy would you say you are? (Very unhappy- very happy)	ESS	0-10
Generally speaking, would you say that most people can be trusted, or that you can't be too careful in dealing with people?	ESS	0-10
All things considered, how satisfied are you with your present job? (Very unsatisfied-very satisfied)	ESS	0-10

The first empirical study showed that responses to the questions about happiness, social trust and job satisfaction were affected by mode. In fact, mode effects were found in both their response distributions and estimated means.

For each variable, I include a particular set of predictors based on existent studies (see previous section) and the availability of measures available in the methodological experiment questionnaire.

In is important to indicate that some of the predictors' categories have been collapsed into fewer categories to simplify the analysis where it has been deemed adequate to do so. In the following lines I present which predictors are included for each situation, show their statistical distribution and indicate whether mode measurement effects affect them.

Predictors

Predictors of happiness:

<u>Negative important events (0-3)</u>: Three open questions ask about the events that have been important in respondents' lives and once respondents have given that

information they have to report whether the experienced event was negative or positive. I merge the three variables into one in order to obtain one variable of important negative events in respondents' lives with four categories including the value 0 for when they do not mention any negative event; 1 for 1 negative event; 2 for 2 negative events and 3 for 3 negative events.

Table 43. Percentages for the measure of accumulation of negative events

Past negative events (0-3)	WEB (%)	MAIL (%)	CATI (%)
0	55.8	52.4	55.0
1	33.6	35.3	32.5
2	8.0	10.0	10.0
3	2.7	2.4	2.6

Income I: This variable measures household income and was collected differently for telephone respondents, who were asked to provide a specific number in contrast to web and mail respondents who had to choose between 10 different monthly income intervals:

Table 44. Income I (10 categories)

Level of Income	%
1	16.45
2	7.11
3	8.70
4	8.28
5	10.62
6	10.62
7	13.91
8	7.75
9	8.70
10	7.86

Income II: New variable that differentiated respondents with a low income (less than 4100 Swiss Francs per month) from those with a higher income (more than 4100 Swiss Francs per month).

Table 45. Income distribution by mode of data collection

	WEB (%)	MAIL (%)	CATI (%)
More than 4000 chf/month	86.36	83.75	55.63
Less than 4000 chf/month	13.64	16.25	44.37

Social-trust predictors:

Happiness (0-10). It is the same variable that works as the dependent variable in the first scenario, with 11 categories.

Life satisfaction (0-10). Another measure of wellbeing measures with 11 categories.

Membership of an association or organization. One of the questions measures whether the individual belongs to a series of organizations or associations. As there is no indication that social trust increases linearly related to the number of belongings, I created a variable with three categories that indicate whether the respondent does not belong to any association or organization, belongs to 1, or whether he or she belongs to 2 or more.

Income I. As described above.

Difficulty of living with income. To simplify the analysis, this question's four categories have been merged into two, measuring whether respondents find it difficult to live with their current household income. This way, there are two categories: 1 if they find it difficult (they do not manage well at all, or not very well), 0 if they do not find it difficult (they manage well or very well).

Job situation. The two categories for this question have been kept as in the original dataset: being employed (1) and not employed (0).

Table 46. Percentages and means for social trust predictors

Predictors		WEB	MAIL	CATI
Happiness (0- 10)		7.8	7.6	8.1
Life satisfaction (0-10)		7.7	7.5	7.9
Membership (%)	No membership	25.0	37.0	27.8
	1 membership	30.5	26.7	25.3
	More than 1 membership	44.5	36.4	46.8
Income (1-10)		5.9	5.6	5.5
Does not have paid job (%)		32.5	39.4	41.2

Job satisfaction predictors:

Works in public or private sector. The measure has been collapsed into two categories: working in the public sector (including public administration workers, public companies and other public sectors) and not working in the public sector (including jobs in private companies but also independents and 'other' categories). The value for the category working in the public sector is 1, 0 otherwise.

Work-life balance has been found to be one of the main drivers of satisfaction at work; the variable we use to measure it is an 11-point scale that goes from 0 to 10.

Stress at work is a scale variable with values that go from 0 for the lowest level of stress up to 6 for the highest level.

Past period of unemployment indicates whether a respondent has been unemployed for a period of three months or more during the past (1 if yes, 0 if not).

Likelihood of losing their job is coded as 1 if the respondent considers that it is likely that they will lose their job in the next 12 months (collapsing the categories 'very likely' and 'quite likely'), and 0 if they do not feel they are at risk of losing their job (including 'not very likely' and 'not likely at all').

Table 47. Percentages and means for job satisfaction predictors

Predictors	WEB	MAIL	CATI
Works in public sector (%)	37.1	30.0	41.3
Work-life balance (0-10)	7.0	7.0	7.4
Level of stress at work (0-6)	2.4	2.2	2.7
Past unemployment (%)	24.9	22.2	22.3
Likely to lose job (%)	8.9	10.7	10.6

In addition to the main predictors, in each of the 3 models I also included a series of socio-demographic variables to control for selection effects. These variables are used to create a propensity score to include in the regression model (which I explain in the last part of this methods section), or they are used to create interaction terms that are widely used in wellbeing research (Diener et al., 2000). These socio-demographic items are:

Sex: male (1) and female (0)

Age in years: I use it as a scale variable when implementing regression analyses, and as a 4-category ordinal variable when looking at the different modes' sample composition. The groups are younger than 30 years old, between 30 and 44 years old, between 45 and 64 years old, and 65 years old and above.

Country of birth: whether respondent was born is Switzerland (1) or not (0).

Marital status indicates whether the person is married or in a civil partnership (coded 0), or not married nor in a civil partnership (including single respondents, widows, divorced and separated respondents) (coded 1).

Job situation: The two categories for this question have been kept as in the original dataset: being unemployed (1) and employed, retired, at military service, stay-at-home partner, manager and other categories (0).

Living area: information about whether the respondent lives in an urban area (centre or outskirts of a city, 1) or in a rural area (isolated or rural village, 0).

Longstanding disability: related to having problems in daily living tasks due to illness, disability, infirmity or mental health problem. If respondents have this type of problem (both a lot or to some extent) they are coded with the number 1, if not, 0.

Education: Although the original variable includes information about 20 categories, our aim was to keep the analysis as simple as possible and so I reduced the categories down to five: no education (1), primary education (2), high school education (3), professional education (4) and university education (5), and then created dummy variables for each.

Controlling for differences in the web, mail and telephone samples

In this chapter I use coarsened exact matching in order to make the different modes' samples as similar as possible. The coarsened exact matching is based on the auxiliary variables described in the previous chapter: nationality (Swiss or another country's nationality), age, whether the respondent has a partner, urbanisation and use of the Internet. In this case, instead of matching the samples of two modes, the three samples are matched according to the set of chosen covariates where differences were found between modes but that are not expected to be affected by mode effects on measurement.

The sample balance test implemented after coarsened exact matching indicated that the samples are more balanced after the matching. The command "imb" in Stata indicates that the overall imbalance has decreased after the matching (multivariate L1 distance: 0.27, compared to 0.30 before the matching). The matched and unmatched units can be seen in the following table:

Table 48. Matched and unmatched units after CEM

	Web	Mail	Telephone
All	364	457	351
Matched	312	419	280
Unmatched	52	38	71

7.4. Results

The results are presented in three steps for each of the four models. First of all, I show the distribution of happiness, social trust, and job satisfaction for the mail, web and telephone modes, and then continue by examining the results for each of the three regression models and whether there are differences between the population subgroups of interest. In the last step, I show the post-test results that compare the regression results for the different modes of data collection.

7.4.1. Model 1. Predictors of happiness I: life events

Figure 13 shows the distribution for the measures of happiness.

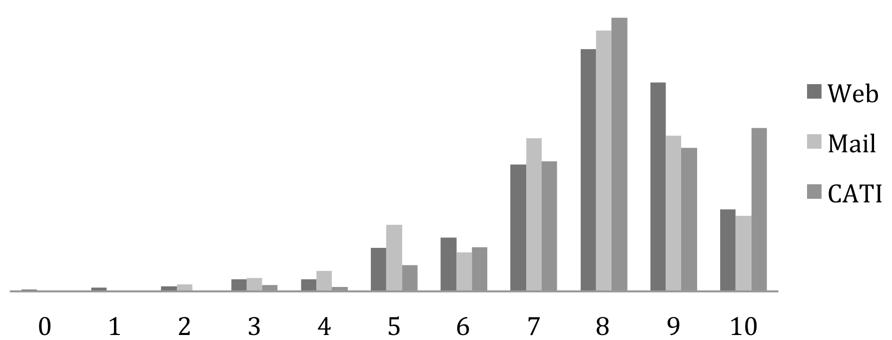


Figure 13. Distribution of happiness, by mode of data collection

None of the telephone respondents chose any of the three first most negative categories, illustrating how lower levels of happiness are less likely to be reported in this mode compared to the self-completion modes. On the other hand, self-completion respondents selected the middle categories more often.

Regression results show the effect the accumulation of negative events during respondents' life course has on happiness varies for web, mail and telephone, as can be seen in the table 49. In addition, the variability explained by the model is low for every sample: for web and for mail it is 2% and telephone 1%.

Table 49 reveals that the influence that important negative events have on happiness varies depending on which mode was examined. The magnitude of the coefficient was different for each mode: having experienced one negative event decreases the level of happiness particularly for the web and mail respondents ($p < 0.001$), but the effect increases for the second and third events. The effect of the third event is the strongest when comparing the odds ratios. In the case of the results for the telephone sample, however, only the proportional odds ratio of comparing respondents that have gone through a second event showed a significant negative effect ($p < 0.001$): neither the first nor the third event have a negative effect on happiness. Although the results change slightly when introducing the control for socio-demographic characteristics, regression coefficients are still different for the different modes' samples, and so are their attached significance levels. Looking at the results for the web sample, it is possible to see how the significance levels change, but the odds ratios remain very similar. In fact, the relationship between negative events and happiness is stronger for the web and mail samples than for telephone. Moreover, although the direction of the effect is the same, the negative events have no effect in happiness after adjusting for multiple testing.

Table 49. Regression results of the happiness model I

Happiness	Web (n=451)			Mail (n=351)			Telephone (n=364)		
Life events (Base = 0)	OR	S.E.	Sig.	OR	S.E.	Sig.	OR	S.E.	Sig.
1 negative event	0.58	0.11	**	0.52	0.12	**	0.69	0.16	
2 negative event	0.20	0.07	***	0.38	0.14	**	0.36	0.14	
3 negative events	0.13	0.08	**	0.08	0.05	***	0.58	0.36	
R2[4]		0.02	***		0.02	***		0.01	
After CEM	(n=457)			(n=313)			(n=323)		
1 negative event	0.60	0.12	*	0.46	0.13	**	0.69	0.18	
2 negative event	0.21	0.07	***	0.45	0.17	*	0.37	0.17	(*)
3 negative events	0.13	0.13	*	0.04	0.03	***	0.95	1.21	

*** p<0.001, **p<0.01, *p<0.05, †p<.10

Previous research on the literature of subjective wellbeing looked at the moderator effect of age and economic situation when examining the relationship between stressful life events and happiness. Looking at the regression coefficients for the interactions between negative events and being respondents older than 65, and between negative events and having a low income, it is possible to observe some differences depending on mode (see table below). The additive effect of the second negative event on the category 4 is an addition of 0.24 for younger than 65 respondents, and 0.11 for older than 65 respondents ($p < 0.05$). This means that respondents having suffered two negative events are more likely to choose category 4, for both older and younger respondents, but with older more likely to choose this category, showing a very different dynamic to the results for the web sample. The

[4] R2 for ordered logit not available for complex survey design analysis (after coarsened exact matching)

interaction effect between income and event on happiness was also significant, and when looking at the effect of one, two and three events.

Table 50. Regression results of the model I for happiness, including interactions age-events and income-events

Happiness	Web (n=457)			Mail (n=313)			Telephone (n=323)		
	O.R.	S.E.	sig	O.R.	S.E.	sig	O.R.	S.E.	sig
No event (base)									
1 event	0.52	0.11	**	0.27	0.13	**	0.89	0.33	
2 events	0.13	0.07	***	0.04	0.04	**	0.29	0.18	(*)
3 events	0.13	0.08	*	0.39	0.13	**	0.35	0.32	
Old respondent	0.78	0.26		1.46	0.70		1.55	0.61	
1 event * old	3.12	1.87	†	0.90	0.60		1.04	0.66	
2 events * old	3.67	3.41		0.12	0.16	†	1.30	1.20	
3 events * old	1.10	1.60		2.09	3.23		0.15	0.31	
Low income	0.97	0.38		0.17	0.10	**	1.32	0.44	
1 event * low income	0.40	0.25		2.17	1.66		0.58	0.32	**
2 events * low income	1.49	1.24		16.05	22.69	(*)	1.25	1.12	**
3 events * low income	0.51	0.95		6.89	11.00		6.51	10.68	**
R^2	0.03	***		0.05	**		0.02		

*** p<0.001, **p<0.01, *p<0.05, †p<.10

Both types of respondent, if having reported one negative life event, are more likely to choose category 6 than a higher level of wellbeing than those that did not report negative events, but this tendency is stronger for low-income respondents (0.21) than for the rest. Having gone through a negative event also has a negative impact on choosing the category 9, but again the effect is bigger for those on low incomes. Results were also significant when looking at the marginal effects of the interaction between income and reporting two negative events, and income and three negative events, for choosing categories 6 and 8. Choosing category 8 is less if three events are reported, but the effect is again stronger for low-income respondents (See annexe B for detailed information on the marginal results). Although there were interaction terms that had a significant effect on happiness in web and paper, it is not in a similar way, and there were no significant interaction effects, nor main effects, when examining the results for telephone.

7.4.2. Model 2. Predictors of happiness II: income

The second regression model predicting happiness (table 51, below) showed that reporting not having difficulties with the available household income is positively related with happiness in all modes of data collection. As was the case for the previous model predicting happiness, the amount of explained variance is low for every mode.

The signs of the estimated regression coefficients for the rest of the studied predictors are consistent across modes, but the strength and significance results show differences between the modes. Not having difficulties with the household income has a medium positive effect of 3.07 (sig = 0.000) in the web sample, and 2.42 in the telephone sample (sig < 0.01).

Table 51. Regression results of the happiness model II[5]

Happiness (After CEM)	Web (n=457)			Mail (n=313)			Telephone (n=323)		
	O.R.	S.E.	sig	O.R.	S.E.	sig	O.R.	S.E.	sig
Household income	1.03	0.04		1.10	0.06		0.97	0.04	
Unemployed	1.19	0.84		0.40	0.23		0.34	0.17	(*)
Has partner	1.06	0.19		2.18	0.61	**	1.20	0.26	
No income difficulties	3.07	0.78	***	1.67	0.53		2.42	0.69	**
R^2	0.02	*		0.04	*		0.02	*	

*** p<0.001, **p<0.01, *p<0.05, †p<.10
Note: results did not differ before CEM

[5] The inclusion of interaction effects between household income and household size was not significant.

Previous research found that the effect of marital status and income situation on happiness depends on whether male or female respondents are being considered. In the regression implemented for this study, no interaction effect was found between age and the predictors in any of the different modes' samples.

7.4.3. Model 3. Predictors of social trust

For social trust (see figure 14), the mid-point option (5) is the preferred one by all respondents.

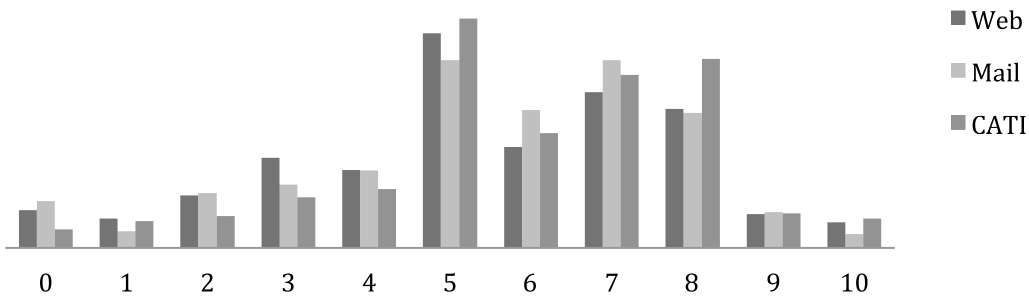


Figure 14. Distribution of responses to the social trust, measure by mode of data collection

For the regression coefficients of the predictors of social trust (table 52, below), there are differences between the modes and in the estimation of the pseudo R-squared, which is bigger for the web and telephone samples than for the mail sample. While happiness and life satisfaction are predictors of social trust in the models for telephone and web respondents, this is not the case for mail respondents, for which household income is the only predictor (OR = 1.14), although not very strong. This is also the case in the web sample. Some differences remain even after controlling for socio-demographic differences, although after adjusting the p-values using the Holm-Bonferroni method, household income remains the only predictor of social trust, although only for the web and mail samples.

Table 52. Regression results of the social trust model

Social Trust	Web	*n=457*		Mail	*n=351*		Telephone	*n=364*	
	O.R.	S.E.	sig	O.R.	S.E.	sig	O.R.	S.E.	sig
Happiness	1.24	0.11	*	1.11	0.12		1.25	0.13	(*)
Life satisfaction	1.24	0.10	*	1.07	0.11		1.27	0.12	**
Membership	1.05	0.21		1.30	0.34		1.25	0.30	
Household income	1.15	0.04	***	1.14	0.06	**	0.96	0.04	
Employed	0.81	0.15		0.60	0.16	†	1.51	0.34	
R^2	0.05	***		0.02	**		0.03	***	
After CEM		(n=457)			(n=313)			(n=323)	
Happiness	1.26	0.11	(*)	1.22	0.16		1.20	0.16	
Life satisfaction	1.23	0.11	(*)	0.98	0.12		1.36	0.19	(*)
Membership	1.03	0.20		1.20	0.34		1.44	0.48	
Household income	1.15	0.04	***	1.20	0.07	**	0.97	0.04	
Employed	0.78	0.15		0.56	0.16	(*)	1.31	0.34	

*** p<0.001, **p<0.01, *p<0.05, †p<.10

7.4.4. Model 4. Predictors of job satisfaction

The distribution of responses to the job satisfaction measure (see figure 17) is similar to the distribution for happiness: telephone respondents were more likely to choose the most extreme positive category (10) more often than respondents from the Web and mail samples, who opted for the middle categories.

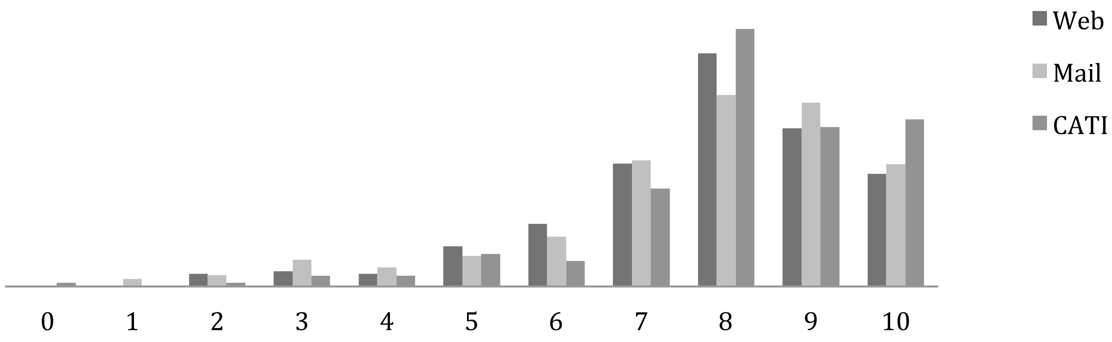


Figure 15. Distribution of responses to the job satisfaction measure, by mode of data collection

The amount of variance explained by the models was similar in the results of all modes of data collection (see table 53), R-squared values being higher than in the previous models for both of the self-completion modes of data collection.

Table 53. Regression results for the job satisfaction model

Job satisfaction	Web *n= 236*			Mail *n=130*			Telephone *n= 149*		
	O.R.	S.E.	sig	O.R.	S.E.	sig	O.R.	S.E.	sig
Works in public sector	0.60	0.19		0.68	0.25		0.80	0.21	
Work-life balance	1.42	0.13	***	1.72	0.17	***	1.72	0.12	***
Non-stressful job	3.09	0.62	***	3.46	0.70	***	2.80	0.40	***
Past period of unemployment	0.86	0.09	(**)	0.92	0.12		0.79	0.07	
Not likely to lose job	0.98	0.18		2.49	0.70	**	1.17	0.22	
Low level of education	0.78	0.13		0.65	0.13	(*)	0.88	0.13	
R^2	0.21	***		0.24	***			0.16	***
After CEM									
Works in public sector	0.64	0.22		0.61	0.23		0.79	0.20	
Work-life balance	1.44	0.15	***	1.56	0.26	(*)	1.72	0.13	**
Non-stressful job	3.14	0.75	***	3.80	0.73	***	2.78	0.60	***
Past period of unemployment	0.85	0.11	(*)	0.87	0.11		0.79	0.08	
Not likely to lose job	1.05	0.19		2.61	0.82	**	1.17	0.24	
Low level education	0.89	0.18		0.60	0.15	(*)	0.89	0.14	

*** p<0.001, **p<0.01, *p<0.05, †p<.10

In the case of telephone, although the predictors are similar to the ones for the web group, the percentage of variance explained by the model is smaller than in the other cases. The odds ratios are similar across modes, indicating that respondents with a higher level of satisfaction with the work-life balance and a smaller level of stress at work are linked to a higher level of job satisfaction. In addition, the expectation of losing one's job is also linked to job satisfaction, but only in the mail sample.

The addition of interaction effects in the regression model showed that the relationship between the predictors and the dependent variable was not moderated by age, gender or education in any of the regression models implemented for the different modes.

7.5. Discussion and conclusion

The study set out with the aim of assessing the importance of mode effects in the estimates of regression coefficients predicting different aspects of subjective wellbeing. A secondary aim was to investigate whether mode effects would also affect the comparison of different population subgroups that are commonly implemented by social science researchers, thus building on the analysis of differential mode effects across subgroups presented in chapter 6.

Previous studies evaluating the impact of mode in substantive analysis have produced inconsistent results concerning differences between modes. However, it was hypothesized that there would be some differences in the results obtained using different modes of data collection: if the way in which respondents answer survey questions depends on the mode of questionnaire administration, it may also be possible to find differences in predictions and relationships between the variables that are related to subjective wellbeing (Dolan & Kavetsos, 2012), which could cause researchers to reach different conclusions about which situations or characteristics are the drivers of happiness or job satisfaction. Given that comparisons and predictions of variables of interest are the *raison d'être* of most research into happiness and wellbeing, the possibility that estimates may vary significantly across surveys depending on how the data is collected was a particular cause for concern.

Relying on a methodological experiment, I showed four basic illustrations of what could happen if three social science researchers aiming to find out the drivers of happiness, social trust and satisfaction at work analysed cross-sectional data using the same statistical models and variables but with data that had been collected through different modes of administration: web, mail and telephone. The results of this study indicate that the presence of mode effects on both outcome and independent variables may provoke different conclusions for researchers. However, this depends on the variables of interest, as differences in the estimated coefficients vary across modes.

An interesting finding is that the results from the regressions did not change a lot after controlling for potential selection effects. For example, the impact of having income difficulties is similar for web and telephone, but not for mail; while the negative effect of having gone through negative life events is only significant in the case of the self-completion samples. The relationship between social trust and happiness and life satisfaction also varied across modes, but the results for the model predicting job satisfaction were quite similar in all cases: work-life balance and the level of stress are important predictors in web, mail and telephone.

This may indicate the already discussed topic of how it is possible that using socio-demographic measures is not enough to control for differences in who responds, which means that differences in estimates across modes remain due to who responds, and not only how they respond. Along the same lines, it is not clear that the strongest differences occur between self-completion and interviewer-administered modes, even though we found measurement effects between these two types of mode of data collection. Though the first descriptive results are consistent with previous research (Dolan & Kavetsos, 2012), showing that telephone respondents present higher level of wellbeing than self-completion respondents, there is a different dynamic of mode effects for each dependent variable. While telephone respondents choose the most extreme positive value to a higher extent than self-completion modes, this tendency is not as high when measuring social trust. Thus, differences may be due more to the composition of the samples than to actual measurement bias, but further research is needed in this respect.

Consistent with the literature, this research found that using measures that appeared to be affected by mode can also have repercussions for the results of commonly used regression models of subjective wellbeing, and particularly in the conclusions one would arrive at comparing different groups of the population. These results are in agreement with those obtained by Holford and Pudney (2015), Conti and Pudney (2011), Dolan and Kavetsos (2012), and Sarracino and his colleagues (2017).

However, the possible interference of other sources of error besides mode of data collection cannot be ruled out, and it is important to note that differences can also be unrelated to mode of data collection. In addition, these models aimed to illustrate what could happen when the simple regression model is replicated across modes of data collection, but social science researchers often use more complex tools by which results could be different to the ones presented here.

Returning to the question posed at the beginning of this study, it is now possible to state that mode effects are important from the standpoint of substantive research as they can influence the way we analyse and understand data on wellbeing and play an important part in the replication (and replicability) of quantitative research. The second aim of this study was to investigate the effect that mode could have in the comparison of regression coefficient between different groups of the population, and I was able to identify differences depending on sex, age, education, and economic situation.

There are different situations in which mode can affect substantive research: when comparisons are made across surveys conducted in different modes of data collection, when comparisons are made across groups in different modes of data collection and when conclusions are drawn from substantive research and differ depending on the mode of data collection. The findings from this study indicate that mode effects can have repercussions in the analyses of uni-dimensional measures of subjective wellbeing data coming from different survey modes, which could potentially render them incomparable. In the next empirical study, I use multi-dimensional measures of different aspects of subjective wellbeing to further investigate the impact of mode effects on measurement at the item level on the relationship between variables and draw conclusions about the comparability of wellbeing data collected using different modes of data collection.

CHAPTER 8. THE IMPACT OF MODE ON THE EQUIVALENCE OF MULTIDIMENSIONAL MEASURES OF WELLBEING

8.1. Introduction

In the previous chapter, I showed how mode effects on measurement affect statistical analyses results, even when controlling for selection differences. Although the relationship between job satisfaction and its predictors appeared to be consistent across modes, results from the happiness and social trust regression models differed between web, paper and telephone. Current analysis of subjective wellbeing require far more complex analyses, based on multidimensional concepts (Oris, Roberts, Joye, & Ernst Stähli, 2016).

In this chapter I focus on multivariate analysis and present results based on multidimensional measures of subjective wellbeing. I analyse the underlying dimensions of wellbeing across different modes of data collection. To do this, I examine whether the latent measures are equivalent in web, paper and telephone. I investigate the potential influence of data collection method on the parameter estimates of two substantive measures: general subjective wellbeing and wellbeing at work. In particular, the aim is to find out whether the question-answer process –

which differs for each mode of data collection – influences the way in which the respondent interprets the question and the repercussions this has on the given response (Hox et al., 2015). The aim of this study is to answer the question: are multidimensional subjective wellbeing measures comparable across modes of data collection? I continue focusing on the problem of measurement bias and look into the potential effect of mode of data collection on the relationships between the quality of life observed measures.

The first part of the chapter examines the work that has been done previously on the impact of mode effects on multivariate analyses and present the way in which the studies were implemented and the results obtained (Hox et al., 2015; Vannieuwenhuyze, 2015). Even though there are numerous studies that compare univariate distributions results from different modes, there is an identified need for researching the impact that mixing modes has on the estimates of the relationships between the variables (Hox et al., 2015), even though most studies do not present conclusive results as yet. In the review of the literature section, I also explain how subjective wellbeing is a multi-dimensional concept that lends itself particularly well to the type of multivariate analysis applied in this study, and how it is based on the findings of the European Social Survey research on wellbeing (European Social Survey, 2015).

In the methodological section, I explain how I implement multivariate analysis that allows taking a deeper look into the measurement equivalence or invariance of subjective wellbeing measures between different modes of data collection, to shed light on whether data that comes from different mode sources are comparable. I implement a multigroup confirmatory factor analysis (MCFA) to compare the subjective wellbeing and wellbeing at work measurement models across telephone, paper and web. Following the description of the methodological approach, I include the description of the co-variance matrices for the different modes and focus on the analysis of the level of measurement equivalence observed. The chapter finishes with the discussion of the results and the limitations faced in this chapter.

8.2. Theoretical framework

8.2.1. Measurement equivalence across modes

The literature on mode effects has highlighted the way in which the question-response process differs for each type of data collection design and how it may affect survey estimates of subjective wellbeing. The extent and sources of measurement error differ depending on survey mode. For instance, self-completed surveys are normally associated to less social desirability and more respondent openness than interviewer-based surveys, especially when the focus of the questions is on sensitive topics (Kaminska & Foulsham, 2013). However, self-completed questionnaires also involve different processing of the questions and skills from respondents that can impact responses (Kreuter, Presser & Tourangeau, 2008).

Much of the literature on mode effects has investigated how such measurement error is often confounded with nonresponse error (see chapter 1) and how this situation may result in different survey estimates depending on which mode of data collection was used. This presents a problem when data comes from a mixed mode design (de Leeuw, Hox & Scherpenzeel, 2010), complicating the analysis of data that has been obtained using different modes of data collection, and compromising the validity of the results obtained, as it can be difficult to disentangle and identify the different effects.

The majority of the examples of research implemented by survey methodologists on the topic of wellbeing involve analysing individual variables (Conti & Pudney, 2011; Holford & Pudney, 2015), and studies examining mode effects in multi-dimensional concepts are rare, for example looking at measurement models comparing different modes of data collection and the level of measurement equivalence. In particular, this type of analysis while controlling for the different sample compositions has not been common (Hox et al., 2015).

8.2.2. Measurement equivalence across mode of data collection

Information on human values, attitudes or behaviours is often gathered using multiple questions that aim to measure complex underlying measures. Such measures are often included in surveys so that they can be later compared across population groups, countries or time periods, but in order to be able to make such comparisons in a successful way it is important that the measurement structures of such latent or underlying measures are the same independently of the group being examined with the same survey items being stable. This is important for analyses that compare means from factor scores, or that examine the relationship between the different items that measure the latent concept. The analysis of measurement invariance has been traditionally used to establish equivalence across surveys that were implemented in different countries in order to make sure that measurements are not different across cultures (de Leeuw et al., 1996; Hox et al., 2015; Vandenberg & Lance, 2000). The idea behind this is that comparing different groups which have different characteristics and responding styles is risky if measurement invariance across them is not established (Vandenberg & Lance, 2000). In spite of this, many studies that implement comparisons across groups – whether countries or modes of data collection

– rarely check the level of equivalence. Despite its traditional use, using measurement invariance comparisons across groups has been found to be adequate for studying the effect of mode of data collection.

8.2.3. Mode effects in multivariate statistics

Research to date that has focused on examining the relationship between multiple variables across modes of data collection shows different results. In some cases, mode appears to influence the relationship between variables in multivariate models, but in other cases it only has an effect in univariate estimations. So far, results have fallen in one of two rival hypotheses (de Leeuw et al., 1996; Vannieuwenhuyze, 2015):

Firstly, there is the 'form-resistant correlation hypothesis'. Its main idea is that univariate statistics may be significantly different in different modes of data collection: this would be the case for point estimates such as the mean of happiness. However, it argues that multivariate statistics, including covariances for the different dimensions of subjective wellbeing – such as happiness and life satisfaction – are consistent across modes, showing no mode effects (de Leeuw et al., 1996).

Differential mode effects, depending on whether we look at univariate or multivariate statistics, are due to the fact that, while univariate distributions reflect change of a specific variable on the x- or y-axis, the shape of a distribution with two or more variables would remain stable.

Conversely, there is the argument on which the alternative hypothesis is based, that if univariate statistics are affected by mode effects, multivariate statistics based on '*higher order moments*' can be even more sensitive to mode, with multivariate analysis results being affected even in a stronger way than univariate statistics. For this reason, the concept of measurement invariance is useful to compare measurement models of multidimensional concepts across modes.

A number of authors have considered the effects of survey design in measurement models. To date, several studies have used confirmatory factor analysis to compare measurement models across groups, for example, countries, or modes of data collection, to determine whether they are generalizable across such groups. Evaluating the comparability of the models is often studied by looking at the level of measurement invariance. This evaluation consists of examining the degree of the equivalence of a latent measure based on factor scores from questionnaire items (Martin & Lynn, 2011). It is possible to identify four main levels of measurement invariance: configural, metric, scalar and full invariance across groups. In addition, there is approximate measurement invariance, which is a more flexible approach to establish measurement equivalence between the groups (Martin & Lynn, 2011). I will explain these types of invariance in depth in the next section. However, it is possible to say that the objective of being able to compare different groups is to obtain the highest level of invariance, or equivalence, possible.

Thus far, several studies found measurement differences depending on mode of data collection. However, the research to date has not been able to establish the effect of mode in multivariate statistical analyses. In the next section, I present some of the findings from previous research on the topic of mode effects in multivariate analyses.

8.2.4. Measurement invariance across modes of data collection

De Leeuw and her colleagues (1996) were some of the first researchers to evaluate measurement invariance across modes of data collection. Their work consisted of an examination of data on the topic of wellbeing and considered a general latent subjective wellbeing measure and another latent, more specific, measure of loneliness. The analysis is based on data from mail, face-to-face and telephone modes. Results showed non-invariance between the self-completion mode and the interviewer-based modes, offering not very promising results for the implementation of mixed-mode surveys.

Other studies have concluded that, depending on the latent measures of interest and the mode comparisons, measurement invariance can be achieved across modes of data collection. Revilla (2010) reports that the equivalence of scales on social trust, media and political trust vary depending on the scale being tested, and discusses results that range from finding measurement equivalence to small differences when examining measures on different types satisfaction items.

Similarly, Martin and Lynn (2011), Heerwegh and Loosveldt (2011), Hox and his colleagues (2015), and Cernat (2015) found mixed evidence about the measurement equivalence across modes. In the following lines I will present the main results these researchers obtained and briefly describe how they implemented their analysis.

Martin and Lynn (2011) examined scale equivalence for the latent variables social trust, political trust, political efficacy, attitude to immigrants, attitude to immigration, religious involvement and the Schwartz Human Values scale performing multi-group confirmatory factor analysis and testing for the different levels of measurement invariance. Their data came from the European Social Survey in the

Netherlands that had been collected using face-to-face, web, and telephone modes. They were able to establish that most latent measures are equivalent across modes, achieving a level of scalar invariance when comparing the groups, except when examining the human values and the attitude towards immigrants scales, which only reaches configural invariance, showing that comparisons across telephone, web and face-to-face can be complicated. However, they point out that not reaching the scalar invariance level could be due to the sample size. Looking at alternative indicator of model fit, Martin and Lynn suggest that the models have an acceptable fit.

That same year, Heerwegh and Loosveldt (2011) established scalar invariance across telephone and self-completion modes (web and mail), while controlling for socio-demographic differences across the sample. They identified social desirability bias as the cause of some differences on the telephone survey sample (Heerwegh & Loosveldt, 2011). Later on, Klausch, Hox and Schouten (2013), in their comparison of face-to-face, telephone, paper and web based on data from the Dutch Crime Victimization Survey (CVS), did not find complete equivalence for measures on different aspects of traffic, policing and police obedience between the interviewer-based modes and the self-completed survey modes. In their work, they take into account the ordinality of the variables used that helps localize measurement effects in ordinal data and also selection effects, focusing only on the measurement effects in the measurement models by using a propensity score adjustment.

Hox and his colleagues (2015) also followed the same type of analysis, which applied to 14 scales of social attitudes on parenthood, loneliness, life satisfaction and partnership conflicts and housework, they further control for possible causes of invariance including propensity scores that account for both socio-demographic characteristics – in particular sex, age, education and urbanization – of respondents and responses from a previous face-to-face wave. Their data came from web, face-to-face and telephone and results showed partial equivalence of the models across modes, although the equivalence level improved when there was a control of the socio-demographic differences. After including the control for the response from the previous wave the level of equivalence was even higher, indicating scalar or metric invariance for most of the scales examined. Their results, therefore, indicate differences according to the measurement model they looked at, to the point that in one scale that measured activities with children, the propensity score accounting for socio-demographic differences did not help to establish a higher level of equivalence. The authors argue that measurement equivalence studies implemented in the last decade confirm configural equivalence for the social science scales and suggest that, for scales with more than 4 to 5 variables, it may be possible to improve the level of equivalence by dropping some items that do not fit the model well. They are well established scales that had already passed validity and reliability tests (Hox et al., 2015).

In a study in which the focus was put on the measurement invariance across modes and time points, Cernat (2015) compared different survey designs based on the type of data collection, examines a scale measuring physical and mental health known as SF12 and tests its equivalence across the modes telephone and face-to-face, as well as across four waves of the Understanding Society Innovation Panel. Evaluating the different levels of measurement equivalence, Cernat established that for the first wave of the survey the measurement model of health is completely equivalent between telephone and face-to-face modes. However, the result changes when looking at the following waves: in waves 2 and 3 scalar invariance does not hold unless the measurement model is modified. As Cernat points out, differences can be due to selection differences, measurement differences or an interaction between both of them, but this could not be further investigated due to the research design. Finally, complete equivalence was again established across modes for wave 4.

Overall, there seems to be some evidence to indicate that mode can also have an impact on multivariate research. Particularly, those studies that looked at measures related to satisfaction and wellbeing found differences across modes, such as those from de Leeuw (1996) and Revilla (2010).

In view of everything that has been mentioned so far, there remain several aspects of measurement invariance across modes about which relatively little is known, particularly when the object of study is subjective wellbeing.

In the next section, I will present some information about the different aspects of subjective wellbeing that make up the multidimensional latent measures of general subjective wellbeing.

8.2.5. Subjective wellbeing as a multi-dimensional concept

In chapter 1, I showed how subjective wellbeing measurement involves non-observed, latent attitudes, often measured through attitudinal questions that ask the respondent to choose between positive and negative dimensions on how they feel about something, for example the level life satisfaction (Alwin & Krosnick, 1991). Subjective wellbeing is, therefore, a complex multi-dimensional concept that can be studied from different approaches. Huppert & So (2013: 843) developed a new inclusive framework of wellbeing that takes into account the following features: feeling accomplished, emotionally stable, engaged, valuable, optimistic, absence of negative emotions, having supporting relationships, self-esteem and vitality. The following graph (figure 16), created by Huppert and her colleagues (Huppert et al.,

2013) illustrates the different components of subjective wellbeing. Within the general concept of wellbeing, it is possible to identify a series of elements or dimensions that correspond to different aspects of life (Michaelson et al., 2009). Following this idea of the multi-dimensional subjective wellbeing, we use the classification presented by the European Social Survey:

Evaluative wellbeing
- Overall estimations of happiness and life satisfaction

Emotional wellbeing
- Abscence of negative feelings such as anxiety or depression

Positive functioning
- Feeling like one is "doing well"
- Feeling of accomplisment , autonomy, learning

Vitality
- Resting well, energy

Community wellbeing
- Trust in other people (e.g. neighbours)

Supportive relationships
- Trust and belonging, social exchange, sources of support

Figure 16. Components of subjective wellbeing

8.2.6. Work-related wellbeing as a multi-dimensional concept

Wellbeing at work can be summarized as the negative or positive evaluation that people make of their job (Weiss & Merlo, 2015), or how much people like their job (Millán, Hessels, Thurik, & Aguado, 2013). It is an additional dimension of an individual's wellbeing that is often studied as a separate measure, often through the item job satisfaction. However, it is in fact a combination of different dimensions that compose a more complex measure (Guest, 2002; Martínez-Martí & Ruch, 2016). Job satisfaction, or wellbeing at work, is also often included in research that studies its link with general wellbeing, or happiness (Martínez-Martí & Ruch, 2016). One-dimensional measures that do not reflect the heterogeneity of wellbeing at work can suffer from shortcomings because, as was the case with other wellbeing measures, different respondents may evaluate their satisfaction based on different indicators. This can make comparisons difficult between different types of workers, such as self-employed and employees (Millán et al., 2013; Muñoz de Bustillo & Fernández Macías, 2005).

There are different dimensions that measure wellbeing at work. Millán and his colleagues (2013) looked at job satisfaction depending on the type of work and job security. Other researchers, Martínez-Martí and Ruch (2016) used 5 observed items measuring satisfaction in different work domains: supervisor behaviors, job security, salary, working conditions and relationships with colleagues. Other researchers (Rode, 2005; Van Horn, Taris, Schaufeli & Schreurs, 2004) used a larger number of items to look the multiple dimensions of wellbeing by having various latent measures that at the same time compose the overall wellbeing at work. Such unobserved measures were, for example, “evaluative judgments about jobs, affective experiences at work, and beliefs about jobs” (Weiss, 2002, p. 173). Van Horn and colleagues included dimensions on enthusiasm and emotional exhaustion (which are related to emotional resources), job satisfaction, or organizational commitment; and complement them with measures on behavioral and cognitive dimension, as suggested by Brief and Weiss (2002). Items on autonomy, aspiration or professional competence, such as seeking challenges, or how effectively they deal with problems, were included.

8.2.7. Research questions

Does survey mode influence the relationship between different dimensions of subjective wellbeing? That is, are multi-dimensional models on subjective wellbeing equivalent across modes?

Individual items of subjective wellbeing appeared to be affected by both selection and measurement mode effects. However, previous studies have shown that mode effects in individual variables do not always have an effect on multivariate statistics, as long as the relationship between different items that measure an underlying concept is equal across modes. However, research that has specifically looked at measures on wellbeing has shown that complete equivalence is not achieved. Based on these findings, I expect to find from metric to scalar measurement equivalence across web, mail and telephone.

8.3. Methods

To investigate the effect of mode of data collection on the relationships between observed items that measure subjective wellbeing, I use two confirmatory factor analysis models: one for general subjective wellbeing, and one for wellbeing at work. The focus of the chapter is to be able to tell whether these two measurement models are equivalent across modes or not, and for that reason I implement a multi-group confirmatory factor analysis that is able to indicate whether differences observed across the modes are significant.

8.3.1. Data

The analysis for the research question involves respondents to the main survey questionnaire, which excludes respondents to the non-response questionnaire and the reserve respondents. The statistical analyses implemented involve respondents that answered through the mode they were assigned to in the first place, as was the case for the previous chapters, and only respondents with a listed telephone number in order to study samples with similar sample compositions.

For the final part of the analysis, only those respondents that have a paid job are selected. Their sample sizes by mode are as follows: there are 457 web respondents, 351 paper respondents, and 364 telephone respondents.

8.3.2. Variables

The subjective wellbeing variables used for this chapter were chosen based on the measurement model provided by the European Social Survey presented above, and also the items that measure different aspects of job satisfaction. The variables are:

Table 54. Subjective wellbeing items

Question	Categories
Taking all things together, how happy would you say you are? (Very unhappy- very happy)	0-10
All things considered, how satisfied are you with life as a whole nowadays? (Very unsatisfied-very satisfied)	0-10
Generally speaking, would you say that most people can be trusted, or that you can't be too careful in dealing with people? (You can't be too careful- most people can be trusted)	0-10
How much time during the past week have you felt depressed? (None or almost none of the time - all or almost all of the time)	1-4
How much of the time during the past week has your sleep been restless? (None or almost none of the time - all or almost all of the time)	1-4
Most days I feel a sense of accomplishment from what I do (Agree strongly – disagree strongly)	1-5
I'm always optimistic about my future (Agree strongly – disagree strongly)	1-5
In general I feel very positive about myself (Agree strongly – disagree strongly)	1-5
In the last month, how often have you felt confident about your ability to handle your personal problems? (Never- very often)	1-4
To what extent do you get support from your close ones if needed? (Not at all - completely)	0-6
To what extent do you give support to your close ones if needed? (Not at all - completely)	0-6
How much of the time do you find your job interesting? (None or almost none of the time - all or almost all of the time)	0-6
How much of the time do you find do you find your job stressful? (None or almost none of the time - all or almost all of the time)	0-6
How likely would you say it is that you will become unemployed in the next 12 months? (Very likely – not likely at all)	1-5
All things considered, how satisfied are you with your present job? (Very unsatisfied-very satisfied)	0-10

8.3.3. *Analytical approach*

The first measurement model consists of a structural equation model about the dimensions of subjective wellbeing. It is derived from the European Social Survey (2015) description on dimensions of subjective wellbeing and is adapted to the available measures. The dimensions are evaluative wellbeing, emotional wellbeing, positive functioning, community wellbeing and having supportive relationships. The main difference from the European Social Survey model is the dimension vitality (for example, feeling energetic) due to the lack of information on this aspect.

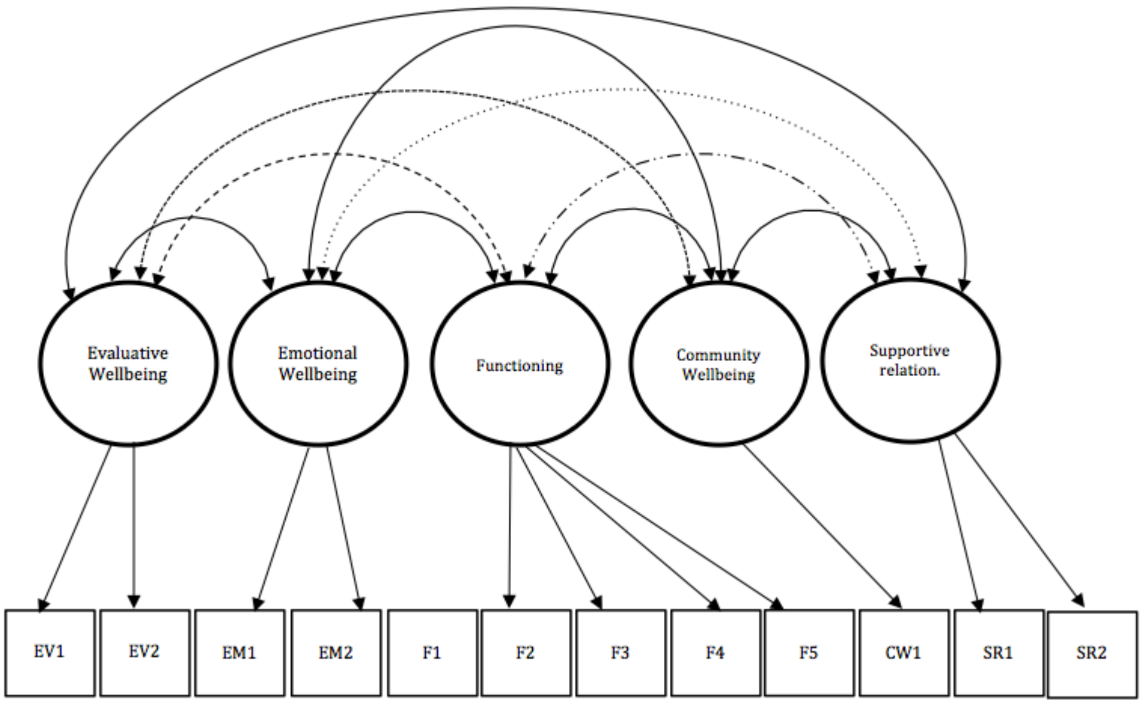


Figure 17. Multidimensional subjective wellbeing

It is possible to distinguish five dimensions, which correspond to the wellbeing elements described in the section above: evaluative wellbeing, emotional wellbeing, positive functioning, vitality, community wellbeing, and supportive relationships.

Table 55. Correspondence between SWB dimensions and survey questions

Subjective wellbeing dimensions	Survey items
Evaluative wellbeing	Happiness Life satisfaction
Emotional wellbeing	Depression Anxiety
Functioning	Freedom Accomplishment Optimism Positivity Handle problems
Community wellbeing	Social trust
Supportive relationships	Get support Give support

The chosen model has, therefore, five latent factors and 12 observed variables. The most general dimension is measured by the widely used happiness and life satisfaction, both of them ranging from 0 (the least positive category) to 10 (the most positive category). The rest of the factors, except community wellbeing (measured by social trust, which ranges from 0 to 10, from negative to positive) are measured by ordinal variables as follows: emotional wellbeing is measured by absence of depression and anxiety; functioning includes feeling free to do what they want, sense of accomplishment, feeling optimistic, feeling positive and being able to handle problems in their lives. Lastly, having supportive relationships is measured by being able to get support from friends and family if needed and not feeling lonely.

The second measurement model corresponds to the multidimensional measure of wellbeing at work. On this occasion, there is only a latent variable for which the observed items are job satisfaction, job interest, job stress, and likelihood to become unemployed in 12 months. These observed items represent different dimensions of multidimensional job satisfaction or wellbeing at work which have been previously studied in the literature, although there were dimensions for which there were no observed items in the data used for this thesis. The items used for this chapter on the topic of wellbeing at work are: subjective evaluative wellbeing (job satisfaction), emotional and behavioural (how stressful and interesting the job is), and job security (likelihood of becoming unemployed in 12 months).

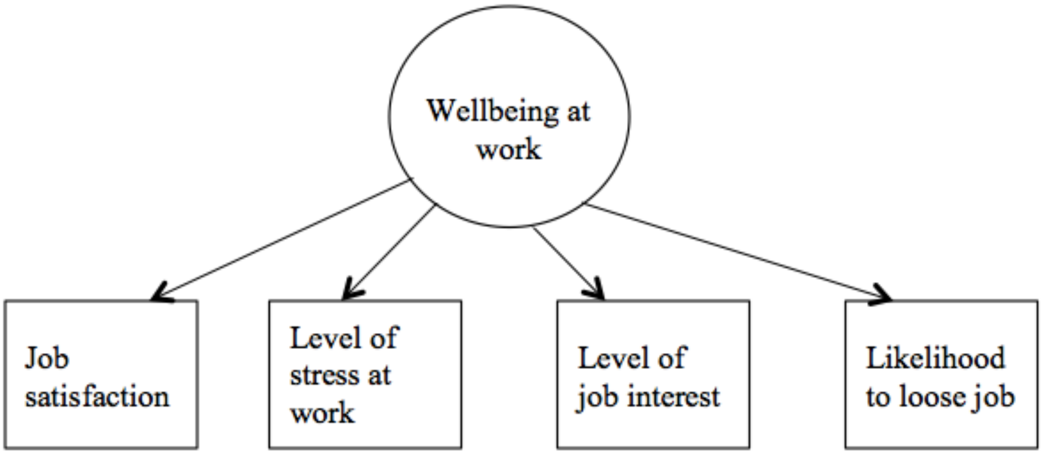


Figure 18 Multidimensional wellbeing at work

Establishing the baseline measurement model

To determine the baseline measurement model in which the rest of the confirmatory factor analysis is going to be based, I implement a confirmatory factor analysis and look at two different types of fit indicators: Root Mean Square Error of Approximation (RMSEA), which is an absolute fit measure which indicates how well the model fits theoretically, and the Comparative Fit Index (CFI), which compares the fit to the base model and takes into account sample size (Hooper, Coughlan & Mullen, 2008).

For the personal and social subjective wellbeing model, the first trial was implemented including all the variables that were asked in the mixed-mode experiment corresponding to the European Social Survey (2015) wellbeing dimensions (see table 56).

Table 56. Subjective wellbeing model fit

Model	X^2(df)	RMSEA	RMSEA 90% CI	CFI	TLI
a)	5697.973(91)	0.041*	[0.034, 0.048]*	0.989*	0.984*
b)	5310.503(66)	0.041 0.061*	[0.033, 0.049] [0.053, 0.068]*	0.964 0.992*	0.947

*Robust

This way, the first model (a) tested included two variables that did not play an important role in the subjective wellbeing factors of our model: the variable that measured vitality (feeling restless, with a loading of 0.22) and the variable that measured the number of intimate social relations one can rely on (loading of 0.23), which is related to the factor having positive relationships. Results from this first base model did not provide robust results, and therefore we chose model (b), for which I present both the robust and non-robust results, and that shows a decent fit (see table 44), as the indicator (RMSEA) is lower than 0.80, and very close to 0.060, which indicates a good fit (Hooper, Coughlan & Mullen, 2008). In addition, the Comparative Fit Index (CFI) indicates good fit, as it is higher than 0.95 (Hu & Bentler, 2009). The Tucker-Lewis Index (TLI) is a relative fit index indicating a good fit if the value is under 0.95 (Hu & Bentler, 1999)

The indicators RMSEA and CFI, however, indicate a poor fit for the wellbeing at work model (see table 45). Although the level CFI is close to 0.95, the indicator RMSEA is higher than 0.80 (0.123). Even though it is not an optimal fit, there were no better fitting models after adding or removing items on wellbeing at work. However, the first measurement model trial I tested included an additional item measuring work-life balance that made the measurement model fit worse (RMSEA = 133). For this reason, I do not include this item in the analysis.

Table 57. Wellbeing at work model fit

X^2(df)	RMSEA	RMSEA 90% CI	CFI	TLI
23.603 (2)	0.123	[0.081, 0.169]	0.929	0.787

Multigroup confirmatory analysis

The second step is to compare model fit across groups to be able to see whether there are differences in the way the different observed items fit the measurement models of subjective wellbeing and wellbeing at work. After this, I present the correlation matrices for the different modes of data collection and two models in order to identify potential differences in the relationship between the variables that measure the two latent measures examined here. After this, I proceeded to test the level of measurement equivalence. For each model, I followed the same analysis strategy: the multiple-group confirmatory factor analysis (MCFA) was the first step in order to determine measurement equivalence across the modes, as we compare the MCFA results for each mode increasing the equality constraints restrictions. This is, we increase the number of equivalence indicators (i.e. loadings, intercepts, variances) (Vandenberg & Lance, 2000) to see what level of equivalence can be established.

Testing for measurement equivalence is the core of this chapter. In spite of being used mainly before implementing cross-country and cross-cultural comparisons, it has previously been used to determine measurement invariance across-modes.

Specifically, the type of analysis is Multigroup Confirmatory Factor Analysis (MCFA), which is integrated in the multigroup Structural Equation Modelling types of analyses.

Once the base model for the CFA has been decided, a series of steps follow (Vandenberg & Lance, 2000) to help establish the level of invariance between the different groups, in this case, the modes of data collection. The measurement invariance tests have been described previously by researchers such as Hox and colleagues (2015), Vandenberg and Lance (2000) or Yong and Pearce (2013). There are three main types of invariance: configural, scalar and strict.

Configural invariance is the most basic form of invariance and describes the situation in which the questionnaire items measure the same concept: the loadings of the latent measure for each observed item must be close for the different groups being compared (Hirschfeld & Von Brachel, 2014). Overall, it shows that the measurement structure is similar across the group: there is the same number of underlying factors and would lead to similar conclusions independently of which group is being analyzed (Yong & Pearce, 2013). Weak or metric invariance assumes configural invariance and in addition, the loadings must not be statistically significantly different. This is important because it shows that even if our groups' data is biased (for example, systematic response biases), it would not be affected (Chen, 2008; Van De Vijver, 2011). In scalar invariance, item loadings and item intercepts have to be the same across groups while the model fit is similar to previous stages (Hirschfeld & Von Brachel, 2014; Martin & Lynn, 2011). Finally, strict invariance: on top of the previous requirements, the residual variances are also similar when comparing the groups (Wu, Li & Zumbo, 2007).

Although strong invariance would be desirable, there is the overall agreement that partial scalar equivalence is enough to compare the different groups (Hox et al., 2015; Martin & Lynn, 2011). One of the main theoretical difficulties in this type of study is to decide how similar data has to be in order to be equivalent (Cieciuch, Davidov, Schmidt, Algesheimer & Schwartz, 2014). While some authors claim that strict invariance is a must if the analysis plan involves mean score comparisons (Hirschfeld & Von Brachel, 2014; Wu et al., 2007), there also exists the argument that scalar measurement invariance is enough to establish comparisons across modes of data collection (Martin & Lynn, 2011).

It is possible to establish measurement invariance if the difference in CFI between the results obtained from the MCFA for each set of constraints is smaller than 0.01 (Cheung & Rensvold, 2002).

The first CFA, for personal and social subjective wellbeing, that I implement here tests a multigroup model in which only three out of the twelve variables used are continuous, having less than five categories. The rest of the variables are ordinal and therefore we use a special approach that takes this into account (Rhemtulla, Brosseau-Liard & Savalei, 2012), and use the weighted least squares means and variance adjusted (WLSMV) estimator to estimate parameters (Finney & DiStefano, 2013). This way, we obtain thresholds that are equivalent to the item loadings (Hirschfeld & Von Brachel, 2014). The same approach is used for the wellbeing at work model, although I do not take a different approach for ordinal measures, as all the items have 5 or more response categories.

8.4. Results

The purpose of this chapter was to know whether subjective wellbeing measures are equivalent across modes of data collection. To assess measurement equivalence between web, paper and telephone, a multigroup confirmatory factor analysis was implemented to study a model on the structure of wellbeing. To complement and illustrate the findings, I present the comparison of model fit between the modes, and the correlation matrices to ease the comprehension of the results obtained.

The first step of the analysis involved looking at the model fit for each mode of data collection. To compare the measurement model fit of the two measures of wellbeing, I compare the fit indicators RMSEA and CFI. The results obtained from the confirmatory factor analysis on personal and social wellbeing on each of the modes helps to get a first impression of the differences observed, in addition to an indication of whether the model fit is adequate.

8.4.1. The personal and social wellbeing model

Table 46 shows how the fit of the first measurement model with the included observed variables is good for all modes. Even though the RMSEA indexes for each mode are slightly different, they are all under 0.080. In addition, CFI levels are correct, all close to 0.95.

Table 58. Indicators of subjective wellbeing model fit across modes

Model	X^2(df)	RMSEA [90% CI]	CFI
All modes	234.458(45)	0.061 [0.053, 0.068]	0.964
Web	137.857(45)	0.068 [0.055, 0.081]	0.955
Paper	123.429(45)	0.071 [0.056, 0.086]	0.951
Telephone	83.105(45)	0.049[0.032, 0.065]	0.980

Only a small amount of differences between the modes were identified in the correlation matrices that show the relationship between the variables that conform the factors. I mention here a few differences that appear in the matrix, then continue on to looking at the test for equivalence. For example, the correlation between social trust and depression is weaker in the paper group (0.01) than in web and telephone groups (0.23 and 0.20 respectively), and the correlation between giving support to their close ones and level of anxiety is smaller and negative (-0.02) in the telephone group compared to the web and paper groups (0.08 and 0.15 respectively). Overall, it is possible to say that the correlations are somewhat smaller for the paper mode, but correlations between items that correspond to same factor appear to be very similar.

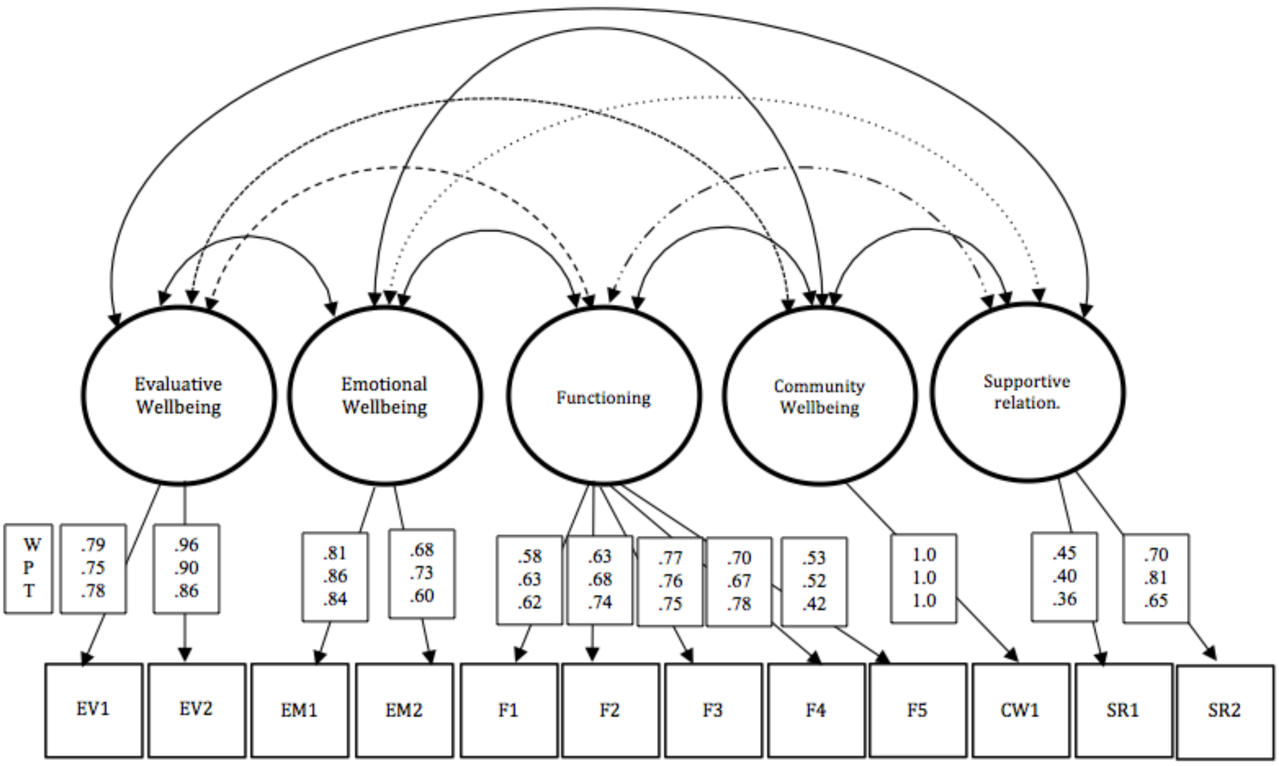


Figure 19. Results from CFA

The loadings on each factor from the different variables also have similar loading values in all the cases (see figure 21). As I will show later in the chapter, when I present the results for the multigroup analysis, the loadings were not statistically different across groups. The results from the wellbeing model show, therefore, that the observed variables which measure the underlying concept of subjective wellbeing behave similarly independently of which mode of data collection is being analysed. In spite of this, it is possible to see some differences on certain variables: when comparing the self-completion modes and telephone the loading of the variable which measures anxiety (M2) is different, although in the same direction as the two variables that measure having supportive relationships (receiving and getting support and not feeling lonely). As does the relationship between the different dimensions of subjective wellbeing (see table 59), which already appears to indicate that at least configural invariance will be established.

Table 59. Relationship between SWB dimensions

SWB dimensions	**All**	Web (n = 457)	Mail (n = 351)	Telephone (n = 364)
Evaluative wellbeing-Emotional wellbeing	0.64	0.63	0.62	0.66
Evaluative wellbeing-Functioning	0.66	0.67	0.64	0.67
Evaluative wellbeing-Community wellbeing	0.32	0.35	0.22	0.38
Evaluative wellbeing-Supportive relationship	0.79	0.80	0.67	0.95
Emotional wellbeing-Functioning	0.67	0.63	0.70	0.70
Emotional wellbeing-Community wellbeing	0.22	0.25	0.14	0.28
Emotional wellbeing-Supportive relationship	0.90	0.89	0.82	1.05
Functioning-Community wellbeing	0.18	0.22	0.11	0.21
Functioning-Supportive relationship	0.63	0.61	0.51	0.82
Community wellbeing-Supportive relationship	0.25	0.26	0.18	0.33

In spite of how similar the relationships between the different factors are, results show that the relationships are smaller for the paper groups than for the telephone and web, in particular, for the following combination of comparisons: evaluative wellbeing-community wellbeing, evaluative wellbeing-supportive relationship, emotional wellbeing-community wellbeing, emotional wellbeing-supportive relationship.

Tests on measurement invariance confirm that the structure of subjective wellbeing is partially equivalent between the different modes of data collection. Following the steps previously mentioned, we look at the different levels of equivalence that may allow comparisons across modes of data collection on the topic of subjective wellbeing.

Specifying different equality constraints between the different modes (Hirschfeld & Von Brachel, 2014) we get the different indicators we are going to look at for establishing invariance: difference in chi-square, degrees of freedom and significant test, and the difference in CFI between the models (see table 60 below). As a reminder, the test of configural equivalence consists of checking that the loading pattern is similar and there is the same number of factors for each mode, as I showed in the previous step's results. Secondly, there is the test of metric equivalence that is implemented by constraining factor loadings to be equal across the modes. Then, I then test of partial scalar invariance, which constraints the measurement intercepts and loadings to be equal across modes. Finally, the test of full scalar invariance (also known as strong invariance, or full uniqueness measurement invariance) allows variances and covariances between latent and observed scores to be different across groups and fixes the residual variances to be equal across groups (van de Schoot, Lugtig & Hox, 2012).

Table 60. Level of invariance between the modes of data collection

Invariance	X^2(df)	RMSEA	CFI	Change	Diff.
Configural	299.17(147)	0.052	0.958	-	-
Metric	314.41(163)	0.049	0.959	<0.01	No
Scalar	331.61(179)	0.047	0.958	<0.01	No
Strict	390.61(201)*	0.047	0.957	<0.01	No

After obtaining the configural invariance indexes, the metric invariance test shows that the factor loadings can be assumed to be equal as the difference between the two CFI is smaller than 0.01. The next model comparison steps show that measurement invariance is also achieved when constraining the loadings and the intercepts to be equal across the modes up to the scalar invariance: the chi-square test is not significantly different and the change in CFI is smaller than 0.01. The last comparison, testing for strong invariance, gives a slightly different result because, even though ΔCFI is smaller than 0.01, the test shows the lack of strong invariance as the output demonstrates that the chi-square is significantly different between the models.

8.4.2. The wellbeing at work model

Table 61 shows the results for those respondents who have a job, on the measure that indicates their level of wellbeing. What stands out in the table is that the model fit is very different for the telephone group compared to the fit of the baseline model with the pooled data from all the modes, the web group and the paper group. For telephone respondents, indicators revealed that the measurement model fits the data: the RMSEA is under 0.08 (0.061) and the CFI is close to 0.95, although a little smaller.

Table 61. Indicators of wellbeing at work model fit across modes

Model	X^2(df)	RMSEA	CFI
All modes	23.603(2)	0.123	0.929
Web	8.809(2)	0.137	0.892
Paper	123.429(2)	0.130	0.933
Telephone	3.526(2)	0.061	0.984

The correlation matrix (see table 62) display results about the relationship between the different items on job wellbeing. It is possible to appreciate some differences. For example, the correlation between finding their job interesting and the level of stress is stronger for the telephone mode (0.21) than for the mail and web respondents. However, the correlation between stress and likelihood of losing the job in the next 12 months is stronger for the self-completion groups than for the telephone group. The rest of the correlations are similar across groups, in some cases the relationship is very weak (interest level and unemployment) because they measure different observed factors of wellbeing at work. The strongest correlations can be found when looking at job satisfaction with the rest of the items.

Table 62. Correlation matrix between items of wellbeing at work

Dimension	Job satisfaction			Interest level			Stress level			Likelihood unemployment		
Mode	W	P	T	W	P	T	W	P	T	W	P	T
Job satisfaction	1	1	1									
Interest level	0.47	0.56	0.55	1	1	1						
Stress level	0.13	0.24	0.20	0.05	0.08	0.21	1	1	1			
Losing job	0.22	0.17	0.23	-0.04	-0.08	0.09	0.20	0.17	0.09	1	1	1

The loadings for the latent construct for wellbeing at work appear to vary a lot from mode to mode. The differences between web, paper, and telephone are present in every observed item. In fact, the value of the loadings for web and paper for the measure job satisfaction is very high possibly indicating there is some mirror and that the measurement model does not fit the data well. The results from the wellbeing at work model for telephone showed that the observed variables that explained a higher amount of the latent variable were job satisfaction and level of job interest, while stress and insecurity do not have a strong relationship with the latent measure. In spite of this, it is possible to see some similarities when comparing the different modes: job satisfaction is the observed measure with the highest loading value, while the loading of insecurity is different on certain variables: when comparing the self-completion modes and telephone the loading of the variable that measures anxiety (M2) is different, although in the same direction. In this case, not even configural invariance is guaranteed, as the loadings were differed greatly across modes.

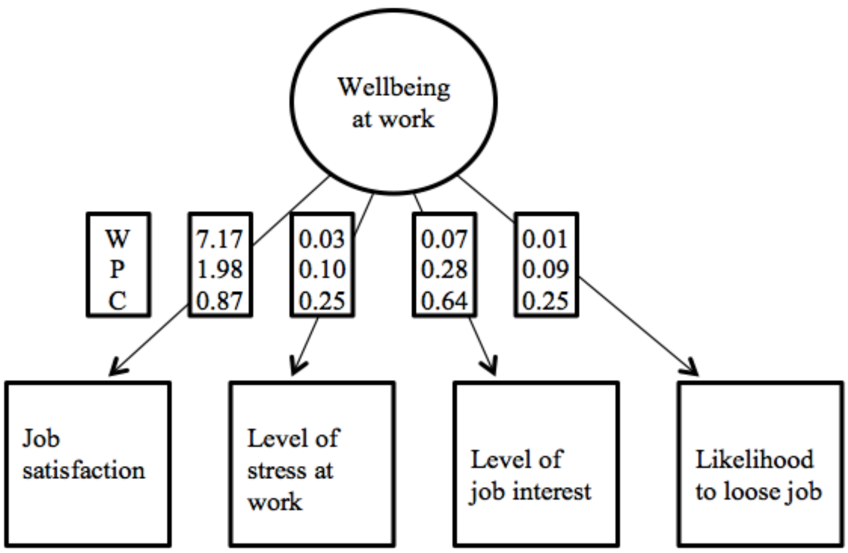


Figure 20. Measurement model of wellbeing at work

Tests on measurement invariance confirm that the structure of subjective wellbeing is partially equivalent between the different modes of data collection did not work properly, confirming that the measurement model did not fit the data for the web and paper groups.

8.5. Discussion and conclusion

This study set out with the aim of assessing the importance of mode in multivariate analysis of wellbeing measures. Several reports have shown that people respond differently to different modes of data collection, and that this impacts the level of measurement equivalence across modes, making comparisons difficult. In particular, and using confirmatory factor analysis, some researchers such as de Leeuw (1996), Martin and Lynn (2011), or Heewegh and Loosveldt (2011) found different levels of measurement invariance, or equivalence for the comparison of interview modes versus self-completion modes. Their results showed different results, with some finding significant differences and invariance between modes for wellbeing studies, and most of the other studies showing a high level of equivalence (metric or scalar).

In addition to these previous studies, in the previous chapters I had identified measurement and selection differences between the modes of data collection, particularly between the self-completion modes and interviewer-based modes. Furthermore, measurement effects were specifically found on widely used measures such as happiness and job satisfaction. This was an important result that lead to examine in closer detail whether the subjective wellbeing measurement structures were equivalent across modes.

In this chapter, therefore, the core idea was to identify the impact that such mode effects can have analysing the relationship between subjective wellbeing variables and their predictors: the aim was to be able to establish the level of measurement equivalence on the topic of subjective wellbeing

By implementing multigroup confirmatory factor analysis of subjective wellbeing it was possible check the construct validity of different measurement models and their equivalence when using multiple modes of data collection. However, this was only the case for the measurement model of personal and social wellbeing. On the one hand, this study confirms scalar invariance for the personal and social wellbeing measure, but it fails to identify a fitting measurement model for wellbeing at work, for which fit indicators only show optimal fit when looking at telephone respondents. However, this is an interesting finding in itself, showing that there is no measurement equivalence for this measurement structure.

One expected finding was the extent to which mode of data collection impacts subjective wellbeing. Even though results display a higher level of equivalence than the previous study on subjective wellbeing structure implemented by de Leeuw and colleagues (1996), this result supports more recent work being done on measurement invariance, even if the substantive research topic is different in our study (Cieciuch et al., 2014; Hox et al., 2015; Martin & Lynn, 2011; Vannieuwenhuyze & Loosveldt, 2012). Consistent with the literature, this research found that the latent measure of personal and social subjective wellbeing keeps a very similar structure independently of mode of data collection. However, the confirmatory factor analysis results did not establish the strictest level of invariance: even though the measurement structure, loading pattern and measurement intercepts are equivalent across modes, the results indicate that there are some differences in the way people respond, and that this may have consequences for substantive analysis.

Perhaps the most unexpected finding is the difference in results between the two measurement models examined in this chapter, although this could equally be due to the lack of a more complete set of observed items that examine wellbeing at work, or the smaller size of the samples. However, and consistently with previous results, it appears that results are highly dependent on the survey design and observed variables measuring subjective wellbeing that are available. Such results may encourage mode comparisons on subjective wellbeing, but it may not be the case for all dimensions, and it is essential that measurement equivalence be tested before comparing data that has been gathered using different modes. In addition to this, measurement invariance is essential if cross-mode comparisons are to be made and establishing scalar invariance is considered to be the minimum, but in order to be able to compare means of measures of subjective wellbeing composite scores, establishing full or strict equivalence would still be a requirement.

The lack of strict invariance could be resolved by future research in which it is possible to have a better control of respondents' characteristics, this is, to control more than using the propensity scores built based on the socio-demographic available – which did not make a difference in the results obtained – but using other types of control such as previous responses if researchers are working with longitudinal data.

Even though the analysis implemented in this chapter is limited in the sense that the analysis looks into only two different latent measures, it is still possible to partially answer the research questions posed at the start of this work process. Using a multigroup confirmatory factor analysis (MCFA) looking at telephone, paper and web, I determined that there is a level of scalar invariance across the examined modes for the personal and subjective wellbeing model. This indicates that a variety of multivariate analysis looking at the relationship between the variables can be implemented without fearing for mode measurements effects. However, researchers must be careful when using composite scores comparing means, for example. In the case of wellbeing at work, it is also strongly recommended to check the level of invariance before implementing other types of analysis with the multidimensional measure, and the level of invariance will very likely depend on the observed items to which the researcher has access.

Differences in fit for the wellbeing at work model may be an indication that previous studies have mainly used data from telephone or face-to-face surveys, and therefore developed a theoretical model that fits well such data. That, however, does not work when the data comes from self-completed questionnaires. This is something that would be important to look into in future studies, as the observed correlations between the observed variables may indicate that telephone survey participants respond in a different way to survey questions.

CHAPTER 9. DISCUSSION AND CONCLUSION

This thesis set out to investigate the effect that the mode of data collection has on measures of subjective wellbeing, with the aim of informing users of survey data on this topic about the possible repercussion of mode of data collection in their analyses and results. Although mode of data collection is just one of the different aspects of survey design that can impact the data obtained, it can greatly impact the way the response process (Cernat, 2015a). For this reason, numerous journal articles and books have looked at the impact of response mode in the quality of the data. However, while previous studies have noted the importance of studying the impact of mode of data collection in measures of subjective wellbeing (Pudney, 2010; Dolan & Kavetsos, 2012), there are a lack of conclusive results about the presence of mode effects in such measures and how much they matter in substantive research.

In a context in which many studies draw on mixed-mode survey designs, or use different sources of data that were collected using different modes, concerns about the comparability of responses across modes is a recurrent theme in survey methods literature (de Leeuw et al., 2010; Kaminska & Foulsham, 2013). Because differences in survey estimates related to mode can be due to differential respondent characteristics and to different ways of responding in each mode, knowing more about the extent and the type of mode effect causing the differences is essential before analysing data that come from different response modes. Indeed, the reason for mixed mode surveys is to take advantage of the selection effect, to improve the overall representativeness of the samples, but there is the drawback of having measurement differences that are hard to control.

A large amount of research has been carried out to investigate the best ways of disentangling such mode effects, with different levels of success (Tourangeau, 2017), the results of which find that it is difficult to predict which variables are going to be affected by mode and why (Martin & Lynn, 2011).

With this thesis, I aimed to contribute to the literature about the effect of mode in substantive research using a methodological experiment that was expressly designed for this purpose. I addressed a series of research questions that focus on different aspects of how mode can impact survey results based on the findings from previous studies on the topic of subjective wellbeing and survey methodology. In this closing chapter, I present the main findings for each question and how they relate to the existing research presented in the literature review.

To recap, the research set out to address the following three over-arching research questions:

RQ1. Do different modes of data collection differentially affect the quality of survey estimates of subjective wellbeing?

RQ2. Do mode effects on measurement affect all respondents equally?

RQ3. Do mode effects on measures of subjective wellbeing impact the results of substantive research into the predictors and correlates of subjective wellbeing measures?

9.1. Mode effects in measures of subjective wellbeing

In Study 1, I showed that not all types of respondent are equally likely to respond to web, mail and telephone surveys. Based on the previous literature in this field (Sakshaug et al., 2010; Vannieuwenhuyze et al., 2010), this was to be expected, and my results confirmed that there were differences in the composition of the samples according to age, nationality, and marital status, among others. These results were used as a base to construct the analytical approach, for which I aimed to separate selection from measurement effects by rendering the different modes' samples as similar as possible, in order to be able to isolate the effect related to how people respond to each mode. This is one of the most popular techniques used in the literature (Tourangeau, 2017) due to availability of socio-demographic data in surveys, it is also easier to implement than other methods and therefore more commonly used than other approaches by data users.

The findings from the same study indicated that some measures of wellbeing are indeed sensitive to mode, even after controlling for differences in the sample compositions between the modes. Mode was found to impact the outcome on wellbeing-related measures when comparing the telephone and the self-completion modes, but there was not a significant difference between responses in mail and web modes on any of the wellbeing measures analysed. Results indicated that 18 out of 27 measures of wellbeing suffer from statistically significant differences when comparing telephone and self-completion modes, after coarsened exact matching. The results after applying different approaches to control the mode effects on selection were the same for all the SWB variables, and there was not an obvious pattern related to a specific response format, although a higher proportion of questions offering eleven and five response alternatives (in the form of agree-disagree scales) were affected by measurement effects than questions with other response formats. Looking at the overall findings, it is possible to say that telephone respondents tended to report significantly higher levels of subjective wellbeing, and that this tendency has an impact on both means and distributions of responses across answer categories. This result complements those of a relatively large number of studies indicating that respondents tend to give more socially desirable answer in telephone interviews compared to other modes (e.g. Holbrook et al., 2003).

These results are consistent with the findings of previous research (e.g. Pudney, 2010), but they also indicated that socio-demographic controls are not particularly useful as a way to control for selection effects. They appear to have a relatively low capacity for doing this, which leaves some doubts about the source of estimate differences across modes of data collection. Investigating other alternatives, such as personality measures or available auxiliary data on the socio-economic position of respondents would likely improve the control of selection effects.

Although in this study, the differences between mail and web disappeared after the coarsened exact matching, most differences remained when the telephone mode was involved in the comparison. I conclude that these differences are due to mode effects on measurement (and the direction of the effects is consistent with this conclusion), but it is still possible that the mode effects that remained were related to differences between the respondents on unobserved variables.

9.2. Mode effects in sensitive open-ended measures

In Study 2, the focus of the analysis was on how mode affects respondents' answers to open-ended questions. The findings indicate that some aspects of responses to open-ended questions on the topic of important life events are also affected by mode of data collection. For the first question (of three) asking about life events significant differences were found in terms of item nonresponse between web and telephone and between paper and telephone, indicating that nonresponse is higher in the self-completion modes than in the telephone mode. Response length was, in all cases, longer in self-completion modes than in the telephone mode, although there may be differences due to the differential processing approach taken to record the answers in each mode. Finally, testing for differences in the more relevant question of the content of answers – the theme of the life events reported and their level of positivity – I showed there few differences in the reporting of some events depending on mode, however, some were reported as more positive than others depending on the mode of data collection. One interesting finding was that telephone respondents reported significantly more positive events than paper and web, which may indicate that, even if the events reported are the same, they may be interpreted in a different way when respondents have to evaluate how positive or negative their impact was.

These results partially support previous findings about the effect of mode in open-ended questions. On the one hand, having an interviewer asking the questions has been found to have a positive effect in obtaining a lower item-nonresponse rate. On the other hand, Schaefer and Dillman (1998) showed the presence of an interviewer might also have had an effect in the responses about the positivity of the event. The overall tendency in my results was to find no differences between mail and web, contradicting the results of research by Denscombe (2008), and a higher number of unanswered items in web surveys when compared to the telephone mode's results. However, this was not the case for all three of the open-ended questions analysed, and the item non-response was lower in the case of web respondents. The examination of response length, although a common technique to examine the quality of responses to open-ended questions, was not very informative, because the response differences that were found between the different modes could also be due to way in which the information was processed: web respondents typed the answers themselves on the computer, while responses to the paper and telephone questionnaire needed to be added to the dataset, which could have led to the reduction of the length of respondents' answers.

9.3. The interaction between respondents' characteristics and mode effects

In the study 3, I explored the extent to which different sub-groups of the population respond differently to the mode effect on measurement. The differential effect of mode on respondents with different levels of cognitive ability or language skills has been mentioned as a concern for some researchers, especially when looking at responses reporting attitudes or sensitive information (Cernat, 2015a; Revilla, 2012). This, however, has rarely been subjected to study in previous research, especially when looking at open-ended questions. The evidence from Study 3 suggests, however, that most respondents are affected by mode in a similar way, supporting Revilla's findings (2012), which looked at the respondents' characteristics of age and education and found differences in the quality of responses. However, the differences I observed between sub-groups were not completely as expected from the literature review, and depended on the measure of interest. I found that responses given by respondents older than 65 tended to be affected by mode when reporting their health situation. There were also some differential mode effects between the motivated and the reluctant respondents in responses to the life satisfaction question. There were no differences depending on respondents' characteristics in responses that reported the level of positivity or negativity of each life event examined, but yes in the length of the respondents comparing younger and older respondents (which tended to give longer answers in telephone than the other group).

9.4. The impact of mode in regression analyses of subjective wellbeing

In addition to examining differences in the estimates and means of measures of subjective wellbeing, one of the main objectives was to find the impact that such differences could have in analyses that are commonly used by social science researchers. I provided the illustration of four different models predicting happiness, social trust and wellbeing at work. The results for the model of job satisfaction predicted by satisfaction with work-life balance and level of stress was very similar independently of the response mode. However, there were some differences when looking at the relationship between social trust and its predictors (for example, household income or life satisfaction were only associated in certain), and happiness and its relationship with household income and with the accumulation of negative live events (which was not significant in the case of the telephone sample).

It was not possible, however, to identify a particular pattern in the effects of mode of data collection depending on whether they were found in the outcome variable, in the dependent variable or in both. Along the same lines, it was also not clear that the strongest differences were always observed between self-completion and interviewer-administered modes, even though the differential impact of mode in the means and distributions of the subjective wellbeing items were found between those two response modes. In regression analyses, comparisons between different subgroups of the population are a popular technique in the social sciences, used to test whether the relationship between the predictor and the outcome varies depending on individuals' characteristics such as age, sex, or educational level. I also found differences in these comparisons between modes, when looking at the effect of the interaction between life events and having a low income, which was perhaps one of the most interesting results was found when examining the impact of live events in happiness.

9.5. The impact of mode in multivariate analyses of subjective wellbeing

In the last empirical study (Study 5), I focused on another widely used analysis in wellbeing studies: a multivariate analysis that allows the identification of the effect of mode in multidimensional measures of wellbeing. I implemented a multigroup confirmatory factor analysis to test for the measurement equivalence across modes of two well-being measures: personal subjective wellbeing and wellbeing at work. The equivalence of the measurement models is key for cross-mode comparisons, and the highest level is necessary if researchers are interested in looking at comparisons of means of composite scores in multi-mode surveys (Hox et al., 2017). Results from the study corroborated the hypothesis proposed by de Leeuw (1996) that when mode effects are found in the statistical distributions of individual variables, this may not affect results from multivariate analyses that include such affected variables. Indeed, I found that mode had no effect on the way the measures performed across modes of data collection in the case of the subjective wellbeing model. However, the fit of the model measuring wellbeing at work was different when comparing results from the telephone sample and the web and mail samples.

The insights gained from this study may be of assistance to researchers working on quality of life studies, who face a series of challenges in the analysis of information related to different aspects of subjective wellbeing and vulnerability

(Oris, Roberts, Joye, & Ernst Stähli, 2016); but also to those interested in developing their own research design to collect new data. The findings of the research and potentially the theoretical conclusions drawn from this study contribute in several ways to our understanding of survey design and provide a basis for researchers to understand how mode of data collection can affect the quality of their research.

9.6. Limitations of the research undertaken

Information about the causes of mode effects is valuable for finding ways to minimise them (Roberts, 2007), and the methodological experiment analysed in this thesis provided the opportunity to use data from three single mode surveys that allowed the investigation of differences in selection and measurement effects. In this thesis, in order to investigate the extent of selection and measurement effects, I decided to disentangle the different types of mode effect by rendering the different modes' samples comparable in terms of composition. Although this approach has been widely used in the literature and is relatively straightforward to implement, it also has some disadvantages, as it is very difficult to be certain that the effects observed are uniquely due to mode effects on measurement rather than selection effects associated with variables other than socio-demographic characteristics. In the past few years, alternative methods such as using multi-trait-multi-method designs to assess measurement quality across modes (Saris & Revilla, 2016), or a combination of various techniques have been implemented at the same time (Cernat, 2015a),. It is possible that the availability of a wider range of auxiliary data with which to control the selection differences between modes would have led different results relating to measurement differences to the ones presented in this thesis. It is not uncommon, however, for social science data users, to be unable to make use of such approaches due to special survey data requirements, complex analytic approaches and rarely available information (such as multiple measures of the same concept and panel data).

Throughout the thesis, differences across modes could be potentially reduced if the method used to disentangle them were different, or if additional measures (free of measurement effects themselves) had been used in the matching of the different modes' samples. This is the case for the identification of mode effects, the analysis of differences across subgroups, the test for differences in regression coefficients, and also the test of the measurement invariance in the last study. One of the main challenges to overcome is related to the fact that different modes of data collection attract different types of people that not only differ on socio-demographic characteristics but also on subjective wellbeing and personality characteristics, for which observed measures from the questionnaire would likely be affected by measurement effects themselves.

The approaches used to disentangle measurement and mode effects here has an additional drawback, which is that the measurement effect calculations only take into account respondents that had a listed telephone number, and not the full sample. The decision to focus on this subsample helped in the effort to control for sample differences and hence, in disentangling the different mode effects by comparing more similar respondents across modes, but also reduced the size and heterogeneity of the studied population. Thus, mode differences are only calculated for a part of the sample, and, therefore, cannot be generalised to the whole population. In addition, the fact that the survey only includes participants living in the French speaking part of Switzerland also limits the generalizability of the findings.

The research implemented in this thesis is focused on comparisons of three single-mode surveys. Although in Study 3 I pool data from the mail and web surveys, the aim of the thesis was to look at the differences between these samples, in order to avoid the confounding of additional types of errors. However, focusing on the individual modes separately only provides a limited view of the possible consequences of using mixed-mode data. Indeed, while it is informative about the effects of using concurrent mixed-mode designs, it is not so informative for researchers interested in analysing data from a sequential mixed-mode design.

While it was possible to detect differences between modes, it was not really possible to draw conclusions about which modes provide better data quality, as the mode effect is always calculated with respect to another mode. Telephone respondents chose more positive answers, but it is also true that web and mail respondents tended to report more mild responses, and it is not possible to tell which is the best option to measure wellbeing. However, the results support the idea that using self-completed questionnaires to raise information about socially desirable traits, and opting for the least exaggerated report may be advisable.

In addition, there are some limitations when testing the extent of mode effects in the responses to open-ended questions. Response length is not necessarily the best way of looking at the quality of the responses, and it is particularly uninformative when researchers aim to recode responses into wider categories that miss the higher level of detail. In this case, this is not a useful indicator of quality.

Lastly, additional steps in the analyses implemented to test for the interactions between mode and the different types of respondent could have provided additional detail about subgroup variations in mode effects. In particular, implementing a canonical correlation analysis (Thompson, 2005), on another type of analysis, which indicates whether mode effects for the different dimensions of wellbeing (the different measures grouped by topic) vary for a combination of respondents' characteristics. In this study, due to the small sample sizes of some of the population sub-groups analysed, it was not possible to obtain satisfactory results, but it is could be useful for studies looking at the effect of mode depending on respondents' characteristics, as it allows analysis of the combination of various characteristics.

9.7. Future work

Based on the discussion of the results and the limitations of the research undertaken for this thesis, a number of possible future studies using the same or a similar set up are apparent. A further study could assess the effect of using mixed-mode data on the results of analyses of subjective wellbeing, instead of focusing on comparisons across modes of data collection. Another possible area of future research would be to investigate if the items of subjective wellbeing examined here appear to be affected by measurement effects when using some of the alternative techniques for controlling for selection effects mentioned above. Moreover, social science researchers in Switzerland often use longitudinal data, particularly in the field of subjective wellbeing and vulnerability across the life course. It would be important to gain a deeper understanding of the potential effect of mode on the results of such research – in the analysis of the relationship between negative life events and happiness, for example. This is an important point for wellbeing studies, because even though some of its aspects remain relatively stable over time, others change more often. Understanding how mode effects would impact the measurement of changes over time would be of particular interest – and a crucial consideration for existing longitudinal surveys planning to introduce or already introducing new modes of data collection, such as Understanding Society in the United Kingdom (Jäckle, Lynn & Burton, 2015), or the European Social Survey (Villar & Fitzgerald, 2017).

Another valuable extension to this research would be to repeat the study in different regions of Switzerland, and also in additional countries, which would allow the examination of the interaction between potential cultural differences and mode of data collection in cross-country or cross-regional analyses.

9.8. Conclusion

Notwithstanding some of the limitations discussed earlier, the research undertaken in this thesis suggests that substantive researchers using data that come from different mode designs would benefit from looking at differences in the survey estimates between the different modes of data collection. Although not many differences were found when comparing survey estimates for the mail and the web samples, some types of statistical analyses can be potentially affected, making it difficult to know whether conclusions are accurate or not. The results of this thesis indicate that the presence of mode effects on both outcome and independent variables may lead to different conclusions for researchers. Mode effects are important from the standpoint of substantive research as they can influence the way we analyse and understand data on wellbeing and play an important part in the replication of quantitative studies before conclusions are accepted, on which policy decisions may be based and theoretical conclusions may be drawn. Because the types of mode effects observed appear to depend on the topic and population of study, it is recommended to test for both measurement and selection differences before implementing data comparisons and mixed-mode data analyses. Decisions of how to address potential differences across modes will depend on the researcher’s objective and the type of analysis that is going to be implemented. If the objective is to compare means across subgroups of the population, or across countries whose data has been collected using different modes, my research suggests that whether you use interviewer-based and self-completion survey designs can have strong repercussions for the results obtained. However, using data collected with different response modes to use regression analyses and structural equation models may not affect the substantive conclusions obtained by the researcher. A key priority should therefore be to test for differences between modes when implementing such analyses and arriving at results that could be influenced by how the data were collected. This is of special importance when key policies are being planned on the basis of data from surveys of subjective wellbeing of the population.

Table 63. Summary of results by research question

RQ1. Do different modes of data collection differentially affect the quality of survey estimates of subjective wellbeing?	Both selection and measurement effects were found: • 21 out of 27 measures were sensitive to mode of data collection. • Telephone respondents report higher levels of subjective wellbeing. • Coarsened exact matching controlled for a higher extent of the selection effect than propensity scores and covariates. • Mode also affects responses to open-ended question (non-response rates, positivity of responses)
RQ2. Do mode effects on measurement affect all respondents equally?	Few differences were found: • Reluctant-motivated respondents. • Older than 65 respondents and younger than 65 respondents. • Self-rated health and life satisfaction • Interaction effects only for some response alternatives (extreme-positive, middle categories)
RQ3. Do mode effects on measures of subjective wellbeing impact the results of substantive research into the predictors and correlates of subjective wellbeing measures?	Regression models based on single measures of wellbeing were affected by mode: • Happiness and its predictors, social trust and its predictors, job satisfaction and its predictors Multivariate measures of general subjective wellbeing were consistent across modes

REFERENCES

AAPOR. (2011). Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. *The American Association for Public Opinion Research*, 61. http://doi.org/Doi 10.1111/J.1600-6143.2005.00812.X

Agresti, A., & Finlay, B. (2009). *Statistical methods for the social sciences*. Pearson Prentice Hall.

Allen, J., & van der Velden, R. (2001). Educational mismatches versus skill mismatches: Effects on wages, job satisfaction, and on-the-job search. *Oxford Economic Papers*, *53*(3), 434–452. http://doi.org/10.1093/oep/53.3.434

Alwin, D. F., & Krosnick, J. A. (1991). The Reliability of Survey Attitude Measurement: The Influence of Question and Respondent Attributes. *Sociological Methods & Research*, *20*(1), 139–181. http://doi.org/0803973233

Amaro Galhano, L. (2016). *Nationalité et recrutement: Le cas des ouvriers portugais dans le domaine gros œuvre romand*. *LIVES working papers*. http://doi.org/10.12682/lives.2296-1658.2016.52

Anderson, C., John, O. P., Keltner, D., & Kring, A. M. (2001). Who attains social status? Effects of personality and physical attractiveness in social groups. *Journal of Personality and Social Psychology*, *81*(1), 116–132.

http://doi.org/10.1037/0022-3514.81.1.116

Angrist, J.D., Imbens, G.W., and Rubin, D.B. (1996). Identification of Causal Effects Using Instrumental Variables. Journal of the American Statistical Association, 91, 444-455. DOI: http://www.dx.doi.org/10.1080/01621459.1996.10476902

Austin, A. (2016). On Well-Being and Public Policy: Are We Capable of Questioning the Hegemony of Happiness? *Social Indicators Research*, *127*(1), 123–138. http://doi.org/10.1007/s11205-015-0955-0

Bahadur, A. V., Ibrahim, M., & Tanner, T. (2010). *The resilience renaissance? Unpacking of resilience for tackling climate change and disasters*. *SCR Discussion Paper*. Retrieved from http://www.fsnnetwork.org/sites/default/files/ids_resilience-renaissance.pdf

Barge, S., & Gehlbach, H. (2012). Using the Theory of Satisficing to Evaluate the Quality of Survey Data. *Research in Higher Education*, *53*(2), 182–200. http://doi.org/10.1007/s11162-011-9251-2

Beatty, P., & Herrmann, D. (2002). To answer or not to answer: Decision processes related to survey item nonresponse. In *Survey nonresponse* (pp. 71–85). Retrieved from https://scholar.google.ch/scholar?hl=fr&q=Beatty+and+Herrmann+2002+item+nonresponse&lr=&oq=Beatty+and+Herrmann+2002+item+nonr

Belli, R. F., Conrad, F. G., & Wright, D. B. (2007). Cognitive psychology and survey methodology: Nurturing the continuing dialogue between disciplines. *Applied Cognitive Psychology*, *21*(2), 141–144. http://doi.org/10.1002/acp.1333

Bem, D. (1972). Self-perception theory. *Advances in Experimental Social Psychology*. Retrieved from http://www.dbem.ws/SP Theory.pdf

Bertrand, M., & Mullainathan, S. (2001). Do People Mean What They Say? Implications for Subjective Survey Data. *American Economic Review*, *91*(2), 67–72. http://doi.org/10.1257/aer.91.2.67

Bevan, J. L., Gomez, R., & Sparks, L. (2014). Disclosures about important life events on Facebook: Relationships with stress and quality of life. *Computers in Human Behavior*, *39*, 246–253. http://doi.org/10.1016/j.chb.2014.07.021

Biemer, P. P. (2001). Nonresponse Bias and Measurement Bias in a Comparison of Face to Face and Telephone Interviewing. *Journal of Official Statistics*, *17*(2), 295–320. Retrieved from http://search.proquest.com/openview/b75ed6369c49a12291be7d83033009a8/1?pq-origsite=gscholar&cbl=105444

Biemer, P. P. (2010). Total survey error: Design, implementation, and evaluation. *Public Opinion Quarterly*, *74*(5), 817–848. http://doi.org/10.1093/poq/nfq058

Biemer, P. P., & Lyberg, L. (2003). *Introduction to Survey Quality*. *Technometrics* (Vol. 45). Wiley. http://doi.org/10.1198/tech.2003.s786

Billings, A. G., & Moos, R. H. (1981). The role of coping responses and social resources in attenuating the stress of life events. *Journal of Behavioral Medicine*, *4*(2), 139–157. http://doi.org/10.1007/BF00844267

Birkmann, J. (ed. . (2013). *Measuring Vulnerability to Natural Hazards: Towards Disaster Resilient Societies*. United Nations University Press. http://doi.org/10.1017/CBO9781107415324.004

Blackwell, M., Iacus, S., King, G., & Porro, G. (2009). Cem: Coarsened exact matching in Stata. *Stata Journal*, *9*(4), 524–546. http://doi.org/The Stata Journal

Bland, J. M., & Altman, D. G. (2000). The odds ratio. *BMJ*, *320*(May), 2000. http://doi.org/10.1136/bmj.320.7247.1468

Böckenholt, U. (2017). Measuring response styles in Likert items. *Psychological Methods*, *22*(1), 69–83. http://doi.org/10.1037/met0000106

Börkan, B. (2010). The Mode Effect in Mixed-Mode Surveys: Mail and Web Surveys. *Social Science Computer Review*, *28*(3), 371–380. http://doi.org/10.1177/0894439309350698

Bowling, A. (2005). Mode of questionnaire administration can have serious effects on data quality. *Journal of Public Health*, *27*(3), 281–291. http://doi.org/10.1093/pubmed/fdi031

Bowyer, B. T., & Rogowski, J. C. (2017). Mode Matters: Evaluating Response Comparability in a Mixed-Mode Survey. *Political Science Research and Methods*, *5*(2), 295–313. http://doi.org/10.1017/psrm.2015.28

Bradburn, N. M., Sudman, S., Blair, E., William, L., Miles, C., Singer, E., & Stocking, C. (1979). *Improving Interview Method and Questionnaire Design : Response Effects to Threatening Questions in Survey Research*. *Oxford Journals* (Vol. 59). Retrieved from https://www.ncjrs.gov/App/abstractdb/AbstractDBDetails.aspx?id=69523

Brändle, K. (2017). *The geography of social links among a young cohort in Switzerland*. *LIVES Working papers* (Vol. 58). LIVES. http://doi.org/10.12682/lives.2296-1658.2017.58

Brief, A. P., & Weiss, H. M. (2002). Organizational Behavior: Affect in the Workplace. *Annual Review of Psychology*, *53*(1), 279–307. http://doi.org/10.1146/annurev.psych.53.100901.135156

Brockmann, H., Delhey, J., Welzel, C., & Yuan, H. (2009). The China puzzle: Falling happiness in a rising economy. *Journal of Happiness Studies*, *10*(4), 387–405. http://doi.org/10.1007/s10902-008-9095-4

Brown, G. W., Bifulco, A., & Harris, T. O. (1987). Life events, vulnerability and onset of depression: Some refinements. *British Journal of Psychiatry*, *150*(JAN.), 30–42. http://doi.org/10.1192/bjp.150.1.30

Buelens, B., & Van Den Brakel, J. A. (2017). Comparing Two Inferential Approaches to Handling Measurement Error in Mixed-Mode Surveys. *Journal of Official Statistics*, *33*, 513–531. http://doi.org/10.1515/JOS-2017-0024

Burden, A., Roche, N., Miglio, C., Hillyer, E. V, Postma, D. S., Herings, R. M., … Price, D. B. (2017). An evaluation of exact matching and propensity score methods as applied in a comparative effectiveness study of inhaled corticosteroids in asthma. *Pragmatic and Observational Research*, *Volume 8*, 15–30. http://doi.org/10.2147/POR.S122563

Butz, A. R., Waiters, B. L., Deatrick, E. E., & Usher, E. L. (2013). *Does Method Matter? Open-Ended Responses and Early Adolescents*. Retrieved from http://sites.education.uky.edu/motivation/files/2014/06/Butzetal_APS2013.pdf

Carreño, M.-L., Cardona, O. D., & Barbat, A. H. (2007). Urban Seismic Risk Evaluation: A Holistic Approach. *Natural Hazards*, *40*(1), 137–172. http://doi.org/10.1007/s11069-006-0008-8

Cernat, A. (2015a). *Evaluating mode differences in longitudinal data : moving to a mixed mode paradigm of survey methodology*. *ISER Working Paper Series.* Retrieved from https://www.iser.essex.ac.uk/research/publications/523665

Cernat, A. (2015b). The Impact of Mixing Modes on Reliability in Longitudinal Studies. *Sociological Methods & Research*, *44*(3), 427–457. http://doi.org/10.1177/0049124114553802

Chen, F. F. (2008). What happens if we compare chopsticks with forks? The impact of making inappropriate comparisons in cross-cultural research. *Journal of Personality and Social Psychology*, *95*(5), 1005–18. http://doi.org/10.1037/a0013193

Chen, H., Cohen, P., & Chen, S. (2010). How big is a big odds ratio? Interpreting the magnitudes of odds ratios in epidemiological studies. *Communications in Statistics - Simulation and Computation*, *39*(4), 860–864. http://doi.org/10.1080/03610911003650383

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating Goodness-of- Fit Indexes for Testing Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*(2), 233–255. http://doi.org/10.1207/S15328007SEM0902

Cieciuch, J., Davidov, E., Schmidt, P., Algesheimer, R., & Schwartz, S. H. (2014). Comparing results of an exact vs. an approximate (Bayesian) measurement invariance test: a cross-country illustration with a scale to measure 19 human values. *Frontiers in Psychology*, *5*, 982. http://doi.org/10.3389/fpsyg.2014.00982

Cleland, C., Kearns, A., Tannahill, C., & Ellaway, A. (2015). Home truths: Are housing-related events more important for residents' health compared with other life events? *Housing Studies*, *3037*(February 2016), 1–24. http://doi.org/10.1080/02673037.2015.1094565

Clingingsmith, D. (2016). Negative emotions, income, and welfare: Causal estimates from the PSID. *Journal of Economic Behavior and Organization*, *130*, 1–19. http://doi.org/10.1016/j.jebo.2016.07.004

Collins, D. (2003). Pretesting survey instruments: An overview of cognitive methods. *Quality of Life Research*, *12*(3), 229–238. http://doi.org/10.1023/A:1023254226592

Conti, G., & Pudney, S. (2011). Survey design and the analysis of satisfaction. *Review of Economics and Statistics*, *93*(3), 1087–1093. http://doi.org/10.1162/REST_a_00202

Couper, M. P. (2017). New Developments in Survey Data Collection. *Annual Review of Sociology*, *43*(1), annurev-soc-060116-053613. http://doi.org/10.1146/annurev-soc-060116-053613

D'Orazio, M., Di Zio, M., & Scanu, M. (2006). *Statistical Matching: Theory and Practice*. *Statistical Matching: Theory and Practice*. Wiley. http://doi.org/10.1002/0470023554

Daniels, E., King, M. A., Smith, I. E., & Shneerson, J. M. (2001). Health-related quality of life in narcolepsy. *Journal of Sleep Research*, *10*(1), 75–81. http://doi.org/10.1046/j.1365-2869.2001.00234.x

Dasoki, N., Morselli, D., & Spini, D. (2015). *La mémoire autobiographique à travers le parcours de vie : évaluation rétrospective du bonheur et de la vulnérabilité* (No. 43). Lausanne. http://doi.org/10.12682/lives.2296-1658.2015.43

Davis, C. G., Thake, J., & Vilhena, N. (2010). Social desirability biases in self-reported alcohol consumption and harms. *Addictive Behaviors*, *35*(4), 302–311. http://doi.org/10.1016/j.addbeh.2009.11.001

de Leeuw, E. D., Mellenbergh, G. J., & Hox, J. J. (1996). The Influence of Data Collection Method on Structural Models: A Comparison of a Mail, a Telephone, and a Face-to-Face Survey. *Sociological Methods & Research*, *24*(4), 443–472. http://doi.org/10.1177/0049124196024004002 de Leeuw, E. D. (2005). To Mix or Not to Mix Data Collection Modes in Surveys. *Journal of Official Statistics*, *21*(2), 233–255.

de Leeuw, E. D., Dillman, D. A., & Hox, J. J. (2007). *International Handbook of Survey Methodology*. Retrieved from

de Leeuw, E., Hox, J. J., & Dillman, D. A. (2008). *International handbook of survey methodology*. Lawrence Erlbaum Associates. Retrieved from https://www.routledge.com/International-Handbook-of-Survey-Methodology/de-Leeuw-Hox-Dillman/p/book/9780805857535

de Leeuw, E. D., Hox, J., & Scherpenzeel, A. (2010). Mode Effect or Question Wording? Measurement Error in Mixed Mode Surveys. In *JSM Proceedings (Survey Research Methods Section)* (pp. 5959–5967). http://doi.org/http://www.amstat.org/sections/srms/proceedings/y2010/Files/400117.pdf

de Neve, J., Diener, E., Tay, L., & Xuereb, C. (2013). The objective benefits of subjective well-being. *CEP Discussion Paper No 1236*, (1236), 1–35. Retrieved from https://poseidon01.ssrn.com/delivery.php?ID=4930060200021250300650230910970641220230210260630300571040211130270830100290951270810960980511270251110140700641020720051241201080840750580760941261180690930240260520320170210001221140201191170161120881070800960

Deaton, A., & Stone, A. A. (2013). *Do context effects limit the usefulness of self-reported wellbeing measures?* Retrieved from https://www.princeton.edu/~deaton/downloads/Context effects and wellbeing july all.pdf

De Vaus, D. A. (2012). *Surveys in social research*. Routledge. Retrieved from https://books.google.ch/books?hl=fr&lr=&id=rnxiAgAAQBAJ&oi=fnd&pg=PP1&dq=sociological+theory+and+survey+methods&ots=6cXEFbsxQQ&sig=a4ZGda96o6bg6mWRJbpMUUejId8#v=onepage&q=sociological theory and survey methods&f=false

DeCastellarnau, A. (2017). A classification of response scale characteristics that affect data quality: a literature review. *Quality and Quantity*, 1–37. http://doi.org/10.1007/s11135-017-0533-4

Deci, E. L., & Ryan, R. M. (2008). Hedonia, eudaimonia, and well-being: An introduction. *Journal of Happiness Studies*, *9*(1), 1–11. http://doi.org/10.1007/s10902-006-9018-1

Dehejia, R. H., & Wahba, S. (2002). Propensity Score-Matching Methods for Nonexperimental Causal Studies. *Review of Economics and Statistics*, *84*(1), 151–161. http://doi.org/10.1162/003465302317331982

DeMaio, T. J. (1984). Social desirability and survey measurement. A review. In C. F. Turner & E. Martin (Eds.), *Surveying subjective phenomena* (Russell Sa). New York: SAGE Publications Ltd. http://doi.org/10.4135/9781446263501

Denscombe, M. (2008). The Length of Responses to Open-Ended Questions: A Comparison of Online and Paper Questionnaires in Terms of a Mode Effect. *Social Science Computer Review*, *26*(3), 359–368. http://doi.org/10.1177/0894439307309671

Dias de Freitas, Kamyla Thais Pinheiro Ferrari, E., Poerschke Vieira, M., da Silva, W. R., Pereira de Carvalho, H., & Cardoso, F. L. (2016). Association of subjective social status and sociodemographic indicators in athletes. *Revista Brasileira de Cineantropometria & Desempenho Humano*, *18*(5), 591–601.

http://doi.org/10.5007/1980-0037.2016v18n5p591

Dickinson, T., Fernandez, M., Thomas, L. A., Mulholland, P., Briggs, P., & Alani, H. (2016). Identifying Important Life Events from Twitter Using Semantic and Syntactic Patterns, 143–150. Retrieved from http://oro.open.ac.uk/48679/1/IDENTIFYING IMPORTANT LIFE EVENTS FROM%25aTWITTER USING SEMANTIC AND SYNTACTIC%25aPATTERNS.pdf

Diener, E. (1984). Subjective well-being. *Psychological Bulletin*, *95*(3), 542–575. http://doi.org/10.1037/0033-2909.95.3.542

Diener, E. (1994). Assessing subjective well-being: Progress and opportunities. *Social Indicators Research*, *31*(2), 103–157. http://doi.org/10.1007/BF01207052

Diener, E. (2009). Introduction – Measuring Well-Being: Collected Theory and Review Works (pp. 1–6). http://doi.org/10.1007/978-90-481-2354-4_1

Diener, E. (2000). Subjective well-being: The science of happiness and a proposal for a national index. American Psychologist, 55(1), 34-43.

Diener, E., & Biswas-Diener, R. (2002). Will money increase subjective well-being? A literature review and guide to needed research. *Social Indicators Research*, *57*(2), 119–169. http://doi.org/10.1023/A:1014411319119

Diener, E., Heintzelman, S. J., Kushlev, K., Tay, L., Wirtz, D., Lutes, L. D., & Oishi, S. (2017). Findings all psychologists should know from the new science on subjective well-being. *Canadian Psychology/Psychologie Canadienne*, *58*(2), 87–104. http://doi.org/10.1037/cap0000063

Diener, E., Lucas, R., Schimmack, U., & Helliwell, J. (2010). *Well-Being for Public Policy*. *Well-Being for Public Policy*. Oxford University Press. http://doi.org/10.1093/acprof:oso/9780195334074.001.0001

Diener, E., Oishi, S., & Lucas, R. E. (2002). Subjective Well-Being: The Science of Happiness and Life Satisfaction. In *The Oxford Handbook of Positive Psychology, (2 Ed.)*. Oxford University Press. http://doi.org/10.1093/oxfordhb/9780195187243.013.0017

Diener, E., Suh, E., & Oishi, S. (1997). Recent Findings on Subjective Well-Being. *Indian Journal of Clinical Psychology*, *24*(1), 25–41. Retrieved from http://psycnet.apa.org/psycinfo/1997-43193-002

Dillman, D. A. The Design And Administration Of Mail Surveys, 17Annual Review of Sociology 225–249 (1991). http://doi.org/10.1146/annurev.soc.17.1.225

Dillman, D. A. (2017). The promise and challenge of pushing respondents to the web in mixed-mode surveys. *Survey Methodology*, *43*(1), 3–30. Retrieved from http://www.statcan.gc.ca/pub/12-001-x/2017001/article/14836-eng.pdf

Dillman, D. A., & Christian, L. M. (2005). Survey Mode as a Source of Instability in Responses across Surveys. *Field Methods*, *17*(1), 30–52. http://doi.org/10.1177/1525822X04269550

Dillman, D. A., Phelps, G., Tortora, R., Swift, K., Kohrell, J., Berck, J., & Messer, B. L. (2009). Response rate and measurement differences in mixed-mode surveys using mail, telephone, interactive voice response (IVR) and the Internet. *Social Science Research*, *38*(1), 1–18. http://doi.org/10.1016/j.ssresearch.2008.03.007

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, phone, mail, and mixed-mode surveys: the tailored method design*. Hoboken, New Jersey: Wiley.

http://doi.org/10.2307/41061275

Disabato, D. J., Goodman, F. R., Kashdan, T. B., Short, J. L., & Jarden, A. (2016). Different types of well-being? A cross-cultural examination of hedonic and eudaimonic well-being. *Psychological Assessment*, *28*(5), 471–482. http://doi.org/10.1037/pas0000209

Dolan, P., & Kavetsos, G. (2012). *Happy talk: mode of administration effects on subjective well-being*. London, UK: London School of Economics and Political Science.

Dolan, P., & Kavetsos, G. (2016). Happy Talk: Mode of Administration Effects on Subjective Well-Being. *Journal of Happiness Studies*, *17*(3), 1273–1291. http://doi.org/10.1007/s10902-015-9642-8

Dolan, P., & Metcalf, R. (2011). Comparing measures of subjective well-being and views about the role they should play in policy. *Office for National Statistics*, *25*(July), 07–11. Retrieved from https://scholar.google.ch/scholar?hl=en&as_sdt=0%2C5&q=Comparing+measures+of+subjective+well-being+and+views+about+the+role+they+should+play+in+policy.+Office+for+National+Statistic&btnG=

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. http://doi.org/10.1016/j.joep.2007.09.001

Emde, M. (2014). *Open-ended questions in Web surveys. Using visual and adaptive questionnaire design to improve narrative responses*. Retrieved from http://tuprints.ulb.tu-darmstadt.de/4219/

Esposito, J. ., & Jobe, J. . (1991). *A General Model of the Survey Interaction Process*. Retrieved from https://www.bls.gov/osmr/abstract/st/st910040.htm

European Social Survey. (2015). *Measuring and Reporting on Europeans ' Wellbeing : Findings From the European Social Survey*. Retrieved from www.esswellbeingmatters.org, accessed dd/mm/yy

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. http://doi.org/10.1111/j.1468-0297.2004.00235.x

Finney, S. J., & DiStefano, C. (2013). Nonnormal and categorical data in structural equation modelling. In Hancock & Mueller (Eds.), *G. R. Hancock & R. O. Mueller [Ed], Structural equation modeling: A second course (2nd ed.)* (pp. 439–492). Greenwich: nformation Age Publishing.

Fleche, S., Smith, C., & Sorsa, P. (2012). Exploring Determinants of Subjective Wellbeing in OECD Countries EVIDENCE FROM THE WORLD VALUE SURVEY. *Working Papers, OECD Statistics*, (2012). http://doi.org/10.1787/5kg0k6zlcm5k-en

FORS. (2016). *Data on wellbeing in the Swiss Social Report 2016*. Retrieved from http://www.bfs.admin.ch/bfs/portal/en/index/infothek/erhebungen__quellen/blank/blank/bha/00.

Frey, B. S., & Stutzer, A. (2000). Happiness, Economy and Institutions. *The Economic Journal*, *110*(466), 918–938. http://doi.org/10.1111/1468-0297.00570

Frost, J. (2013). Regression Analysis: How Do I Interpret R-squared and Assess the Goodness-of-Fit? Retrieved September 15, 2017, from http://blog.minitab.com/blog/adventures-in-statistics-2/regression-analysis-how-do-i-interpret-r-squared-and-assess-the-goodness-of-fit

Garnefski, N., Kraaij, V., & Spinhoven, P. (2001). Negative life events, cognitive emotion regulation and emotional problems. *Personality and Individual Differences*, *30*(8), 1311–1327. http://doi.org/10.1016/S0191-8869(00)00113-6

Groves, R. M., Fowler, F. J., Couper, M., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey methodology*. Wiley.

Groves, R. M., & Lyberg, L. (2010). Total survey error: Past, present, and future. *Public Opinion Quarterly*, *74*(5), 849–879. http://doi.org/10.1093/poq/nfq065

Groves, R. M., & Peytcheva, E. (2008). The impact of nonresponse rates on nonresponse bias: A meta-analysis. *Public Opinion Quarterly*. Oxford University Press. http://doi.org/10.1093/poq/nfn011

Guest, D. E. (2002). Perspectives on the Study of Work-life Balance. *Social Science Information*, *41*(2), 255–279. http://doi.org/10.1177/0539018402041002005

Halleröd, B., & Seldén, D. (2013). The Multi-dimensional Characteristics of Wellbeing: How Different Aspects of Wellbeing Interact and Do Not Interact with Each Other. *Social Indicators Research*, *113*(3), 807–825. http://doi.org/10.1007/s11205-012-0115-8

Hanappi, D., & Mancini-Vonlanthen, N. (n.d.). *Spouses and families of expatriates: A systematic literature review by correspondence analysis based on semantic segments*. http://doi.org/10.12682/lives.2296-1658.2016.49

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*(12), 1171–9. http://doi.org/10.1097/MLR.0b013e3181354828

Heckman, J. (1997). Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations. The Journal of Human Resources, 32(3), 441-462. doi:10.2307/146178

Heerwegh, D., & Loosveldt, G. (2011). Assessing Mode Effects in a National Crime Victimization Survey using Structural Equation Models : Social Desirability Bias and Acquiescence. *Journal of Official Statistics*, *27*(1), 49–63.

Helliwell, J. F., & Putnam, R. D. (2004). The social context of well-being. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *359*(1449), 1435–46. http://doi.org/10.1098/rstb.2004.1522

Hewitt, M. (2002). Attitudes toward Interview Mode and Comparability of Reporting Sexual Behavior by Personal Interview and Audio Computer-assisted Self-interviewing: Analyses of the 1995 National Survey of Family Growth. *Sociological Methods & Research*, *31*(1), 3–26. http://doi.org/10.1177/0049124102031001001

Hirschfeld, G., & Von Brachel, R. (2014). Multiple-Group confirmatory factor analysis in R – A tutorial in measurement invariance with continuous and ordinal. *Practical Assessment, Research & Evaluation*, *19*(7), 1–11.

Holbrook, A. L., Green, M. C., & Krosnick, J. A. (2003). Telephone versus face-to-face interviewing of national probability samples with long questionnaires. *Public Opinion Quarterly*, *67*(1), 79–125. http://doi.org/10.1086/346010

Holford, A., & Pudney, S. (2015). Survey Design and the Determinants of Subjective Wellbeing. An Experimental Analysis. *IZA Discussion Paper*, (8760).

Hooper, D., Coughlan, J., & Mullen, M. R. (2008). Structural Equation Modelling: Guidelines for Determining Model Fit. *The Electronic Journal of Business Research Methods*, *6*(1), 53–60.

Hope, S., Campanelli, P., Nicolaas, G., Lynn, P., & Jäckle, A. (2014). *The Role of the Interviewer in Producing Mode Effects: Results from a Mixed Modes Experiment Comparing Face-to-Face, Telephone and Web Administration*. Economic &

Social Research Council (Vol. 20141). *ISER Working Paper Series*. Retrieved from https://www.iser.essex.ac.uk/research/publications/working-papers/iser/2014-20

Hox, J., E.D. de Leeuw, and T. Klausch. 2017. “Mixed mode research: Issues in design and analysis.” In Total Survey Error in Practice: Improving Quality in the Era of Big Data, edited by P.P. Biemer, E.D. de Leeuw, S. Eckman, B. Edwards, F. Kreuter, L.E. Lyberg, Tucker, C. and B.T. West, 511–530. Hoboken, NJ: John Wiley and Sons, Inc.

Hox, J. J., de Leeuw, E. D., & Zijlmans, E. A. O. (2015). Measurement equivalence in mixed mode surveys. *Frontiers in Psychology*. http://doi.org/10.3389/fpsyg.2015.00087

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling, 6, 1-55.

Hu, L., & Bentler, P. M. (2009). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *http://dx.doi.org/10.1080/10705519909540118*.

Huppert, F. A., Marks, N., Michaelson, J., Vázquez, C., & Vittersø, J. (2013). ESS6 - 2012/3 Question Module Design Final Template, 1–36. Retrieved from https://www.europeansocialsurvey.org/docs/round6/questionnaire/ESS6_final_personal_and_social_well_being_module_template.pdf

Huppert, F. A., & So, T. T. C. (2013). Flourishing Across Europe: Application of a New Conceptual Framework for Defining Well-Being. *Social Indicators Research*, *110*(3), 837–861. http://doi.org/10.1007/s11205-011-9966-7

Iacus, S. M., King, G., & Porro, G. (2012). Causal inference without balance checking: Coarsened exact matching. *Political Analysis*, *20*(1), 1–24. http://doi.org/10.1093/pan/mpr013

Ihle, A., Oris, M., Fagot, D., Baeriswyl, M., Guichard, E., & Kliegel, M. (2015). The Association of Leisure Activities in Middle Adulthood with Cognitive Performance in Old Age: The Moderating Role of Educational Level. *Gerontology*, *61*(6), 543–550. http://doi.org/10.1159/000381311

Jäckle, A., Lynn, P., & Burton, J. (2015). Going Online with a Face-to-Face Household Panel: Effects of a Mixed Mode Design on Item and Unit Non-Response. *Survey Research Methods European Survey Research Association ISSN*, *9*(1), 57–70. http://doi.org/10.18148/srm/2015.v9i1.5475

Jäckle, A., Roberts, C., & Lynn, P. (2010). Assessing the effect of data collection mode on measurement. *International Statistical Review*, *78*(1), 3–20. http://doi.org/10.1111/j.1751-5823.2010.00102.x

Jans, M. (2008). Mode Effects. In P. J. Lavrakas (Ed.), *Encyclopedia of Survey Research Methods* (pp. 476–480). 2455 Teller Road, Thousand Oaks California 91320 United States of America: Sage Publications, Inc. http://doi.org/10.4135/9781412963947.n302

Joye, D., Pollien, A., Sapin, M., & Stähli, M. E. (2012). Who can be contacted by phone? Lessons from Switzerland. In *Telephone Surveys in Europe: Research and Practice* (Vol. 9783642254, pp. 85–102). Berlin, Heidelberg: Springer Berlin Heidelberg. http://doi.org/10.1007/978-3-642-25411-6_8

Kaminska, O., & Foulsham, T. (2013). *Understanding Sources of Social Desirability Bias in Different Modes : Evidence from Eye-tracking*.

King, G., & Nielsen, R. (2015). *Why Propensity Scores Should Not Be Used for Matching*. Retrieved from https://gking.harvard.edu/presentations/why-propensity-scores-should-not-be-used-matching-6

King, G., Nielsen, R., Coberley, C., & Pope, J. (2010). Comparative Effectiveness of Matching Methods for Causal Inference. *Director of Health Research and Outcomes*. Retrieved from http://people.fas.harvard.edu/

Klausch, T., Hox, J. J., & Schouten, B. (2013). Measurement Effects of Survey Mode on the Equivalence of Attitudinal Rating Scale Questions. *Sociological Methods & Research*, *42*(3), 227–263. http://doi.org/10.1177/0049124113500480

Klausch, T., Schouten, B., & Hox, J. (2014). The use of within-subject experiments for estimating measurement effects in mixed-mode surveys. Retrieved October 27, 2017, from https://www.cbs.nl/en-gb/background/2014/11/the-use-of-within-subject-experiments-for-estimating-measurement-effects-in-mixed-mode-surveys

Knecht, M., Wiese, B. S., & Freund, A. M. (2016). Going beyond work and family: A longitudinal study on the role of leisure in the work-life interplay. *Journal of Organizational Behavior*, *37*(7), 1061–1077. http://doi.org/10.1002/job.2098

Kolenikov, S., & Kennedy, C. (2014). Evaluating three approaches to statistically adjust for mode effects. *Journal of Survey Statistics and Methodology*, *2*(2), 126–158. http://doi.org/10.1093/jssam/smu004

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. http://doi.org/10.1093/poq/nfn063

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, *5*(3), 213–236.

http://doi.org/10.1002/acp.2350050305

Krosnick, J. A. (1999). Survey research. *Annual Review of Psychology*, *50*(1), 537–567. http://doi.org/10.1146/annurev.psych.50.1.537

Krosnick, J. A., Holbrook, A. L., Berent, M. K., Carson, R. T., Hanemann, W. M., Kopp, R. J., … Conaway, M. (2012). The impact of “no opinion” response options on data quality: Non-attitude reduction or an invitation to satisfice? *Public Opinion Quarterly*, *66*(3), 371–403. http://doi.org/10.1086/341394

Krosnick, J. A., & Presser, S. (2010). *Question and Questionnaire Design*. *Handbook of Survey Research*. http://doi.org/10.1111/j.1432-1033.1976.tb10115.x

Krosnick, J. A., Villar, A., & MacInnis, B. (2011). *Do Introductory Sentences Cause Acquiescence Response Bias in Survey Questions?* Retrieved from https://woods.stanford.edu/sites/default/files/files/Krosnick-Survey-Acquiescence.pdf

Krueger, A. B., & Schkade, D. A. (2008). The reliability of subjective well-being measures. *Journal of Public Economics*, *92*(8–9), 1833–1845. http://doi.org/10.1016/j.jpubeco.2007.12.015

Kubiszewski, I., Costanza, R., Franco, C., Lawn, P., Talberth, J., Jackson, T., & Aylmer, C. (2013). Beyond GDP: Measuring and achieving global genuine progress. *Ecological Economics*, *93*, 57–68. http://doi.org/10.1016/j.ecolecon.2013.04.019

Kushlev, K., Dunn, E. W., & Lucas, R. E. (2015). Higher Income Is Associated With Less Daily Sadness but not More Daily Happiness. *Social Psychological and Personality Science*, *6*(5), 483–489. http://doi.org/10.1177/1948550614568161

Kwak, N., & Radler, B. (2002). A Comparison Between Mail and Web Surveys: Response Pattern, Respondent Profile, and Data Quality. *Journal of Official*

Statistics, *18*(2), 257–273. Retrieved from http://search.proquest.com/openview/a4767f2c969229aa031a451392da9ce6/1?pq-origsite=gscholar&cbl=105444

Layard, R. (2005). *Happiness: Lessons from a new science*. Penguin Press.

Lipps, O. (2016). Non-Observation Bias in an Address- Register-Based CATI/CAPI Mixed Mode Survey. *Methods, Data, Analyses*, *10*(1), 5–24. http://doi.org/10.12758/mda.2016.001

Liu, M. (2014). *Response Style and Rating Scales: The Effects of Data Collection Mode, Scale Format, and Acculturation*. Retrieved from http://psm.isr.umich.edu/sites/default/files/Mingnan Liu Dissertation Defense.pdf

Liu, M., Conrad, F. G., & Lee, S. (2017). Comparing acquiescent and extreme response styles in face-to-face and web surveys. *Quality and Quantity*, *51*(2), 941–958. http://doi.org/10.1007/s11135-016-0320-7

Lugtig, P. J., Lensvelt-Mulders, G. J., Frerichs, R., Greven, A. (2011). Estimating nonresponse bias and mode effects in a mixed-mode survey. *International Journal of Market ResearchJournal of Market Research*, *53*(5), 669–686. http://doi.org/10.2501/ijmr-53-5-669-686

Luhmann, M., Hawkley, L. C., Eid, M., & Cacioppo, J. T. (2012). Time frames and the distinction between affective and cognitive well-being. *Journal of Research in Personality, 46*(4), 431–441.

Lydeard, S., & Jones, R. (1989). Life events vulnerability and illness: A selected review. *Family Practice*, *6*(4), 307–315. http://doi.org/10.1093/fampra/6.4.307

Lyubomirsky, S., & Lepper, H. S. (1999). A measure of subjective happiness: Preliminary reliability and construct validation. *Social Indicators Research*, *46*(2), 137–155. http://doi.org/10.1023/A:1006824100041

Maggiori, C., Rossier, J., Krings, F., Johnston, C., & Massoudi, K. (2014). Career pathways and professional transitions: preliminary results from the 1st wave of a 7-year longitudinal study (pp. 1–34).

Martin, P., & Lynn, P. (2011). The effects of mixed mode survey designs on simple and complex analyses. *ISER Working Paper Series*. Retrieved from http://ideas.repec.org/p/ese/iserwp/2011-28.html

Martínez-Martí, M. L., & Ruch, W. (2016). The Relationship Between Orientations to Happiness and Job Satisfaction One Year Later in a Representative Sample of Employees in Switzerland. *Journal of Happiness Studies*, 1–15. http://doi.org/10.1007/s10902-016-9714-4

McFall, S. L. (Ed). (2012). *Understanding Society: Findings 2012*. Retrieved from https://www.understandingsociety.ac.uk/research/publications/findings/2012

Mclauchlan, C., & Schonlau, M. (2016). Are Final Comments in Web Survey Panels Associated with Next-Wave Attrition ? *Survey Research Methods*, *10*(3), 211–224. http://doi.org/10.18148/srm/2016.v10i3.6217

Michaelson, J., Abdallah, S., Thompson, N., Hompson, S., Marks, N., Aked, J., … Potts, R. (2009). *National Accounts of Well-being: bringing real wealth onto the balance sheet*. London. Retrieved from http://www.nationalaccountsofwellbeing.org/learn/download-report.html

Millán, J. M., Hessels, J., Thurik, R., & Aguado, R. (2013). Determinants of job satisfaction: A European comparison of self-employed and paid employees. *Small Business Economics*, *40*(3), 651–670. http://doi.org/10.1007/s11187-011-9380-1

Mohorko, A., de Leeuw, E., & Hox, J. J. (2013). Coverage Bias in European Telephone Surveys: Developments of Landline and Mobile Phone Coverage across Countries and over Time. *Survey Methods: Insights from the Field.*, 1–27. http://doi.org/10.13094/SMIF-2013-00002

Muñoz de Bustillo Llorente, R., & Fernández Macías, E. (2005). Job satisfaction as an indicator of the quality of work. *Journal of Socio-Economics*, *34*(5), 656–673. http://doi.org/10.1016/j.socec.2005.07.027

Nandi, A., & Platt, L. (2017). Are there differences in responses to social identity questions in face-to-face versus telephone interviews? Results of an experiment on a longitudinal survey. *International Journal of Social Research Methodology*, *20*(2), 151–166. http://doi.org/10.1080/13645579.2016.1165495

Narum, S.R. (2006). Beyond Bonferroni: Less conservative analyses for onservation genetics. Conservation Genetics, 7, 783–787.

Niemi, I. (1993). Systematic error in behavioural measurement: Comparing results from interview and time budget studies. *Social Indicators Research*, *30*(2–3), 229–244. http://doi.org/10.1007/BF01078729

O'Brien, K., Leichenko, R., Kelkar, U., Venema, H., Aandahl, G., Tompkins, H., … West, J. (2004). Mapping vulnerability to multiple stressors: Climate change and globalization in India. *Global Environmental Change*, *14*(4), 303–313. http://doi.org/10.1016/j.gloenvcha.2004.01.001

OECD. (2013). *OECD Guidelines on Measuring Subjective Well-being*. *Report*. OECD Publishing. http://doi.org/10.1787/9789264191655-en

OECD. (2015). *How's life? 2015: Measuring Well-being*. *How's Life? 2015: Measuring well-being*. OECD Publishing. http://doi.org/10.1787/how_life-2015-en

Oris, M., Roberts, C., Joye, D., & Ernst Stähli, M. (2016). Surveying Human Vulnerabilities Across the Life Course: Balancing Substantive and

Methodological Challenges (pp. 1–25). Springer, Cham. http://doi.org/10.1007/978-3-319-24157-9_1

Oris, M. (2017). La vulnérabilité, une approche par le parcours de vie. Revue de droit comparé du travail et de la sécurité sociale (1), pp. 36–45.

Paggi, M. E., Jopp, D., & Hertzog, C. (2016). The Importance of Leisure Activities in the Relationship between Physical Health and Well-Being in a Life Span Sample. *Gerontology*, *62*(4), 450–458. http://doi.org/10.1159/000444415

Park, C. L. (2010). Making sense of the meaning literature: an integrative review of meaning making and its effects on adjustment to stressful life events. *Psychological Bulletin*, *136*(2), 257–301. http://doi.org/10.1037/a0018301

Paulhus, D. (2002). Socially desirable responding: The evolution of a construct. In H. I. Braun, D. N. Jackson, & D. E. Wiley (Eds.), *The role of constructs in psychological and educational measurement* (pp. 49–69). Routledge. Retrieved from https://books.google.ch/books?hl=en&lr=&id=JnmRAgAAQBAJ&oi=fnd&pg=PA51&dq=socially+desirable+responding+paulhus+2002&ots=45dLQ3jOsJ&sig=IAwj5EGE35fVQVtiYG2cKXRcTY0

Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology*, *46*(3), 598–609. http://doi.org/10.1037/0022-3514.46.3.598

Paulhus, D. L., & Reid, D. B. (1991). Enhancement and denial in socially desirable responding. *Journal of Personality and Social Psychology*, *60*(2), 307–317. http://doi.org/10.1037/0022-3514.60.2.307

Pearl, J. (2009). Chapter 11.3.5 Understanding Propensity Scores. In *Causality Models, Reasoning and Inference (Second Edition)* (Vol. 8, pp. 405–408).

Retrieved from http://bayes.cs.ucla.edu/BOOK-09/ch11-3-5-final.pdf

Pouwels, B., Siegers, J., & Vlasblom, J. D. (2008). Income, working hours, and happiness. *Economics Letters*, *99*(1), 72–74. http://doi.org/10.1016/j.econlet.2007.05.032

Pudney S. (2010). An experimental analysis of the impact of survey design on measures and models of subjective well-being. ISER Working Paper 2010-20

Reja, U., Manfreda, K. L., Hlebec, V., & Vehovar, V. (2003). Open-ended vs. Close-ended Questions in Web Questionnaires. *Developments in Applied Statistics*, *19*, 159–177. Retrieved from http://www.stat-d.si/mz/mz19/reja.pdf

Revilla, M. (2012). Impact of the Mode of Data Collection on the Quality of Answers to Survey Questions Depending on Respondent Characteristics. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique*, *116*(1), 44–60. http://doi.org/10.1177/0759106312456510

Revilla, M. (2010). *Impact of the mode of data collection on the quality of survey questions in social sciences*. Universitat Pompeu Fabra. Retrieved from https://repositori.upf.edu/handle/10230/19815

Revilla, M. (2013). Measurement invariance and quality of composite scores in a face-to-face and a web survey. *Survey Research Methods*, *7*(1), 17–28. http://doi.org/10.18148/srm/2013.v7i1.5098

Revilla, M., & Saris, W. E. (2013, June 1). A Comparison of the Quality of Questions in a Face-to-face and a Web Survey. *International Journal of Public Opinion Research*. Oxford University Press. http://doi.org/10.1093/ijpor/eds007

Revilla, M., & Ochoa, C. (2016). Open narrative questions in PC and smartphones: is the device playing a role? *Quality and Quantity*, *50*(6), 2495–2513. http://doi.org/10.1007/s11135-015-0273-2

Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods*, *17*(3), 354–373. http://doi.org/10.1037/a0029315

Roberts, C. (2007). *Mixing modes of data collection in surveys : A methodological review ESRC National Centre for Research Methods*. *NCRM Methods Review Paper*. Retrieved from http://eprints.ncrm.ac.uk/418/

Roberts, C. (2016). Response Styles in Surveys: Understanding their Causes and Mitigating their Impact on Data Quality. In C. Wolf, D. Joye, T. W. Smith, & Y. Fu (Eds.), *The SAGE Handbook of Survey Methodology* (pp. 579–596). 1 Oliver's Yard, 55 City Road London EC1Y 1SP: SAGE Publications Ltd. http://doi.org/10.4135/9781473957893.n36

Roberts, C., Joye, D., & Ernst Stähli, M. (2016). *Mixing modes of data collection in Swiss social surveys: Methodological report of the LIVES-FORS mixed mode experiment*. *LIVES Working Papers* (Vol. 2016). http://doi.org/10.12682/lives.2296-1658.2016.48

Roberts, C., & Vandenplas, C. (2017). Estimating Components of Mean Squared Error to Evaluate the Benefits of Mixing Data Collection Modes. *Journal of Official Statistics*, *33*(2), 303–334. http://doi.org/10.1515/jos-2017-0016

Rode, N. (2005). Translation of Measurement Instruments and their Reliability: An Example of Job-Related Affective Well-Being Scale. *Metodološki Zvezki*, *2*(1), 15–26. Retrieved from https://www.stat-d.si/mz/mz2.1/rode.pdf

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 41–55. http://doi.org/10.1093/biomet/70.1.41

Ross, C. E., & Mirowsky, J. (1984). Socially-Desirable Response and Acquiescence in a Cross-Cultural Survey of Mental Health. *Journal of Health and Social Behavior*, *25*(2), 189–197. http://doi.org/10.2307/2136668

Ross, M. E., Kreider, A. R., Huang, Y. S., Matone, M., Rubin, D. M., & Localio, A. R. (2015). Propensity Score Methods for Analyzing Observational Data Like Randomized Experiments: Challenges and Solutions for Rare Outcomes and Exposures. *American Journal of Epidemiology*, *181*(12), 989–995. http://doi.org/10.1093/aje/kwu469

Rothenbühler, M., & Voorpostel, M. (2016). Attrition in the Swiss Household Panel: Are Vulnerable Groups more Affected than Others?, 221–242. http://doi.org/10.1007/978-3-319-24157-9_10

Ruch, W., Martínez-Martí, M. L., Heintz, S., & Brouwers, S. A. (2014). Short form of the orientations to happiness questionnaire for the german-speaking countries: Development and analysis of the psychometric properties. *Swiss Journal of Psychology*, *73*(4), 225–234. http://doi.org/10.1024/1421-0185/a000141

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*(5), 907–933. http://doi.org/10.1093/poq/nfq057

Saris, W. E., & Gallhofer, I. N. (2007). Estimation of the effects of measurement characteristics on the quality of survey questions. *Survey Research Methods*, *1*(1), 29–43. http://doi.org/10.18148/srm/2007.v1i1.49

Saris, W. E., & Revilla, M. (2016). Correction for Measurement Errors in Survey Research: Necessary and Possible. *Social Indicators Research*, *127*(3), 1005–1020. http://doi.org/10.1007/s11205-015-1002-x

Sarracino, F., Riillo, C. A. F., & Mikucka, M. (2017). Comparability of web and telephone surveys for the measurement of subjective well-being. *Survey Research Methods*, *11*(2), 141–169. http://doi.org/10.18148/srm/2017.v11i2.6740

Schaefer, D. R., & Dillman, D. A. (1998). Development of a standard em-mail methodology: Results of an experiment. *The Public Opinion Quarterly*, *62*(3), 378–397. http://doi.org/10.1086/297851

Scholz, E., & Zuell, C. (2012). Item non-response in open-ended questions: Who does not answer on the meaning of left and right? *Social Science Research*, *41*(6), 1415–1428. http://doi.org/10.1016/j.ssresearch.2012.07.006

Schonlau, M., & Couper, M. P. (2016). Semi-automated categorization of open-ended questions. *Survey Research Methods*, *10*(2), 143–152. http://doi.org/10.18148/srm/2016.v10i2.6213

Schonlau, M., & Guenther, N. (2016). Text Mining Using N-Grams. *SSRN Electronic Journal*. http://doi.org/10.2139/ssrn.2759033

van de Schoot, R., Lugtig, P., and Hox, J. (2012). A checklist for testing measurement invariance. Eur. J. Dev. Psychol. 9, 486–492. doi: 10.1080/17405629.2012.686740

Schouten, B., van den Brakel, J., Buelens, B., van der Laan, J., & Klausch, T. (2013). Disentangling mode-specific selection and measurement bias in social surveys. *Social Science Research*, *42*(6), 1555–1570. http://doi.org/10.1016/j.ssresearch.2013.07.005

Schwarz, N., & Strack, F. (1991). *Evaluating one's life: A judgment model of subjective well-being*. *Subjective Well-Being: An Iterdisciplinary Perspective* (Vol. 1988/05). Deutschland. http://doi.org/10.1037//0003-066X.55.1.34

Shaked, D., Williams, M., Evans, M. K., & Zonderman, A. B. (2016). Indicators of subjective social status: Differential associations across race and sex. *SSM - Population Health*, *2*, 700–707. http://doi.org/10.1016/j.ssmph.2016.09.009

Singhapakdi, A., Sirgy, M. J., Lee, D. J., Senasu, K., Yu, G. B., & Nisius, A. M. (2014). Gender disparity in job satisfaction of Western versus Asian managers. *Journal of Business Research*, *67*(6), 1257–1266. http://doi.org/10.1016/j.jbusres.2013.04.004

Smyth, J. D., Dillman, D. A., & Christian, L. M. (2008). *Context Effects in Internet Surveys : New Issues and Evidence*. http://doi.org/10.1093/oxfordhb/9780199561803.013.0027

Smyth, J. D., Dillman, D. A., Christian, L. M., & Mcbride, M. (2009). Open-ended questions in web surveys: Can Increasing the Size of Answer Boxes and Providing Extra Verbal Instructions Improve Response Quality? *Public Opinion Quarterly*, *73*(2), 325–337. http://doi.org/10.1093/poq/nfp029

Spence, R., Bunn, A., Nunn, S., Hosang, G. M., Kagan, L., Fisher, H. L., … Bifulco, A. (2015). Measuring Life Events and Their Association With Clinical Disorder: A Protocol for Development of an Online Approach. *JMIR Research Protocols*, *4*(3), e83. http://doi.org/10.2196/resprot.4085

Spini, D., Bernardi, L., & Oris, M. (2017). Vulnerability Across the Life Course. *Research in Human Development*, *14*(1), 1–4. http://doi.org/10.1080/15427609.2016.1268891

Spini, D., Hanappi, D., Bernardi, L., Oris, M., & Bickel, J.-F. (2013). *Vulnerability across the life course: A theoretical framework and research directions*. *LIVES Working papers* (Vol. 27). http://doi.org/10.12682/lives.2296-1658.2013.27

Springer, K. W., & Hauser, R. M. (2006). An assessment of the construct validity of

Ryff's Scales of Psychological Well-Being: Method, mode, and measurement effects. *Social Science Research*, *35*(4), 1080–1102. http://doi.org/10.1016/j.ssresearch.2005.07.004

StataCorp. 2013. Stata Statistical Software: Release 13. College Station, TX: StataCorp LP.

Stern, M. J., Dillman, D. A., & Smyth, J. D. (2007). Visual Design, Order Effects, and Respondent Characteristics in a Self-Administered Survey. *Survey Research Methods*, *1*(3), 121–138. Retrieved from http://www.surveymethods.org

Stevenson, B., & Wolfers, J. (2013). Subjective well-being and income: Is there any evidence of satiation? In *American Economic Review* (Vol. 103, pp. 598–604). http://doi.org/10.1257/aer.103.3.598

Stone, A. A. (2012). Subjective Wellbeing Assessment for Survey Research, (December). Retrieved from http://www.who.int/healthinfo/sage/SAGE_Meeting_Dec2012_StoneA.pdf?ua=1

Stone, A. A., Mackie, C., Subjective Well-Being and Policy. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK179214/

Struffolino, E., Voorpostel, M., & Bernardi, L. (2015). Self-reported health among lone mothers: Do employment and education matter? *Doi.org*, *2015*(44). http://doi.org/10.12682/lives.2296-1658.2015.44

Sudman, S., & Bradburn, N. M. (1974). *Response effects in surveys; a review and synthesis* (6th ed.). Chicago: Aldine Pub. Co. Retrieved from http://www.worldcat.org/title/response-effects-in-surveys-a-review-and-synthesis/oclc/1021300?page=citation

Sumner, A., & Mallett, R. (2011). Snakes and Ladders, Buffers and Passports:

Rethinking Poverty, Vulnerability and Wellbeing. *Working Papers*, (83). Retrieved from http://econpapers.repec.org/paper/ipcopager/131.htm

Szumilas, M. (2010). Explaining odds ratios. *Journal of the Canadian Academy of Child and Adolescent Psychiatry*, *19*(3), 227–229. http://doi.org/10.1136/bmj.c4414

Thompson, B. (2005). Canonical Correlation Analysis. In *Encyclopedia of statistics in behavioral sciences*. Chichester, UK: John Wiley & Sons, Ltd. http://doi.org/10.1002/0470013192.bsa068

Tourangeau, R., & Smith, T. W. (1996). Asking sensitive questions: the impact of data collection mode, question format, and question context. *Public Opinion QuarterlyQ*, *60*. http://doi.org/10.1086/297751

Tourangeau, R. (2017). Mixing Modes. In P. P. Biemer, E. D. de Leeuw, S. Eckman, B. Edwards, F. Kreute, L. E. Lyberg, … B. West (Eds.), *Total Survey Error in Practice* (pp. 115–132). Hoboken, NJ, USA: John Wiley & Sons, Inc. http://doi.org/10.1002/9781119041702.ch6

Tourangeau, R., Groves, R. M., & Redline, C. D. (2010). Sensitive topics and reluctant respondents: Demonstrating a link between nonresponse bias and measurement error. *Public Opinion Quarterly*, *74*(3), 413–432. http://doi.org/10.1093/poq/nfq004

Tourangeau, R., Rips, L. J., & Rasinski, K. A. (2000). *The psychology of survey response*. Cambridge University Press. Retrieved from https://books.google.ch/books?hl=fr&lr=&id=bjVYdyXXT3oC&oi=fnd&pg=PR11&dq=the+psychology+of+survey+response+tourangeau&ots=ZZ4hJPazuV&sig=uJyQfrxkB1BWWtdeIwR-_5iVWeo#v=onepage&q=the psychology of survey response tourangeau&f=false

Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin*, *133*(5), 859–883. http://doi.org/10.1037/0033-2909.133.5.859

Trappmann, M., Gramlich, T., & Mosthaf, A. (2015). The effect of events between waves on panel attrition. *Survey Research Methods*, *9*(1), 31–43. http://doi.org/10.18148/srm/2015.v9i1.5849

Turner, C. F., Lessler, J. T., & Gfroerer, J. C. (1992). *Survey measurement of drug use : methodological studies*. National Institute on Drug Abuse, Division of Epidemiology and Prevention Research, U.S. Dept. of Health and Human Services, Public Health Service, Alcohol, Drug Abuse, and Mental Health Administration.

Tutz, G., & Berger, M. (2016). Response Styles in Rating Scales: Simultaneous Modeling of Content-Related Effects and the Tendency to Middle or Extreme Categories. *Journal of Educational and Behavioral Statistics*, *41*(3), 239–268. http://doi.org/10.3102/1076998616636850

UCLA, S. C. G. (n.d.-a). Discriminant Function Analysis | SPSS Data Analysis Examples. Retrieved October 15, 2017, from https://stats.idre.ucla.edu/stata/dae/discriminant-function-analysis/

UCLA, S. C. G. (n.d.-b). Ordered Logistic Regression.

Van De Vijver, F. J. R. (2011). Capturing bias in structural equation modeling. In J. B. E. Davidov, P. Schmidt (Ed.), *Cross-cultural analysis. Methods and applications* (pp. 3–34). New York: Routledge.

Van der Meer, P. H. (2014). Gender, Unemployment and Subjective Well-Being: Why Being Unemployed Is Worse for Men than for Women. *Social Indicators Research*, *115*(1), 23–44. http://doi.org/10.1007/s11205-012-0207-5

Van Horn, J. E., Taris, T. W., Schaufeli, W. B., & Schreurs, P. J. G. (2004). The structure of occupational well-being: A study among Dutch teachers. *Journal of Occupational and Organizational Psychology*, *77*(3), 365–375. http://doi.org/10.1348/0963179041752718

Van Vaerenbergh, Y., & Thomas, T. D. (2013). Response styles in survey research: A literature review of antecedents, consequences, and remedies. *International Journal of Public Opinion Research*, *25*(2), 195–217. http://doi.org/10.1093/ijpor/eds021

Vandenberg, R. J., & Lance, C. E. (2000). A Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices, and Recommendations for Organizational Research. *Organizational Research Methods*, *3*(1), 4–70. http://doi.org/10.1177/109442810031002

Vannieuwenhuyze, J., Loosveldt, G., & Molenberghs, G. (2010). A method for evaluating mode effects in mixed-mode surveys. *Public Opinion Quarterly*, *74*(5), 1027–1045. http://doi.org/10.1093/poq/nfq059

Vannieuwenhuyze, J. T. A. (2015). Mode Effects on Variances, Covariances, Standard Deviations, and Correlations. *Journal of Survey Statistics and Methodology*, *3*(3), 296–316. http://doi.org/10.1093/jssam/smv009

Vannieuwenhuyze, J. T. A., & Loosveldt, G. (2012). Evaluating Relative Mode Effects in Mixed-Mode Surveys:: Three Methods to Disentangle Selection and Measurement Effects. *Sociological Methods & Research*, *42*(1), 82–104. http://doi.org/10.1177/0049124112464868

Vannieuwenhuyze, J. T. A., & Loosveldt, G. (2013). Evaluating Relative Mode Effects in Mixed-Mode Surveys: *Sociological Methods & Research*, *42*(1), 82–104. http://doi.org/10.1177/0049124112464868

Vannieuwenhuyze, J. T. A., Loosveldt, G., & Molenberghs, G. (2014). Evaluating mode effects in mixed-mode survey data using covariate adjustment models. *Journal of Official Statistics*, *30*(1), 1–21. http://doi.org/10.2478/jos-2014-0001

Veenhoven, R. (2000). Wellbeing in the welfare state: Level not higher, distribution not more equitable. *Journal of Comparative Policy Analysis: Research and Practice*, *2*(1), 91–125. http://doi.org/10.1080/13876980008412637

Veenhoven, R. (2008). Sociological Theories of Subjective Well-Being. In M. Eid & R. Larsen (Eds.), *The Science of Subjective Well-being: A tribute to Ed Diener* (Guilford P, pp. 44–61). New York: Guilford Publications. Retrieved from https://personal.eur.nl/veenhoven/Pub2000s/2008c-full.pdf

Villar, A., & Fitzgerald, R. (2017). Chapter 16 Using mixed modes in survey research: Evidence from six experiments in the ESS. In *Values and Identities in Europe: Evidence from the European Social Survey* (pp. 259–293). Retrieved from https://www.europeansocialsurvey.org/docs/methodology/Mixed-modes-in-the-ESS-6-experiments_in-Breen-et-al-2017.pdf

Walsh, E., & Brinker, J. K. (2016). Short and Sweet? Length and Informative Content of Open-Ended Responses Using SMS as a Research Mode. *Journal of Computer-Mediated Communication*, *21*(1), 87–100. http://doi.org/10.1111/jcc4.12146

Weijters, B., Schillewaert, N., & Geuens, M. (2008). Assessing response styles across modes of data collection. *Journal of the Academy of Marketing Science*, *36*(3), 409–422. http://doi.org/10.1007/s11747-007-0077-6

Weiss, H. M. (2002, June 1). Deconstructing job satisfaction. Separating evaluations, beliefs and affective experiences. *Human Resource Management Review*. JAI. http://doi.org/10.1016/S1053-4822(02)00045-1

Weiss, H. M., & Merlo, K. L. (2015). Job Satisfaction. In *International Encyclopedia of the Social & Behavioral Sciences* (Vol. 12, pp. 833–838). http://doi.org/10.1016/B978-0-08-097086-8.22029-1

Weiting, N., & Diener, E. (2014). What matters to the rich and the poor? Subjective well-being, financial satisfaction, and postmaterialist needs across the world. *Journal of Personality and Social Psychology*, *107*(2), 326–338. http://doi.org/10.1037/a0036856

Williams, R. (2008). Ordinal regression models: Problems, solutions, and problems with the solutions. In German Stata User Group Meetings, June (Vol. 27, p. 2008). Retrieved from https://www.stata.com/meeting/germany08/GSUG2008.pdf

Williams, R. (2017). Scalar Measures of Fit: Pseudo R2 and Information Measures (AIC & BIC) . Retrieved from https://www3.nd.edu/~rwilliam/stats3/L05.pdf

Wu, A. D., Li, Z., & Zumbo, B. D. (2007). Decoding the meaning of factorial invariance and updating the practice of multi-group confirmatory factor analysis: A demonstration with TIMSS data. *Practical Assessment, Research & Evaluation*, *12*(3), 1–26.

Yan, T. (2015). Assessing Quality of Answers to a Global Subjective Well-being Question Through Response Times. *Survey Research Methods*, *9*(2), 101–109. http://doi.org/10.18148/srm/2015.v9i2.6107

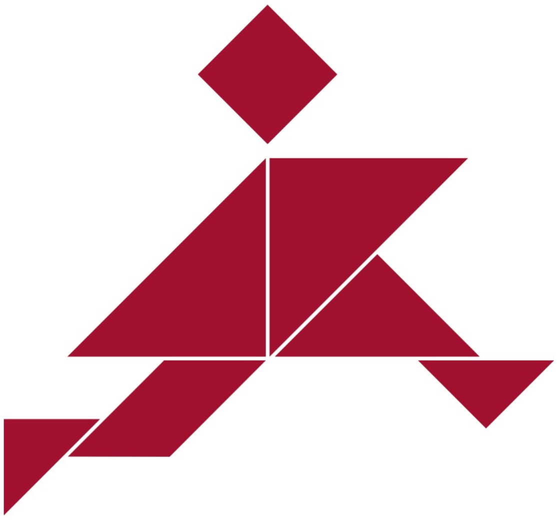
Yong, G. A., & Pearce, S. (2013). A Beginner's Guide to Factor Analysis: Focusing on Exploratory Factor Analysis. *Tutorials in Quantitative Methods for Psychology*, *9*(2), 79–94.

ANNEXE A: QUESTIONNAIRE

Bien-être et mal-être en Suisse romande

QUESTIONNAIRE

Une initiative pour comprendre comment la vie se passe pour les habitants de notre région



Merci de retourner le questionnaire rempli pour une saisie anonyme à:

M.I.S. Trend S.A.
Pont Bessières 3
CH-1005 Lausanne

Le code-barres sur la couverture de ce questionnaire contient un numéro d'identification unique, pour que nous puissions supprimer votre nom de la liste de distribution après le retour de votre questionnaire. La liste de noms sera détruite pour que les individus ne puissent jamais être liés avec les résultats de quelque manière que ce soit. Les données issues de vos réponses seront totalement anonymes.

- Merci de prendre une demi-heure de votre temps pour compléter ce questionnaire.
- Veuillez utiliser un stylo foncé et bien lisible.
- Cochez la réponse sélectionnée avec une croix dans le cercle correspondant: ⊗
- En cas d'erreur, veuillez colorier tout le cercle coché et faire une nouvelle croix dans le cercle qui correspond à la bonne réponse: ● ⊗
- Aux endroits où vous trouvez un texte bleu (p.ex. aller à la Q.25), nous vous prions de passer à la question indiquée.
- S'il n'y a pas d'autre indication, vous devez choisir une seule réponse. Veuillez bien respecter cette règle, s'il vous plaît. Par contre, il y a quelques questions où la possibilité de cocher plusieurs réponses est indiquée explicitement.

Les premières questions portent sur la société en général.

1. Diriez-vous que l'on peut généralement faire confiance à la plupart des personnes ou que l'on n'est jamais trop prudent dans ses contacts avec les autres gens ?

Merci de choisir un chiffre entre 0 et 10, où 0 signifie que l'on n'est jamais trop prudent et 10 signifie que l'on peut faire confiance à la plupart des personnes. Les chiffres intermédiaires vous permettent de nuancer votre jugement.

On n'est jamais trop prudent										**On peut faire confiance à la plupart des personnes**
0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

2. Vous-même ou un membre de votre ménage a-t-il été victime d'un cambriolage ou d'une agression ces 5 dernières années ?

○1 oui
○2 non

3. Dans quelle mesure vous sentez-vous - ou vous sentiriez-vous - en sécurité seul/e le soir à pied dans le quartier où vous habitez ? Vous sentez-vous - ou vous sentiriez-vous...

○1 tout à fait en sécurité
○2 en sécurité
○3 en insécurité
○4 tout à fait en insécurité

4. Quel intérêt avez-vous pour la politique ?

○1 très intéressé
○2 assez intéressé
○3 peu intéressé
○4 pas du tout intéressé

5. Il y a plusieurs moyens afin d'essayer d'améliorer la situation en Suisse ou d'éviter que les choses ne se dégradent. Durant les 12 derniers mois, avez-vous fait l'une des actions suivantes :

Une réponse par ligne

	Oui	**Non**
5a. Avez-vous signé une pétition ?	○1	○2
5b. Avez-vous boycotté certains produits ?	○1	○2
5c. Avez-vous voté lors d'une votation ou d'une élection ?	○1	○2

6. Durant les 12 derniers mois, avez-vous activement participé à l'une de ces organisations ou associations ?

Une réponse par ligne.

	Oui	Non
6a. Un club de sport ou d'activités à l'extérieur ?	○1	○2
6b. Une organisation culturelle ou liée à un hobby ?	○1	○2
6c. Une organisation d'aide humanitaire, de défense des droits de l'homme, des minorités ou des immigrants	○1	○2
6d. Une organisation pour la protection de l'environnement, des animaux ou pour la paix	○1	○2
6e. Une organisation religieuse ou liée à une église	○1	○2
6f. Un groupe de jeunes, une association de personnes âgées, de femmes, une amicale	○1	○2
6g. Une autre association ou organisation bénévole	○1	○2

7. Dans tous les pays, il existe des divergences –voire des conflits- entre les divers groupes sociaux. Quelle est l'importance en Suisse, selon vous, des divergences entre les groupes sociaux suivants :

Entre hommes et femmes

○1 très importante
○2 assez importante
○3 peu importante
○4 pas du tout importante

8. Entre Suisses et étrangers

○1 très importante
○2 assez importante
○3 peu importante
○4 pas du tout importante

9. Entre riches et pauvres

○1 très importante
○2 assez importante
○3 peu importante
○4 pas du tout importante

10. Entre Romands et Alémaniques

○1 très importante
○2 assez importante
○3 peu importante
○4 pas du tout importante

11. Dans quelle mesure pensez-vous que la Suisse doit autoriser des gens d'une origine ethnique différente de la plupart des Suisses à venir vivre ici ? Pensez-vous que la Suisse ...

- ○1 doit autoriser un grand nombre d'entre eux à venir vivre ici
- ○2 doit autoriser certains d'entre eux
- ○3 ne doit autoriser que peu d'entre eux
- ○4 ne doit autoriser aucun d'entre eux

12. Diriez-vous que c'est généralement bon ou mauvais pour l'économie suisse que des gens d'autres pays viennent vivre ici ?

Merci de choisir un chiffre entre 0 et 10, où 0 signifie 'Mauvais pour l'économie' et 10 signifie 'Bon pour l'économie'. Les chiffres intermédiaires vous permettent de nuancer votre jugement.

Mauvais pour l'économie										**Bon pour l'économie**
0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

13. Diriez-vous que la vie culturelle en Suisse est généralement appauvrie ou enrichie par les gens d'autres pays qui viennent vivre ici ?

Merci de choisir un chiffre entre 0 et 10, où 0 signifie La vie culturelle est appauvrie' et 10 signifie 'La vie culturelle est enrichie'. Les chiffres intermédiaires vous permettent de nuancer votre jugement.

La vie culturelle est appauvrie										**La vie culturelle est enrichie**
0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Maintenant quelques questions à propos de vous et de votre vie.

14. Dans l'ensemble, dans quelle mesure êtes-vous satisfait/e de votre vie actuelle?

Merci de choisir un chiffre entre 0 et 10, où 0 signifie 'Très insatisfait/e' et 10 signifie 'Très satisfait/e'
Les chiffres intermédiaires vous permettent de nuancer votre jugement.

Très insatisfait/e										**Très satisfait/e**
0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

15. Quel est votre état de santé en général ?

○1 très bon
○2 bon
○3 passable
○4 mauvais
○5 très mauvais

16. Etes-vous gêné/e d'une quelconque manière dans vos activités quotidiennes par une maladie de longue durée, un handicap, une infirmité ou un problème de santé mentale ?

SI OUI, ce problème vous gêne-t-il fortement ou dans une certaine mesure seulement ?

○1 oui, fortement
○2 oui, dans une certaine mesure
○3 non, pas gêné(e) du tout

17. Avez-vous bu de l'alcool au cours des 7 derniers jours ?

○1 oui **Aller à la Q. 18**
○2 non **Aller à la Q. 19**

Si OUI, merci d'indiquer le nombre de jour(s): _____

18. Au total sur ces 7 derniers jours, environ combien de verres d'alcool avez-vous bu ?

(Si vous ne savez pas, veuillez donner une approximation)

Merci d'indiquer le nombre de verres : _____

19. Est-ce que vous fumez des cigarettes ?

○1 oui **Aller à la Q. 20**
○2 non **Aller à la Q. 21**

20. Environ combien de cigarettes fumez-vous par jour ?

(Si vous ne savez pas, veuillez donner une approximation)

Merci d'indiquer le nombre de cigarettes : ______

21. Combien de jours parmi les 7 derniers avez-vous été actif/ve physiquement pendant au moins 20 minutes d'affilée ?

Vous pouvez inclure les tâches domestiques, comme le ménage ou le jardinage, à condition que l'activité ait duré au moins vingt minutes.

- ○ 0 aucun jour
- ○ 1 un jour
- ○ 2 deux jours
- ○ 3 trois jours
- ○ 4 quatre jours
- ○ 5 cinq jours
- ○ 6 six jours
- ○ 7 sept jours

22. Quel est votre poids actuel ?

(Si vous ne savez pas, veuillez donner une approximation)

Merci d'indiquer votre poids en kilogrammes (p.ex. 78) : ______ Kg

23. Quelle est votre taille ?

(Si vous ne savez pas, veuillez donner une approximation)

Merci d'indiquer votre taille en centimètres (p.ex. 178) : ______ cm

Maintenant nous voulons vous poser quelques questions sur la manière dont vous vous voyez et percevez votre vie.

24. Tout bien considéré, dans quelle mesure diriez-vous que vous êtes heureux ?

Merci de choisir un chiffre entre 0 et 10, où 0 signifie 'Très malheureux' et 10 signifie 'Très heureux'. Les chiffres intermédiaires vous permettent de nuancer votre jugement.

Très malheureux										**Très heureux**
0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Dans quelle mesure êtes-vous d'accord ou non avec les propositions suivantes :

25. Je suis toujours optimiste quand je pense à mon avenir.

- O 1 tout à fait d'accord
- O 2 plutôt d'accord
- O 3 ni d'accord, ni en désaccord
- O 4 plutôt en désaccord
- O 5 tout à fait en désaccord

26. En général, j'ai une image très positive de moi-même.

- O 1 tout à fait d'accord
- O 2 plutôt d'accord
- O 3 ni d'accord, ni en désaccord
- O 4 plutôt en désaccord
- O 5 tout à fait en désaccord

27. Je me sens libre de décider moi-même comment vivre ma vie.

- O 1 tout à fait d'accord
- O 2 plutôt d'accord
- O 3 ni d'accord, ni en désaccord
- O 4 plutôt en désaccord
- O 5 tout à fait en désaccord

28. La plupart du temps, ce que je fais me donne un sentiment de réussite.

- O 1 tout à fait d'accord
- O 2 plutôt d'accord
- O 3 ni d'accord, ni en désaccord
- O 4 plutôt en désaccord
- O 5 tout à fait en désaccord

Maintenant quelques questions sur les sentiments et sensations que vous avez pu éprouver la semaine dernière.

29. Dans quelle mesure vous est-il arrivé la semaine dernière de vous sentir déprimé/e ?

- O 1 à aucun moment ou presque
- O 2 de temps en temps
- O 3 la plupart du temps
- O 4 tout le temps ou presque

30. Dans quelle mesure vous est-il arrivé la semaine dernière d'avoir un sommeil agité ?

- O 1 à aucun moment ou presque
- O 2 de temps en temps
- O 3 la plupart du temps
- O 4 tout le temps ou presque

31. Dans quelle mesure vous est-il arrivé la semaine dernière de vous sentir seul/e ?

- O 1 à aucun moment ou presque
- O 2 de temps en temps
- O 3 la plupart du temps
- O 4 tout le temps ou presque

32. Dans quelle mesure vous est-il arrivé la semaine dernière de vous sentir inquiet/ète ?

- O 1 à aucun moment ou presque
- O 2 de temps en temps
- O 3 la plupart du temps
- O 4 tout le temps ou presque

33. Dans quelle mesure prenez-vous le temps de faire ce dont vous avez vraiment envie ?

Merci de choisir un chiffre entre 0 et 10, où 0 signifie 'Pas du tout' et 10 signifie 'Complètement'. Les chiffres intermédiaires vous permettent de nuancer votre jugement.

Pas du tout										**Complètement**
0	1	2	3	4	5	6	7	8	9	10
O	O	O	O	O	O	O	O	O	O	O

Voici une série de questions portant sur votre vécu et vos pensées au cours du dernier mois. Veuillez indiquer dans quelle mesure chacune s'applique à vous.

34. Au cours du dernier mois, avez-vous eu le sentiment de n'avoir aucune prise, aucun contrôle, sur des aspects importants de votre vie ?

- O 1 jamais
- O 2 presque jamais
- O 3 de temps en temps
- O 4 assez souvent
- O 5 très souvent

35. Au cours du dernier mois, avez-vous eu confiance en votre capacité à surmonter vos problèmes personnels?

- O 1 jamais
- O 2 presque jamais
- O 3 de temps en temps
- O 4 assez souvent
- O 5 très souvent

36. Au cours du dernier mois, avez-vous eu le sentiment que tout allait bien pour vous ? (ou allait dans votre sens)

- O 1 jamais
- O 2 presque jamais
- O 3 de temps en temps
- O 4 assez souvent
- O 5 très souvent

37. Et au cours du dernier mois, avez-vous eu le sentiment que les difficultés s'accumulaient tellement que vous ne parviendriez jamais à les surmonter ?

- ○ 1 jamais
- ○ 2 presque jamais
- ○ 3 de temps en temps
- ○ 4 assez souvent
- ○ 5 très souvent

38. À quelle fréquence rencontrez-vous des amis, de la famille ou des collègues <u>en dehors du travail</u> ?

- ○ 1 jamais
- ○ 2 moins d'une fois par mois
- ○ 3 une fois par mois
- ○ 4 plusieurs fois par mois
- ○ 5 une fois par semaine
- ○ 6 plusieurs fois par semaine
- ○ 7 chaque jour

39. Avez-vous des personnes avec qui vous pouvez parler de sujets intimes et personnels, et si oui combien ?

- ○ 1 aucune
- ○ 2 1
- ○ 3 2
- ○ 4 3
- ○ 5 4-6
- ○ 6 7-9
- ○ 7 10 ou plus

40. En vous comparant à d'autres personnes de votre âge, à quelle fréquence prenez-vous part à des activités sociales ?

- ○ 1 beaucoup moins souvent que la plupart
- ○ 2 moins souvent que la plupart
- ○ 3 à peu près la même chose
- ○ 4 plus souvent que la plupart
- ○ 5 beaucoup plus souvent que la plupart

41. Dans quelle mesure recevez-vous le soutien de vos proches en cas de besoin ?

Merci de choisir un chiffre entre 0 et 6, où 0 signifie 'Pas du tout' et 6 signifie 'Complètement'.
Les chiffres intermédiaires vous permettent de nuancer votre jugement.

Pas du tout						**Complètement**
0	1	2	3	4	5	6
○	○	○	○	○	○	○

42. Et dans quelle mesure apportez-vous du soutien à vos proches en cas de besoin ?

Merci de choisir un chiffre entre 0 et 6, où 0 signifie 'Pas du tout' et 6 signifie 'Complètement'. Les chiffres intermédiaires vous permettent de nuancer votre jugement.

Pas du tout						**Complètement**
0	1	2	3	4	5	6
○	○	○	○	○	○	○

Dans quelle mesure les propositions suivantes vous correspondent-elles ?

43. Je me moque de ce que les autres pensent réellement de moi.

Merci d'indiquer un chiffre entre 0 et 6, où 0 signifie que cette proposition ne correspond pas du tout à vous, et 6 signifie qu'elle correspond totalement à vous. Les chiffres intermédiaires vous permettent de nuancer votre jugement.

Ne me correspond pas du tout						**Me correspond totalement**
0	1	2	3	4	5	6
○	○	○	○	○	○	○

44. Une fois que je me suis décidé/e, les autres arrivent rarement à me faire changer d'opinion.

Merci d'indiquer un chiffre entre 0 et 6, où 0 signifie que cette proposition ne correspond pas du tout à vous, et 6 signifie qu'elle correspond totalement à vous. Les chiffres intermédiaires vous permettent de nuancer votre jugement.

Ne me correspond pas du tout						**Me correspond totalement**
0	1	2	3	4	5	6
○	○	○	○	○	○	○

Maintenant quelques questions sur les choses qui vous sont arrivées dans votre vie.

45. Tout d'abord, nous aimerions aborder les événements qui ont marqué votre vie, puis nous aimerions savoir si son impact a été positif ou négatif.

Quels ont été les 3 événements les plus marquants de votre vie ?

Evénement 1 : Merci d'inscrire un événement dans l'encadré. Veuillez donner autant de précisions que possible.

46. **Evénement 2 :** Merci d'inscrire un événement dans l'encadré. Veuillez donner autant de précisions que possible.

47. **Evénement 3 :** Merci d'inscrire un événement dans l'encadré. Veuillez donner autant de précisions que possible.

48. **Est-ce que l'impact de l'événement 1 sur votre vie a été plutôt positif ou plutôt négatif ?**

- ○ 1 très négatif
- ○ 2 plutôt négatif
- ○ 3 tant négatif que positif
- ○ 4 plutôt positif
- ○ 5 très positif

49. **Est-ce que l'impact de l'événement 2 sur votre vie a été plutôt positif ou plutôt négatif ?**

- ○ 1 très négatif
- ○ 2 plutôt négatif
- ○ 3 tant négatif que positif
- ○ 4 plutôt positif
- ○ 5 très positif

50. Est-ce que l'impact de l'événement 3 sur votre vie a été plutôt positif ou plutôt négatif ?

- ○ 1 très négatif
- ○ 2 plutôt négatif
- ○ 3 tant négatif que positif
- ○ 4 plutôt positif
- ○ 5 très positif

51. Avez-vous connu ou jamais connu personnellement les événements ou problèmes suivants ?

Veuillez répondre par 'Oui, connu', ou 'Non, jamais connu' pour chacun.

	Oui, connu	Non, jamais connu
51a. grave maladie ou accident	○ 1	○ 2
51b. graves difficultés financières, dettes importantes, faillite	○ 1	○ 2
51c. exil, déracinement, migration forcée	○ 1	○ 2
51d. décès d'un enfant, du(de la) conjoint(e)/ compagnon(compagne)	○ 1	○ 2
51e.· divorce, séparation / problèmes sérieux avec votre conjoint(e) ou compagnon(compagne)	○ 1	○ 2

Merci de répondre, dans cette seconde partie, à quelques questions sur votre situation sociale.

1. Êtes-vous un homme ou une femme ?

O 1 un homme
O 2 une femme

2. En quelle année êtes-vous né(e) ?

Merci d'indiquer votre année de naissance : ___ ___ ___ ___

3. Est-ce que vous-même vous vous sentez appartenir à une religion ou confession particulière ?

O 1 oui
O 2 non

Si oui, laquelle ?

O 0 sans religion
O 1 catholique
O 2 protestant
O 3 orthodoxe
O 4 autre religion chrétienne
O 5 juif
O 6 musulman
O 7 bouddhiste
O 8 hindouiste
O 9 autre religion asiatique
O 10 autre religion

Si autre religion, merci d'indiquer à quelle religion vous vous sentez appartenir :

____

4. Êtes-vous citoyen/ne Suisse ?

O 1 oui
O 2 non

5. Quelle(s) autre(s) nationalité(s) avez-vous ?

Merci d'indiquer le(s) pays de votre nationalité : ____________________

O 666 Aucune autre nationalité

6. Dans quel pays êtes-vous né/e ?

Merci d'indiquer dans quel pays vous êtes né/e ______________________

7. **Si vous n'êtes pas né(e) en Suisse, en quelle année êtes-vous venu/e vivre en Suisse pour la première fois ?**

Merci d'indiquer l'année avec 4 chiffres (p.ex. 1971) : ___ ___ ___ ___

○ 66 (ne s'applique pas)
○ 88 (ne sait pas)

8. **Quelle(s) langue(s) parlez-vous le plus souvent à la maison ?**

Merci d'inscrire la/les langue(s) que vous parlez à la maison ici :

____

9. **Dans quel pays votre père est-il né ?**

Merci d'inscrire le pays de naissance de votre père : ___________________________

10. **Dans quel pays votre mère est-elle née ?**

Merci d'inscrire le pays de naissance de votre mère : ___________________________

11. **Considérez-vous que vous appartenez à un groupe discriminé en Suisse ?**

○ 1 oui **Aller à la Q. 12**
○ 2 non **Aller à la Q. 13**

12. **Si oui, pour quelle(s) raison(s) ce groupe est-il discriminé ?**

____
____
____
____

13. Quel est le plus haut niveau de formation que vous avez terminé ?

Veuillez cocher une seule case

Ecole primaire	O 1 O 2	Ecole primaire inachevée Ecole primaire (4 à 6 ans de scolarité)
Scolarité obligatoire	O 3 O 4	Cycle d'orientation, école secondaire (et école primaire de 8-9 ans) 10. année, préapprentissage, cours préparatoire, école préprofessionnelle
Ecoles de culture générale (ECG)	O 5	Ecoles de culture générale (3 ans, certificat d'ECG, maturité spécialisée), Ecoles de degré diplôme (EDD), Ecole de commerce
Ecole de maturité	O 6 O 7 O 8	Maturité gymnasiale, Gymnase, Collège Ecole normale, Etudes pédagogiques (niveau préscolaire et primaire) Maturité professionnelle
Formation professionnelle	O 9 O 10 O 11 O 12 O 13 O 14	Formation professionnelle initiale (Attestation fédérale de formation professionnelle, Apprentissage court (2 ans), Ecole commerciale (1 an), Ecole de formation générale (1-2 ans) Apprentissage 3-4 ans (CFC) en entreprise formatrice ou en école professionnelle Deuxième apprentissage ou apprentissage en tant que deuxième formation Maîtrise professionnelle, brevet fédéral et autres examens professionnels supérieurs Diplôme ou postgrade d'une école professionnelle supérieure, p.ex. dans les domaines techniques, administration, santé, travail social, arts appliqués Diplôme ou postgrade d'une des écoles supérieures suivantes: écoles d'ingénieurs ETS écoles supérieures de cadres pour l'économie et l'administration (ESCEA) écoles supérieures d'arts appliqués (ESAA) écoles supérieures d'économie familiale (ESEF) école hôtelière de Lausanne (EHL, diplômes décernés en 1998, 1999 et 2000)
Hautes écoles spécialisées	O 15	Bachelor
(HES), Hautes écoles pédagogiques (HEP)	O 16	Master, diplôme, postgrade
Hautes écoles universitaires, Ecoles polytechniques fédérales (EPF)	O 17 O 18 O 19 O 20	Bachelor, licence en 3 ans Licence exigeant 4 ans ou plus Master, diplôme, postgrade Doctorat, PhD
Autre (veuillez préciser) :	O 77	____

14. Avez-vous actuellement un travail rémunéré, en avez-vous eu un dans le passé ou n'avez-vous jamais eu de travail rémunéré ?

O 1	j'ai actuellement un travail rémunéré	**Aller à la Q. 15**
O 2	je n'ai actuellement pas de travail rémunéré mais j'en ai eu un autrefois	**Aller à la Q. 15**
O 3	je n'ai jamais eu de travail rémunéré	**Aller à la Q. 27**

15. Combien d'heures travaillez-vous / avez-vous travaillé habituellement par semaine, y compris les heures supplémentaires, payées et non payées?

Si vous avez / aviez plusieurs emplois en même temps, veuillez répondre, pour cette question et les suivantes, à propos de celui qui vous occupe / occupait le plus d'heures. S'ils sont / étaient à égalité de temps, veuillez répondre au sujet de celui qui est / était le mieux rémunéré.

Merci d'indiquer le nombre d'heures: ___ ___

O 88 (ne sait pas)

16. Dans votre emploi principal êtes-vous (étiez-vous)

O 1	employé/e	**Aller à la Q. 18**
O 2	indépendant/e	**Aller à la Q. 17**
O 3	ou collaborateur/trice dans l'entreprise familiale ?	**Aller à la Q. 18**

17. Si vous êtes (vous étiez) indépendant/e, combien d'employés av(i)ez-vous ?

Veuillez noter le nombre d'employés (sans vous compter vous-même) : ___ ___ ___ ___ employé(s)

Puis aller à la Q. 20

18. Dans votre emploi principal, av(i)ez-vous la responsabilité de superviser le travail d'autres employés ?

O 1	oui	**Aller à la Q. 19**
O 2	non	**Aller à la Q. 20**

19. Si vous av(i)ez la responsabilité de superviser le travail d'autres employés, de combien de personnes êtes-vous (étiez-vous) responsable ?

Veuillez noter le nombre de personnes : ___ ___ ___ ___ employé(s)

20. Dans quel type d'organisation travaillez-vous ou avez-vous travaillé ?

O 1 administration publique (Confédération, canton ou commune)
O 2 autre secteur public (comme écoles et hôpitaux)
O 3 entreprise publique
O 4 entreprise privée
O 5 indépendant
O 6 autre

Si 'autre', merci de préciser lequel :

____

____

21. Quel est/était le nom ou le titre de votre emploi principal et que faites-vous / faisiez-vous la plus grande partie du temps?

Merci de noter avec un maximum de précision

____

____

____

____

____

22. <u>Si vous avez actuellement un travail rémunéré</u>, dans quelle mesure, tout bien considéré, êtes-vous satisfait/e de votre travail actuel ?

Merci de choisir un chiffre entre 0 et 10, en sachant que 0 signifie très insatisfait/e et 10 très satisfait/e. Les chiffres intermédiaires vous permettent de nuancer votre jugement.

Très insatisfait/e										**Très satisfait/e**
0	1	2	3	4	5	6	7	8	9	10
O	O	O	O	O	O	O	O	O	O	O

O 66 je n'ai pas actuellement un travail rémunéré **[Aller à la Q. 27]**

23. Si vous avez actuellement un travail rémunéré, dans quelle mesure êtes-vous satisfait/e de la répartition de votre temps entre votre travail rémunéré et les autres aspects de votre vie ?

Merci de choisir un chiffre entre 0 et 10, en sachant que où 0 signifie très insatisfait/e et 10 très satisfait/e. Les chiffres intermédiaires vous permettent de nuancer votre jugement.

Très insatisfait/e										Très satisfait/e
0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

24. Si vous avez actuellement un travail rémunéré, dans quelle mesure vous arrive-t-il de trouver votre travail intéressant ?

Merci de choisir un chiffre entre 0 et 6, en sachant que où 0 signifie 'à aucun moment' et 6 signifie 'tout le temps'. Les chiffres intermédiaires vous permettent de nuancer votre jugement.

A aucun moment						Tout le temps
0	1	2	3	4	5	6
○	○	○	○	○	○	○

25. Si vous avez actuellement un travail rémunéré, dans quelle mesure vous arrive-t-il de trouver votre travail stressant ?

Merci de choisir un chiffre entre 0 et 6, en sachant que où 0 signifie 'à aucun moment' et 6 signifie 'tout le temps'.

A aucun moment						Tout le temps
0	1	2	3	4	5	6
○	○	○	○	○	○	○

26. Si vous avez actuellement un travail rémunéré, d'après vous, quel est le risque que vous perdiez votre emploi au cours des 12 prochains mois ?

○ 1 très probable
○ 2 plutôt probable
○ 3 plutôt pas probable
○ 4 pas du tout probable

27. Laquelle de ces descriptions décrit le mieux ce que vous avez fait ces 7 derniers jours ?

Une seule réponse

- O 1 travail rémunéré (ou interruption temporaire, en congé) (employé, indépendant, collaborateur d'une entreprise familiale)
- O 2 en formation (non payée par l'employeur), même si vous êtes actuellement en vacances
- O 3 sans travail mais cherchant activement un emploi
- O 4 sans travail et voulant trouver un emploi mais sans le chercher activement
- O 5 malade ou handicapé/e de manière durable
- O 6 retraité/e
- O 7 service militaire ou service civil
- O 8 travail ménager, s'occuper des enfants ou d'autres personnes
- O autre

Si 'autre', merci de préciser :

✍ ____

____

28. Avez-vous déjà été sans emploi et à la recherche d'un travail pendant une période de plus de trois mois ?

- O 1 oui **Aller à la Q. 29**
- O 2 non **Aller à la Q. 31**

29. L'une ou l'autre de ces périodes a-t-elle duré 12 mois ou plus ?

- O 1 oui
- O 2 non

30. Et l'une ou l'autre de ces périodes a-t-elle eu lieu ces 5 dernières années ?

- O 1 oui
- O 2 non

31. Avez-vous actuellement un/e époux/épouse ou un partenaire régulier et, si oui, partagez-vous le même domicile ?

O 1	oui, j'ai un/e époux/épouse ou un partenaire régulier et nous partageons le même domicile	**Aller à la Q. 32**
O 2	oui, j'ai un/e époux/épouse ou un partenaire régulier mais nous ne partageons pas le même domicile	**Aller à la Q. 32**
O 3	non, je n'ai pas d'époux/épouse ou partenaire régulier	**Aller à la Q. 38**

32. <u>Si vous avez un/e époux (une épouse) ou un partenaire régulier,</u> quel est le plus haut niveau de formation que votre époux(se)/partenaire ait achevé ?

Veuillez cocher une seule case

Ecole primaire	O 1 O 2	Ecole primaire inachevée Ecole primaire (4 à 6 ans de scolarité)
Scolarité obligatoire	O 3 O 4	Cycle d'orientation, école secondaire (et école primaire de 8-9 ans) 10. année, préapprentissage, cours préparatoire, école préprofessionnelle
Ecoles de culture générale (ECG)	O 5	Ecoles de culture générale (3 ans, certificat d'ECG, maturité spécialisée), Ecoles de degré diplôme (EDD), Ecole de commerce
Ecole de maturité	O 6 O 7 O 8	Maturité gymnasiale, Gymnase, Collège Ecole normale, Etudes pédagogiques (niveau préscolaire et primaire) Maturité professionnelle
Formation professionnelle	O 9 O 10 O 11 O 12 O 13 O 14	Formation professionnelle initiale (Attestation fédérale de formation professionnelle, Apprentissage court (2 ans), Ecole commerciale (1 an), Ecole de formation générale (1-2 ans) Apprentissage 3-4 ans (CFC) en entreprise formatrice ou en école professionnelle Deuxième apprentissage ou apprentissage en tant que deuxième formation Maîtrise professionnelle, brevet fédéral et autres examens professionnels supérieurs Diplôme ou postgrade d'une école professionnelle supérieure, p.ex. dans les domaines techniques, administration, santé, travail social, arts appliqués Diplôme ou postgrade d'une des écoles supérieures suivantes: écoles d'ingénieurs ETS écoles supérieures de cadres pour l'économie et l'administration (ESCEA) écoles supérieures d'arts appliqués (ESAA) écoles supérieures d'économie familiale (ESEF) école hôtelière de Lausanne (EHL, diplômes décernés en 1998, 1999 et 2000)
Hautes écoles spécialisées	O 15	Bachelor
(HES), Hautes écoles pédagogiques (HEP)	O 16	Master, diplôme, postgrade
Hautes écoles universitaires, Ecoles polytechniques fédérales (EPF)	O 17 O 18 O 19 O 20	Bachelor, licence en 3 ans Licence exigeant 4 ans ou plus Master, diplôme, postgrade Doctorat, PhD
Autre (veuillez préciser) :	O 77	____

Les questions suivantes reviennent sur le travail de votre époux/épouse ou partenaire. Par travail, nous entendons un travail procurant un revenu, comme employé, indépendant ou collaborateur de l'entreprise familiale pendant au moins une heure par semaine. Si il/elle ne travaille pas à cause de vacances, de maladie, ou d'un congé parental, merci de répondre en pensant à sa condition de travail habituelle.

33. Votre époux/épouse ou partenaire, a-t-il/elle actuellement un travail rémunéré, en a-t-il/elle eu un dans le passé ou n'a-t-il/elle jamais eu de travail rémunéré ?

O 1 il/elle a actuellement un travail rémunéré **Aller à la Q. 34**
O 2 il/elle n'a actuellement pas de travail **Aller à la Q. 34**
O 3 il/elle n'a jamais eu de travail rémunéré **Aller à la Q. 37**

O 8 (ne sait pas) **Aller à la Q. 37**

34. Si votre époux/épouse ou partenaire a actuellement un travail rémunéré, combien d'heures travaille-t-il/elle habituellement par semaine (dans son emploi principal) ? Y compris les heures supplémentaires payées et non payées.

Merci d'indiquer le nombre d'heures: ___ ___

O 88 (ne sait pas)

35. Dans son travail principal, votre époux/épouse ou partenaire régulier est-il/elle (était-il/elle) ...

O 1 employé/e
O 2 indépendant/e
O 2 ou collaborateur/trice dans l'entreprise familiale ?

O 88 (ne sait pas)

36. Quel est/était le nom ou le titre de son emploi principal et que fait-il/elle / faisait-il/elle la plus grande partie du temps?

Merci de noter avec un maximum de précision

___

___

___

___

O 88 (ne sait pas)

37. Laquelle des descriptions sur cette carte décrit le mieux sa situation (durant les sept derniers jours) ?

Une seule réponse

- O 1 travail rémunéré (ou interruption temporaire, en congé) (employé, indépendant, collaborateur d'une entreprise familiale)
- O 2 en formation (non payée par l'employeur), même si vous êtes actuellement en vacances
- O 3 sans travail mais cherchant activement un emploi
- O 4 sans travail et voulant trouver un emploi mais sans le chercher activement
- O 5 malade ou handicapé/e de manière durable
- O 6 retraité/e
- O 7 service militaire ou service civil
- O 8 travail ménager, s'occuper des enfants ou d'autres personnes
- O 9 autre

Si 'Autre', merci de préciser :

____

____

- O 88 (ne sait pas)

38. En vous comptant, combien de personnes, enfants y compris, habitent dans votre ménage ?

Merci d’indiquer le nombre de personnes, enfants y compris: ___ ___

39. Parmi ceux-ci, combien sont des enfants âgés entre 6 et 17 ans ?

Merci d’indiquer le nombre d'enfants âgés entre 6 et 17 ans : ___ ___

40. Et combien sont des enfants de moins de 6 ans ?

Merci d’indiquer le nombre d'enfants de moins de 6 ans: ___ ___

41. Parmi les catégories suivantes, laquelle correspond à votre état civil officiel actuel ?

- O 1 marié/e
- O 2 en partenariat enregistré (fédéral)
- O 3 séparé/e légalement (mais encore marié/e, lié/e par un partenariat enregistré)
- O 4 divorcé/e, partenariat enregistré dissout
- O 5 Veuf/ve, partenaire enregistré/e décédé/e
- O 6 célibataire, JAMAIS marié/e ni lié/e par un partenariat enregistré

42. Les revenus des gens peuvent provenir de différentes sources telles que salaires, rentes, allocations sociales, revenus d'épargne ou d'investissement et ainsi de suite.

En cumulant toutes les sources de votre revenu, indiquez s'il vous plaît VOTRE REVENU PERSONNEL NET. Si vous ne connaissez pas le chiffre exact, veuillez donner une approximation.

Utilisez la partie de la liste que vous connaissez le mieux : revenu mensuel ou annuel.

	Approximation MENSUELLE	Approximation ANNUELLE
O 00	Aucun revenu personnel, d'aucune source	
O 01	Moins de CHF 1'100	Moins de CHF 13'000
O 02	CHF 1'100 à moins de CHF 1'800	CHF 13'000 à moins de CHF 21'500
O 03	CHF 1'800 à moins de CHF 2'700	CHF 21'500 à moins de CHF 32'000
O 04	CHF 2'700 à moins de CHF 2'900	CHF 32'000 à moins de CHF 34'500
O 05	CHF 2'900 à moins de CHF 3'600	CHF 34'500 à moins de CHF 43'000
O 06	CHF 3'600 à moins de CHF 4'100	CHF 43'000 à moins de CHF 49'500
O 07	CHF 4'100 à moins de CHF 4'400	CHF 49'500 à moins de CHF 52'500
O 08	CHF 4'400 à moins de CHF 5'100	CHF 52'500 à moins de CHF 61'500
O 09	CHF 5'100 à moins de CHF 6'200	CHF 61'500 à moins de CHF 75'000
O 10	CHF 6'200 à moins de CHF 7'300	CHF 75'000 à moins de CHF 88'000
O 11	CHF 7'300 à moins de CHF 8'700	CHF 88'000 à moins de CHF 105'000
O 12	CHF 8'700 à moins de CHF 9'400	CHF 105'000 à moins de CHF 113'000
O 13	CHF 9'400 à moins de CHF 10'200	CHF 113'000 à moins de CHF 122'500
O 14	CHF 10'200 à moins de CHF 12'100	CHF 122'500 à moins de CHF 145'000
O 15	CHF 12'100 à moins de CHF 15'400	CHF 145'000 à moins de CHF 184'500
O 16	CHF 15'400 ou plus	CHF 184'500 ou plus
O 88	(Ne sait pas)	
O 77	(Préfère ne pas répondre)	

43. En cumulant toutes les sources de revenu, indiquez s'il vous plaît le REVENU NET TOTAL DE VOTRE MENAGE (en gros). Si vous ne connaissez pas le chiffre exact, veuillez donner une approximation

	Approximation MENSUELLE	Approximation ANNUELLE
O 01	Moins de CHF 1'100	Moins de CHF 13'000
O 02	CHF 1'100 à moins de CHF 1'800	CHF 13'000 à moins de CHF 21'500
O 03	CHF 1'800 à moins de CHF 2'700	CHF 21'500 à moins de CHF 32'000
O 04	CHF 2'700 à moins de CHF 2'900	CHF 32'000 à moins de CHF 34'500
O 05	CHF 2'900 à moins de CHF 3'600	CHF 34'500 à moins de CHF 43'000
O 06	CHF 3'600 à moins de CHF 4'100	CHF 43'000 à moins de CHF 49'500
O 07	CHF 4'100 à moins de CHF 4'400	CHF 49'500 à moins de CHF 52'500
O 08	CHF 4'400 à moins de CHF 5'100	CHF 52'500 à moins de CHF 61'500
O 09	CHF 5'100 à moins de CHF 6'200	CHF 61'500 à moins de CHF 75'000
O 10	CHF 6'200 à moins de CHF 7'300	CHF 75'000 à moins de CHF 88'000
O 11	CHF 7'300 à moins de CHF 8'700	CHF 88'000 à moins de CHF 105'000
O 12	CHF 8'700 à moins de CHF 9'400	CHF 105'000 à moins de CHF 113'000
O 13	CHF 9'400 à moins de CHF 10'200	CHF 113'000 à moins de CHF 122'500
O 14	CHF 10'200 à moins de CHF 12'100	CHF 122'500 à moins de CHF 145'000
O 15	CHF 12'100 à moins de CHF 15'400	CHF 145'000 à moins de CHF 184'500
O 16	CHF 15'400 ou plus	CHF 184'500 ou plus
O 88	(Ne sait pas)	
O 77	(Préfère ne pas répondre)	

44. Laquelle de ces descriptions correspond au mieux à ce que vous pensez du revenu actuel de votre ménage ?

- O 1 on peut vivre confortablement du revenu actuel
- O 2 le revenu actuel suffit
- O 3 il est difficile de vivre avec le revenu actuel
- O 4 il est très difficile de vivre avec le revenu actuel

45. Avez-vous un abonnement de téléphone "fixe" ?

- O 1 oui **Aller à la Q. 46**
- O 2 non **Aller à la Q. 47**

46. Ce numéro est-il inscrit dans l'annuaire téléphonique ?

- O 1 oui
- O 2 non

- O 8 (ne sait pas)

47. Avez-vous un téléphone mobile (natel) ?

- O 1 oui **Aller à la Q. 48**
- O 2 non **Aller à la Q. 49**

48. Ce numéro de mobile est-il inscrit dans l'annuaire téléphonique ?

- O 1 oui
- O 2 non

- O 8 (ne sait pas)

49. Combien de fois utilisez-vous Internet, le World Wide Web ou une messagerie email pour votre usage personnel, que ce soit à la maison ou au travail ?

- O 1 pas d'accès à la maison ou au bureau
- O 2 a un accès mais ne l'utilise jamais
- O 3 moins d'une fois par mois
- O 4 une fois par mois
- O 5 plusieurs fois par mois
- O 6 une fois par semaine
- O 7 plusieurs fois par semaine
- O 8 chaque jour

Pour finir, quelques questions sur cette enquête et les enquêtes en général.

50. Dans quelle mesure diriez-vous que les enquêtes comme celle-ci constituent une intrusion dans la vie privée des gens ?

Merci de choisir un chiffre entre 0 et 10, où 0 signifie que ces enquêtes ne constituent pas du tout une intrusion dans la vie privée et 10 signifie que ces enquêtes constituent une intrusion totale dans la vie des gens. Les chiffres intermédiaires vous permettent de nuancer votre jugement.

Pas du tout une intrusion dans la vie privée des gens										Une intrusion totale dans la vie privée des gens
0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

51. A quel point faites-vous confiance aux résultats obtenus par des enquêtes comme celle-ci ?

Merci de choisir un chiffre entre 0 et 10, où 0 signifie que vous ne faites aucune confiance et 10 signifie que vous faites une confiance totale aux enquêtes comme celle-ci. Les chiffres intermédiaires vous permettent de nuancer votre jugement.

Aucune confiance										Confiance totale
0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

52. A quel point trouvez-vous intéressant répondre à des enquêtes comme celle-ci ?

Merci de choisir un chiffre entre 0 et 10, où 0 signifie pas intéressant du tout et 10 signifie extrêmement intéressant. Les chiffres intermédiaires vous permettent de nuancer votre jugement.

Pas intéressant du tout										Extrêmement intéressant
0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

53. A quel point trouvez-vous utiles les enquêtes comme celle-ci pour réunir des informations sur la société ?

Merci de choisir un chiffre entre 0 et 10, où 0 signifie pas utiles du tout et 10 signifie extrêmement utiles. Les chiffres intermédiaires vous permettent de nuancer votre jugement.

Pas utiles du tout										Extrêmement utiles
0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

54. Si l'on vous demandait de participer à une enquête qui durerait environ une demi-heure, comment préféreriez-vous répondre aux questions ?

Une seule réponse

- O 1 un entretien à votre domicile
- O 2 un entretien par téléphone
- O 3 en complétant un questionnaire sur papier
- O 4 en complétant un questionnaire sur internet
- O 5 une autre manière

Si 'Une autre manière, veuillez indiquer quelle autre manière vous préféreriez.

____

____

55. Parfois les questions ont des effets différents sur les gens. Concernant les questions de cette enquête, que diriez-vous ? Pensez-vous que des personnes que vous connaissez seraient amenées à donner des réponses fausses ou exagérées aux questions suivantes ou pas ?

La question sur 'combien de gens d'une origine ethnique différente de la plupart des Suisses il faut autoriser à venir vivre ici'.

- O 1 oui, cette question amènerait des personnes que je connais à donner des réponses fausses ou exagérées
- O 2 non, cette question n'amènerait personne que je connais à donner des réponses fausses ou exagérées

56. La question sur 'combien ils sont heureux ou malheureux'.

- O 1 oui, cette question amènerait des personnes que je connais à donner des réponses fausses ou exagérées
- O 2 non, cette question n'amènerait personne que je connais à donner des réponses fausses ou exagérées

57. La question sur 'le nombre de jours où ils ont bu de l'alcool sur les 7 derniers jours'

- O 1 oui, cette question amènerait des personnes que je connais à donner des réponses fausses ou exagérées
- O 2 non, cette question n'amènerait personne que je connais à donner des réponses fausses ou exagérées

58. Maintenant vous avez terminé le questionnaire. Nous aimerions juste vous poser une petite question sur la longueur de ce questionnaire. Auriez-vous été prêt/e à continuer …

- O 1 beaucoup plus longtemps
- O 2 un peu plus longtemps
- O 3 plus du tout

59. Est-ce que vous avez d'autres commentaires par rapport aux thèmes que nous avons touchés pendant l'entretien ?

Vous pouvez inscrire tous vos commentaires ici :

Merci d'avoir répondu à ces questions.

ANNEXE B

Marginal effect for interaction effects between negative life events and age, and negative life events and having a low income.

Happiness category*low income* negative events	Margin	Std. Err.	z	P>z	[95% Conf. Interval]	
1 0 0	0.0043789	0.0027901	1.57	0.117	-0.0010895	0.0098474
1 0 1	0.0112577	0.0070581	1.59	0.111	-0.002576	0.0250914
1 0 2	0.0312631	0.0258935	1.21	0.227	-0.0194873	0.0820134
1 0 3	0.0925016	0.0791623	1.17	0.243	-0.0626537	0.2476569
1 1 0	0.0232363	0.0172089	1.35	0.177	-0.0104925	0.0569652
1 1 1	0.0279815	0.0205244	1.36	0.173	-0.0122455	0.0682085
1 1 2	0.0164054	0.0180793	0.91	0.364	-0.0190293	0.0518401
1 1 3	0.0683007	0.087641	0.78	0.436	-0.1034724	0.2400739
2 0 0	0.004939	0.0030812	1.6	0.109	-0.0011001	0.010978
2 0 1	0.0125113	0.007658	1.63	0.102	-0.002498	0.0275206
2 0 2	0.0315756	0.0236542	1.33	0.182	-0.0147857	0.0779369
2 0 3	0.0850534	0.0662632	1.28	0.199	-0.0448201	0.2149268
2 1 0	0.0251434	0.0180936	1.39	0.165	-0.0103194	0.0606063
2 1 1	0.0299874	0.0211151	1.42	0.156	-0.0113974	0.0713721
2 1 2	0.017633	0.0185124	0.95	0.341	-0.0186507	0.0539167
2 1 3	0.0656386	0.0721652	0.91	0.363	-0.0758026	0.2070797
3 0 0	0.0127737	0.0055444	2.3	0.021	0.0019069	0.0236406
3 0 1	0.0314748	0.0130473	2.41	0.016	0.0059025	0.057047
3 0 2	0.0687904	0.0334556	2.06	0.04	0.0032187	0.1343622
3 0 3	0.1591699	0.0822687	1.93	0.053	-0.0020739	0.3204137
3 1 0	0.0602872	0.0323676	1.86	0.063	-0.0031522	0.1237266
3 1 1	0.0706939	0.036587	1.93	0.053	-0.0010153	0.142403
3 1 2	0.0418801	0.0364672	1.15	0.251	-0.0295942	0.1133544
3 1 3	0.1310631	0.098735	1.33	0.184	-0.0624539	0.3245801
4 0 0	0.0402968	0.011939	3.38	0.001	0.0168969	0.0636968
4 0 1	0.091654	0.0240522	3.81	0	0.0445125	0.1387955
4 0 2	0.1528295	0.0429477	3.56	0	0.0686536	0.2370055
4 0 3	0.2518674	0.0601501	4.19	0	0.1339754	0.3697593
4 1 0	0.1547616	0.0554649	2.79	0.005	0.0460524	0.2634707
4 1 1	0.1739714	0.0586109	2.97	0.003	0.0590961	0.2888468
4 1 2	0.1065597	0.06885	1.55	0.122	-0.0283839	0.2415032
4 1 3	0.23448	0.0764064	3.07	0.002	0.0847261	0.3842338
5 0 0	0.0254904	0.0090291	2.82	0.005	0.0077936	0.0431872

5 0 1	0.052654	0.0172292	3.06	0.002	0.0188854	0.0864227
5 0 2	0.0708699	0.024813	2.86	0.004	0.0222374	0.1195024
5 0 3	0.0819165	0.0325326	2.52	0.012	0.0181538	0.1456793
5 1 0	0.0772957	0.027573	2.8	0.005	0.0232536	0.1313378
5 1 1	0.0830224	0.0283062	2.93	0.003	0.0275432	0.1385016
5 1 2	0.0539469	0.0304857	1.77	0.077	-0.005804	0.1136978
5 1 3	0.0854308	0.0329741	2.59	0.01	0.0208028	0.1500589
6 0 0	0.1001173	0.0213482	4.69	0	0.0582757	0.141959
6 0 1	0.1750734	0.030654	5.71	0	0.1149927	0.2351541
6 0 2	0.1927817	0.0385937	5	0	0.1171395	0.268424
6 0 3	0.154265	0.0666224	2.32	0.021	0.0236876	0.2848424
6 1 0	0.2119846	0.0361011	5.87	0	0.1412279	0.2827414
6 1 1	0.215247	0.0348947	6.17	0	0.1468547	0.2836392
6 1 2	0.1594308	0.0664696	2.4	0.016	0.0291528	0.2897088
6 1 3	0.1774301	0.0790954	2.24	0.025	0.0224058	0.3324543
7 0 0	0.3683001	0.0367537	10.02	0	0.2962642	0.440336
7 0 1	0.3902902	0.0368859	10.58	0	0.3179952	0.4625853
7 0 2	0.3106166	0.0540529	5.75	0	0.204675	0.4165583
7 0 3	0.1362285	0.0867721	1.57	0.116	-0.0338417	0.3062987
7 1 0	0.3169346	0.0706495	4.49	0	0.1784641	0.455405
7 1 1	0.2900977	0.0743081	3.9	0	0.1444566	0.4357389
7 1 2	0.3445751	0.0581255	5.93	0	0.2306513	0.4584989
7 1 3	0.1805743	0.1432395	1.26	0.207	-0.10017	0.4613186
8 0 0	0.3238973	0.04354	7.44	0	0.2385605	0.4092342
8 0 1	0.1855123	0.0348818	5.32	0	0.1171453	0.2538793
8 0 2	0.113631	0.0414665	2.74	0.006	0.0323583	0.1949038
8 0 3	0.032103	0.0268405	1.2	0.232	-0.0205035	0.0847094
8 1 0	0.1055194	0.0466344	2.26	0.024	0.0141176	0.1969212
8 1 1	0.0886771	0.0407233	2.18	0.029	0.0088608	0.1684934
8 1 2	0.1988157	0.1390505	1.43	0.153	-0.0737183	0.4713497
8 1 3	0.0467888	0.052672	0.89	0.374	-0.0564465	0.150024
9 0 0	0.1198064	0.0308266	3.89	0	0.0593874	0.1802254
9 0 1	0.0495723	0.0159549	3.11	0.002	0.0183012	0.0808433
9 0 2	0.027642	0.0135259	2.04	0.041	0.0011317	0.0541523
9 0 3	0.0068948	0.0063318	1.09	0.276	-0.0055154	0.0193049
9 1 0	0.0248372	0.013811	1.8	0.072	-0.002232	0.0519063
9 1 1	0.0203216	0.0114621	1.77	0.076	-0.0021437	0.042787
9 1 2	0.0607534	0.0698617	0.87	0.385	-0.076173	0.1976797
9 1 3	0.0102937	0.0127913	0.8	0.421	-0.0147769	0.0353643

happiness category*old respondent* negative events	Margin	Std. Err.	z	P>z	[95% Conf.	Interval]
1 0 0	0.007322	0.0046774	1.57	0.117	-0.0018455	0.0164896
1 0 1	0.0141376	0.0088098	1.6	0.109	-0.0031292	0.0314045
1 0 2	0.0155495	0.0111305	1.4	0.162	-0.0062659	0.0373649
1 0 3	0.1043354	0.0886865	1.18	0.239	-0.069487	0.2781578
1 1 0	0.0069563	0.0052199	1.33	0.183	-0.0032745	0.0171871
1 1 1	0.0126271	0.0089453	1.41	0.158	-0.0049054	0.0301596
1 1 2	0.0747185	0.0814459	0.92	0.359	-0.0849125	0.2343496
1 1 3	0.0361822	0.0465833	0.78	0.437	-0.0551194	0.1274839
2 0 0	0.0080909	0.0050455	1.6	0.109	-0.001798	0.0179799
2 0 1	0.0155319	0.0094226	1.65	0.099	-0.0029362	0.0339999
2 0 2	0.0170984	0.0118937	1.44	0.151	-0.0062128	0.0404096
2 0 3	0.0950617	0.0735722	1.29	0.196	-0.0491371	0.2392605
2 1 0	0.007705	0.0056375	1.37	0.172	-0.0033443	0.0187542
2 1 1	0.0139046	0.0095562	1.46	0.146	-0.0048252	0.0326345
2 1 2	0.0714428	0.0685209	1.04	0.297	-0.0628557	0.2057414
2 1 3	0.0381234	0.0458975	0.83	0.406	-0.0518339	0.1280808
3 0 0	0.0201848	0.008611	2.34	0.019	0.0033076	0.0370619
3 0 1	0.0383047	0.0155897	2.46	0.014	0.0077495	0.0688599
3 0 2	0.0421885	0.021707	1.94	0.052	-0.0003564	0.0847335
3 0 3	0.1748183	0.088649	1.97	0.049	0.0010695	0.3485671
3 1 0	0.0192828	0.010803	1.78	0.074	-0.0018906	0.0404562
3 1 1	0.0344274	0.0175115	1.97	0.049	0.0001056	0.0687493
3 1 2	0.1411859	0.0891733	1.58	0.113	-0.0335907	0.3159624
3 1 3	0.0872926	0.0854277	1.02	0.307	-0.0801425	0.2547278
4 0 0	0.0582227	0.0157362	3.7	0	0.0273804	0.0890651
4 0 1	0.1063747	0.0269719	3.94	0	0.0535108	0.1592385
4 0 2	0.116574	0.042347	2.75	0.006	0.0335754	0.1995726
4 0 3	0.2636004	0.0567112	4.65	0	0.1524485	0.3747523
4 1 0	0.0557333	0.0230241	2.42	0.015	0.0106068	0.1008597
4 1 1	0.0965433	0.0348317	2.77	0.006	0.0282744	0.1648122
4 1 2	0.2451006	0.059733	4.1	0	0.1280261	0.3621751
4 1 3	0.2004324	0.1231997	1.63	0.104	-0.0410347	0.4418994
5 0 0	0.0337151	0.0112287	3	0.003	0.0117072	0.055723
5 0 1	0.0583526	0.0188823	3.09	0.002	0.021344	0.0953612
5 0 2	0.0631105	0.0238365	2.65	0.008	0.0163917	0.1098292
5 0 3	0.0805187	0.0347244	2.32	0.02	0.0124601	0.1485773
5 1 0	0.0320976	0.0135276	2.37	0.018	0.005584	0.0586111
5 1 1	0.0535483	0.0203204	2.64	0.008	0.0137211	0.0933754
5 1 2	0.0859378	0.0327738	2.62	0.009	0.0217022	0.1501733
5 1 3	0.0890061	0.034574	2.57	0.01	0.0212424	0.1567699
6 0 0	0.1188796	0.0229898	5.17	0	0.0738205	0.1639388
6 0 1	0.1838263	0.0318334	5.77	0	0.1214339	0.2462186
6 0 2	0.1926585	0.0388633	4.96	0	0.1164878	0.2688292

6 0 3	0.1415489	0.0737115	1.92	0.055	-0.002923	0.2860208
6 1 0	0.1109819	0.0302354	3.67	0	0.0517216	0.1702421
6 1 1	0.1721326	0.0366878	4.69	0	0.1002258	0.2440393
6 1 2	0.1709662	0.076462	2.24	0.025	0.0211035	0.320829
6 1 3	0.2128901	0.0456541	4.66	0	0.1234097	0.3023704
7 0 0	0.3668229	0.0358165	10.24	0	0.2966239	0.4370219
7 0 1	0.374191	0.0368977	10.14	0	0.3018728	0.4465092
7 0 2	0.3610297	0.0515849	7	0	0.2599252	0.4621343
7 0 3	0.1111146	0.0803556	1.38	0.167	-0.0463794	0.2686086
7 1 0	0.3390728	0.0458956	7.39	0	0.249119	0.4290266
7 1 1	0.3781657	0.0406756	9.3	0	0.2984429	0.4578884
7 1 2	0.1620611	0.1082372	1.5	0.134	-0.0500799	0.3742021
7 1 3	0.251075	0.179516	1.4	0.162	-0.1007699	0.6029198
8 0 0	0.2866137	0.0405291	7.07	0	0.2071781	0.3660493
8 0 1	0.1660085	0.0339053	4.9	0	0.0995552	0.2324617
8 0 2	0.152061	0.0545546	2.79	0.005	0.045136	0.2589859
8 0 3	0.0239708	0.0204305	1.17	0.241	-0.0160723	0.0640139
8 1 0	0.304921	0.0592461	5.15	0	0.1888008	0.4210413
8 1 1	0.1871872	0.0549417	3.41	0.001	0.0795034	0.294871
8 1 2	0.0399192	0.0337793	1.18	0.237	-0.026287	0.1061255
8 1 3	0.0695091	0.0770774	0.9	0.367	-0.0815598	0.2205781
9 0 0	0.1001482	0.0269319	3.72	0	0.0473628	0.1529337
9 0 1	0.0432728	0.0145687	2.97	0.003	0.0147187	0.0718268
9 0 2	0.0397299	0.0217394	1.83	0.068	-0.0028785	0.0823383
9 0 3	0.0050312	0.0045622	1.1	0.27	-0.0039104	0.0139729
9 1 0	0.1232494	0.0506585	2.43	0.015	0.0239605	0.2225382
9 1 1	0.0514638	0.0230641	2.23	0.026	0.006259	0.0966687
9 1 2	0.0086678	0.0079944	1.08	0.278	-0.0070009	0.0243366
9 1 3	0.015489	0.0190668	0.81	0.417	-0.0218813	0.0528593