The Happiness Analyzer

A New Technique for Measuring Subjective Well-Being

Kai Ludwigs

ISBN: 978-94-6361-075-9

Layout and printed by: Optima Grafische Communicatie, Rotterdam, The Netherlands (www.ogc.nl)

The Happiness Analyzer

A New Technique for Measuring Subjective Well-Being

The Happiness Analyzer

Een nieuwe techniek voor het meten van subjectief welbevinden

Thesis

to obtain the degree of Doctor from the Erasmus University Rotterdam by command of the rector magnificus
Prof.dr. H.A.P. Pols
and in accordance with the decision of the Doctorate Board. The public defence shall be held on
Friday the 23rd of March 2018 at 9.30 hrs
by
Kai Ludwigs
born in Neuss, Germany

Erasmus University Rotterdam

DOCTORAL COMMITTEE

Promoter:	Prof.dr. L.R. Arends
	Prof.dr. H.R. Commandeur
Other members:	Dr. F.M. van der Veen
	Prof.dr. M. Luhmann
	Prof.dr. R. Schöb
Copromotor:	Dr. M.J. Burger

Table of Contents

Glossary

Glossary of well-being concepts

Quality of Life
The term quality of life usually defines the weighted average of factors that have an impact on society (OECD Better Life Index; Durand, 2015). For example, the OECD sees quality of life as the high-level construct that needs to be improved to increase objective and subjective well-being in the world.

Well-Being
The term well-being includes both subjective and objective indicators that describe the recent state of a human or a society (Durand, 2015).

Objective well-being
Objective well-being indicators are, for example, health indicators such as sleep quality or chronic conditions. Objective well-being is not the main focus of this dissertation.

Subjective well-being
The term subjective well-being describes the subjective feeling to feel good in the sense of having many positive emotions (affective balance), a high satisfaction with life and different aspects of life (life evaluation) and the feeling that life makes sense (Eudaemonia) (OECD, 2013). The term is often used synonymously with the term life satisfaction or happiness (OECD, 2013). Subjective well-being is the main focus of this dissertation.

Glossary of subjective well-being measurement methods

Global self-report

Subjective well-being is most of the time measured by a global self-report method (World Database of Happiness, 2017). A participant has to report how high his or her subjective well-being is overall. On the one hand, the global self-report does not require many resources from either the researcher or the participant and gives a first indicator to monitor differences in people's subjective well-being. On the other hand, it does not include enough information to understand differences in detail.

Day Reconstruction Method (DRM)

The Day Reconstruction Method (DRM; Kahneman et al., 2004) was developed by Nobel-prize laureate Daniel Kahneman and colleagues to collect better data on people's subjective well-being by collecting data on their everyday life and everyday feelings. The DRM asks participants to reconstruct their previous day in episodes reporting their activities, social environment and location and then asks them to reconstruct how they felt during these episodes. This procedure takes up to 45 minutes if it is done by paper and pencil and requires high resources from the researcher as well to transcribe and analyze the collected data.

Experience Sampling Method (ESM)

The Experience Sampling Method (ESM; Csikzentmihalyi & Hunter, 2003) is a method to collect more affective rather than cognitive data on people's subjective well-being. Participants usually carry a beeper and a questionnaire with them and if the beeper vibrates they are asked to report how they feel right now, what they do, with whom and where. This procedure does not take as long as the DRM but also only helps to collect some momentary assessments of people's subjective well-being rather than having the full context of people's time-use.

OECD Gold Standard on measuring subjective well-being

The OECD (2013) published guidelines that represent the gold standard for measuring subjective well-being.

According to the OECD, subjective well-being is a construct consisting of three elements: i) life evaluation – a reflective, cognitive judgement of a person's life or specific parts of it; ii) affect – a person's positive and negative emotions and feelings; iii) eudaimonia – according to Aristotle's 2000-year-old construct, a person's judgement of his life in terms of meaning and purpose in life (for more details on these definitions, see OECD 2013).

To measure these different elements, the OECD suggests using six different modules: i) a core module about happiness and life-satisfaction with a single question; ii) an affect module with multiple specific questions; iii) a life evaluation module with multiple specific questions; iv) an eudaimonic well-being module with multiple specific questions; v) a domain evaluation module with multiple specific questions about satisfaction in specific life domains (e.g., health); and vi) an experienced well-being module for which the guidelines recommend using the ESM and/or DRM. In addition to these modules, the OECD generally advocates for more longitudinal studies instead of cross-sectional studies and for linking SWB data to objective data, including location data, economic variables or biological markers such as heart rate variability, face emotion recognition and others.

The requirements can be ordered according to an onion model, as displayed in Figure 1, consisting of three layers within a frame. i) General Measurement: A general SWB questionnaire including OECD modules one to five that is designed to obtain a cognitive measurement. ii) Activity-based Measurement: Daily life and daily affective experience measurement employing affective time-use diaries via techniques such as the DRM to comprehensively capture people's time use to obtain more contextual information. iii) Experience-based Measurement: Affective measurement in the moment using, for

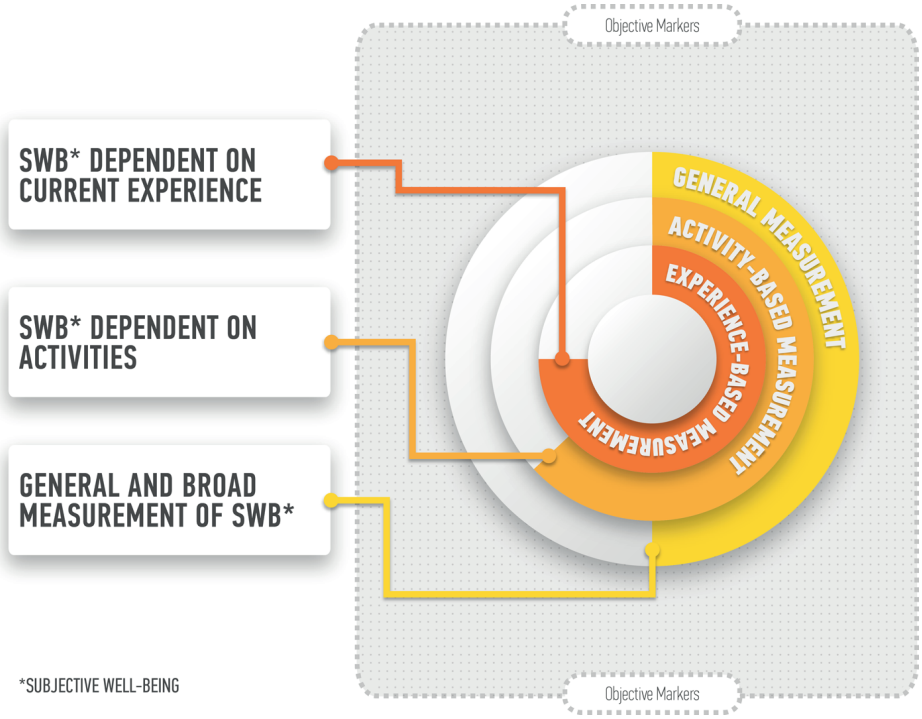

Figure 1: The Onion Model

example, the ESM. iv) Objective Markers: Around the subjective layers is an objective frame integrating other objective markers to increase validity, for example, location data, economic variables, and biological markers such as heart rate variability or face emotion recognition as noted above. Unlike the OECD, we separate the DRM and ESM into two different layers. We agree that both methods primarily measure experienced well-being, but the DRM provides far more contextual information about a person's life and activities than the ESM because the DRM collects information about the entire 24 hours in a day, rather than simply a few moments. Thus, the DRM helps in obtaining a more detailed understanding of the underlying mechanisms of SWB.

Unfortunately, most existing studies do not measure subjective well-being according to this standard because considerable resources are needed (from both researchers and participants) to capture the following information: i) people's SWB at multiple points in time using general questionnaires; ii) people's everyday life and everyday life feelings; iii) people's direct feelings in the moment; and iv) a combination of subjective and objective well-being measurements such as people's subjective ratings of SWB and their objective stress level indicated by, for example, heart rate variability.

Table 1 displays a set of questions for the subjective well-being layers that are in line with the OECD guidelines.

Table 1: Measures of subjective well-being according the OECD gold standard

Measure	Items	Scale Range
Happiness Core (HC)	Taking all things together, how happy would you say you are?	0: Extremely unhappy 10: Extremely happy
Life Satisfaction Core (LC)	All things considered, how satisfied are you with your life as a whole nowadays?	0: Extremely dissatisfied 10: Extremely satisfied
Scale of Positive and Negative Experience (SPANE)	How often did the interviewed person experience the following emotions in the last two weeks: 1: Negative 2: Unpleasant 3: Good 4: Bad 5: Happy 6: Afraid 7: Pleasant 8: Contended 9: Sad 10: Angry 11: Joyful 12: Positive	0: never 7: always
Satisfaction With Life Scale (SWLS)	Indicate your agreement which each item: 1: In most ways, my life is close to my ideal 2: The conditions of my life are excellent 3: I am satisfied with my life 4: So far, I have gotten the important things I want in life 5: If I could live my life over, I would change almost nothing.	1: Strongly disagree 7: Strongly agree

Table 1 (*continued*)

Measure	Items	Scale Range
Flourishing Scale (FS)	Indicate your agreement with each item: 1: I lead a purposeful and meaningful life 2: My social relationships are supportive and rewarding 3: I am engaged and interested in my daily activities 4: I actively contribute to the happiness and well-being of others 5: I am competent and capable in the activities that are important to me 6: I am a good person and live a good life 7: I am optimistic about my future 8: People respect me	1: Strongly disagree 7: Strongly agree
Domain Evaluation Questionnaire (DEQ)	The following questions ask you how satisfied you feel about specific aspects in your life: 1: Standard of Living 2: Health 3: Productivity 4: Personal relationships 5: Safety 6: Community 7: Personal Security 8: Free time 9: Environment 10: Job	0: Not at all satisfied 10: Completely satisfied
Day Reconstruction Method (DRM)	What did you do in this period? Where have you been in this period? Who was with you in this period? How did you feel during this episode?	0: Unhappy 10: Happy
Experience Sampling Method (ESM)	How do you feel right now? What are you doing right now? Where are you right now? Who is with you right now?	0: Unhappy 10: Happy

Glossary of other concepts

DIKW Model

One of the "taken-for-granted" models in information science is the "Data-Information-Knowledge-Wisdom" (DIKW) model (Ackoff, 1989). This model describes the differences and hierarchy of the constructs data, information, knowledge and wisdom. We will outline the different constructs in the following.

Data

Data are defined as symbols that represent properties of objects, events and their environment. They are the products of observation. But are of no use until they are in a useable (i.e. relevant) form. The difference between data and information is functional, not structural (Ackoff, 1989). Zeleny (1987) summarizes that data are associated to "know nothing".

Information

Information is contained in descriptions, answers to questions that begin with such words as who, what, when and how many. Information systems generate, store, retrieve and process data. Information is inferred from data (Ackoff, 1989). Zeleny (1987) summarizes that information is associated to "know what".

Knowledge

Knowledge is know-how, and is what makes possible the transformation of information into instructions. Knowledge can be obtained either by transmission from another one who has it, by instruction, or by extracting it from experience (Ackoff, 1989). Zeleny (1987) summarizes that knowledge is associated to "know how".

Wisdom

Wisdom is the ability to increase effectiveness. Wisdom adds value, which requires the mental function that we call judgement. The ethical and aesthetic values that this implies are inherent to the actor and are unique and personal (Ackoff, 1989). Zeleny (1987) summarizes that wisdom is associated to "know why".

More data

More data by itself can just give more chances to find information, knowledge and wisdom.

Better data

Better data means that these data have a higher chance of finding information, knowledge and wisdom because they are collected to match this purpose by adding to existing data sources.

1 Introduction

BACKGROUND

In the past few decades, a growing number of people have been surveyed about their subjective well-being with general retrospective questions such as "All things considered, how satisfied are you with your life as a whole nowadays?" (ESS; 2013). This is generally a positive development; however, when analyzing data from these surveys, researchers and decision makers have realized that better data are needed to understand what differentiates people with a higher subjective well-being from people with a lower subjective well-being. Better data can be achieved by adding different data sources rather than only retrospective and general questions about subjective well-being such as the one mentioned above. A method to collect better data by a multi-method approach is explained by the term "triangulation" (Denzin, 1978; Jick, 1979). Triangulation is defined by Denzin (1978, p. 291) as "the combination of methodologies in the study of the same phenomenon". If different methods are used to study subjective well-being it can help to not only collect more data, meaning more general subjective well-being ratings from more people, it can help to collect better data, meaning different data sources to evaluate people's subjective well-being in order to find information on what defines subjective well-being, maybe knowledge about how subjective well-being is defined and possibly wisdom on why subjective well-being is higher or lower in different groups to maybe find or choose the right interventions to improve subjective well-being. The three examples below should help to outline this statement.

For example, we know that most migrants report lower subjective well-being than locals in general surveys, but we do not know if they feel worse in every life situation

or if there are specific moments in their everyday life when they feel worse than locals (Hendriks, Ludwigs & Veenhoven, 2016). Thus, it is difficult to have information on what defines migrants' subjective well-being and to know how to decide which interventions should be funded to improve migrants' subjective well-being and difficult to evaluate the effects of those interventions.

Another example is unemployed people who often do not report a lower affective balance than employed people but often do report a lower life satisfaction and thus lower subjective well-being (Knabe et al., 2010). Collecting better data on unemployed and employed people's everyday life and everyday feelings shows that unemployed people spend more time with free-time activities such as doing sports and thus have often a good affective balance in general but when comparing the subjective well-being while doing sports to employed people shows that they report a lower subjective well-being (Knabe et al., 2010). In this example, better data on subjective well-being, gives the information to prevent policy makers from thinking that it is not urgent to fund interventions to improve unemployed people's subjective well-being such as job-trainings or mental coaching to help them prepare for their challenge to find a new job.

A more individual example is a psychotherapist who surveys a client about his or her general subjective well-being but does not know in which situations the client has a below average subjective well-being and in which situations he or she is on an average or above average level. By having better data on everyday life and everyday feelings a therapist could have better information on what defines the client's subjective well-being, maybe suggest better fitting interventions to the client and evaluate those interventions in more detail to learn for the next decision, which would result in a higher probability for fitting interventions and more knowledge on how subjective well-being is defined for the client with every cycle.

In conclusion, having better data on people's subjective well-being can help us to better understand what defines subjective well-being, how subjective well-being is defined and maybe why subjective well-being is higher or lower in different groups to maybe increase the chances to find more solutions to improve subjective well-being.

PROBLEM STATEMENT AND RESEARCH GOAL

In 2013 the OECD published a guideline that represents the gold standard for measuring subjective well-being in greater detail to collect data in the quality needed as a basis for more information and knowledge about people's subjective well-being to increase the chances for efficient decisions to improve subjective well-being and enrich the evaluation of those decisions to enable continuous learning. Unfortunately, most existing studies nevertheless do not measure subjective well-being according to this standard,

as traditional methods (e.g., paper and pencil or personal interviews) require considerable resources (from both researchers and participants) to capture i) people's subjective well-being at multiple points in time using general questionnaires; ii) people's everyday life and everyday feelings; iii) people's specific feelings in the moment; and iv) a combination of subjective and objective well-being measurements. To resolve this issue, we developed an app as a mobile assessment tool, the "Happiness Analyzer". The main goal of this dissertation is to explain and evaluate this new tool.

STRUCTURE OF THE DISSERTATION

Figure 1.1 gives an overview on the structure of the dissertation and table 1.1 gives an overview of the different chapters.

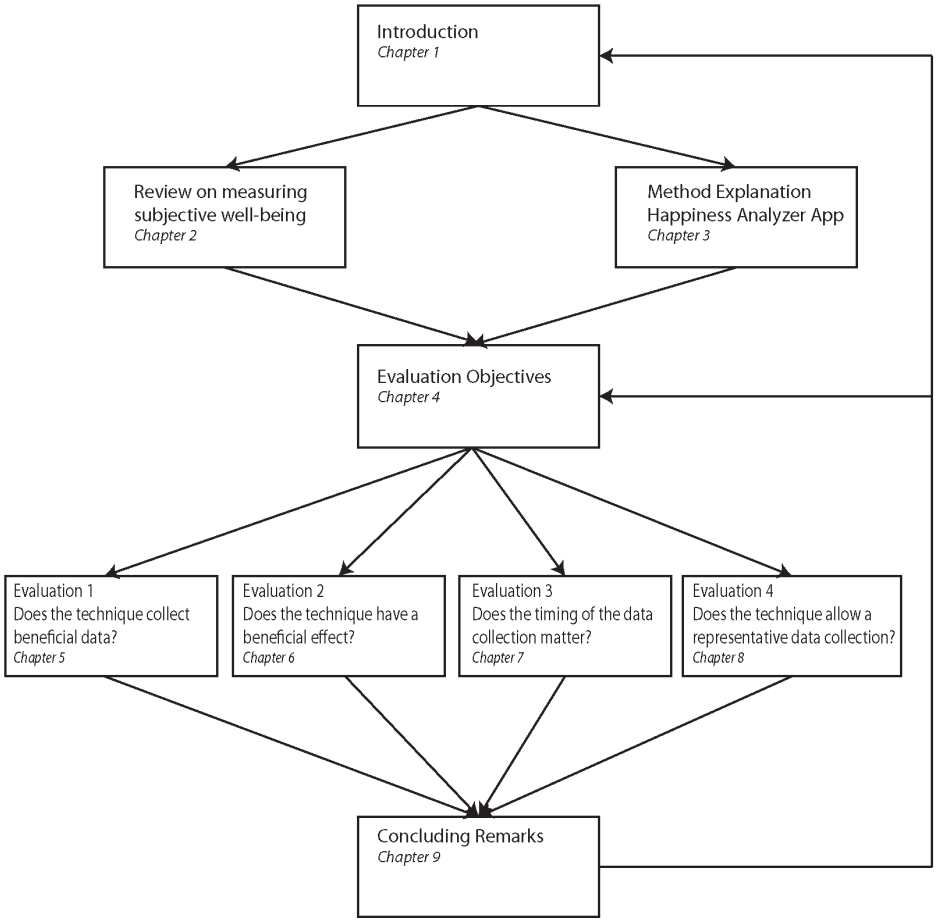

Figure 1.1: Overview of the structure of the dissertation

Table 1.1: In-depth overview of the individual dissertation chapters

Chapter	Title	Central question	Sub-questions	Methodology	Data	Conclusions	Co-Authors	Status & Outlet
1	Introduction							
2	Measuring Happiness: A Practical Review	- What are the benefits and problems of different subjective well-being assessment methods?	- Which subjective well-being measures are commonly used - Which subjective well-being measures should be used for which purpose?	- Literature Review - Analysis of the World Database of Happiness	- World Database of Happiness	- Global Self-Report surveys can be enough for certain occasions but in general more detailed assessments are very beneficial but rarely used.	- Lena Henning - Lidia Arends	Accepted as a book chapter in: *Perspectives on Community Well-Being* (2018)
3	The Happiness Analyzer – A New Technique For Measuring Subjective Well-Being	- Does the Happiness Analyzer help to fulfil the OECD gold standard on measuring subjective well-being with less resources?	- Are there any other tools that have the same features?	- Literature Review - App Development - Evaluation Studies	- Evaluation study with German students ($N = 112$) - Different Data collections from other researchers using the Happiness Analyzer	- The Happiness Analyzer helps to fulfil the OECD gold standard on measuring subjective well-being with less ressources.	- Stephan Erdtmann	Submitted to: *Journal of Happiness Studies*
4	Evaluation Goals							
5	Why are locals happier than internal migrants? The role of daily life.	- Does the technique collect beneficial data?	- Do domestic migrants and locals differ in their everyday life and everyday life feelings?	- Micro-level analysis - MANCOVA - OLS regression analyses	- Self-Collected data on young German adults ($N = 150$)	- The technique collects beneficial data to explain subjective well-being in more detail.	- Martijn Hendriks - Ruut Veenhoven	Published in: *Social Indicators Research* (2016)

Table 1.1: In-depth overview of the individual dissertation chapters (*continued*)

Chapter	Title	Central question	Sub-questions	Methodology	Data	Conclusions	Co-Authors	Status & Outlet
6	How Does More Attention to Subjective Well-Being Affect Subjective Well-Being?	- Does the technique have a beneficial effect?	- Does the attention to subjective well-being have a positive effect or negative effect on people's subjective well-being?	- ANOVA and MANOVA	- Self-Collected data on young German adults (N = 129 & N = 120)	- The technique seems to have a beneficial effect on the user. - It seems to be beneficial to pay attention to individual subjective well-being.	- Richard Lucas - Ruut Veenhoven - Martijn Burger - Lidia Arends	Accepted in: *Applied Research in Quality of Life*
7	Using the Day Reconstruction Method - Same results when used at the end of the day or on the next day?	- Does the timing of the data collection matter?	- Are people happier in the evening of the same day compared to the morning of the next day?	- Multi-Level Analysis	- Data collected in cooperation with the German Socio Economic Panel (N = 374)	- The timing of the data collection does matter. - People seem to be happier on the evening of the same day compared to the morning of the next day.	- Lena Henning - Lidia Arends	Submitted to: *Journal of Well-Being Assessment*
8	Can happiness apps generate representative datasets? - A case study collecting data on people's happiness using the German Socio Economic Panel	- Does the technique allow a representativdata collection?	- Can apps generate representative datasets?	- Chi-Square Analysis	- Data collected in cooperation with the German Socio Economic Panel (N = 2135 & N = 1869)	- The technique allows a representative data collection with quota sampling and a high reward.	- Richard Lucas - Ruut Veenhoven - David Richter - Lidia Arends	*Submitted to: Journal of Happiness Studies*
9	Concluding Remarks							

In chapter two, we will introduce and evaluate different subjective well-being assessment methods in a review. Because research on subjective well-being remains a young discipline with many different definitions for the term and many different measurement techniques, this chapter aims to review the different definitions and measurements. With the help of the World Database of Happiness we review which measurements are used, how frequently they are used and how high their psychometric quality is given the published research thus far. In the end, the chapter presents a guideline for measuring subjective well-being, which goes into more detail than the OECD guidelines mentioned above, as they mainly describe which modules should be used without giving detailed recommendations on different psychometric qualities.

In the third chapter, we will describe the Happiness Analyzer method in detail. We will discuss why it is difficult to match the OECD gold standard without the use of modern technologies. After reviewing existing online tools that try to capture people's subjective well-being in more detail, we outline what the "Happiness Analyzer" adds to the field. In the rest of the section, we explain the different functionalities. At the end of the section, we review the evaluation feedback that we received in the most recent evaluation study and conclude that the tool works and has all the main features to help researchers collect data on people's subjective well-being according to the OECD gold standard.

As an introduction to chapters 5, 6, 7 and 8 we will summarize in chapter 4: i) this introduction (chapter 1); ii) the review article (chapter 2) and iii) the method paper (chapter 3) and then outline the evaluation objectives, which we try to achieve with four example studies (chapters 5, 6, 7 and 8).

The first example (chapter 5) shows the benefit of the data collected with the Happiness Analyzer by analyzing data on migrants' subjective well-being compared to locals' subjective well-being. As research shows that migrants report lower subjective well-being than locals, it seemed interesting to elucidate the reasons for this gap in more detail. Based on a survey of young adults using the Happiness Analyzer, we show that migrants spend less time with happiness-producing activities, such as active leisure, social drinking/parties, and activities outside home/work/transit, and that they report lower subjective well-being than locals when spending time with friends and eating. Additionally, we show that it is useful to capture subjective well-being according to the OECD gold standard because it can explain more variance in the happiness gap between migrants and locals, which highlights the potential of data collected with the Happiness Analyzer to get a better understanding in which situations certain groups have a lower subjective well-being and to evaluate interventions to improve subjective well-being for these groups more accurately to enable continuous learning.

In the second example (chapter 6), we evaluated whether it is beneficial to use the Happiness Analyzer. To address this question, we ran two longitudinal studies that included two groups: one group merely answered three subjective well-being surveys

over 4 weeks, and the other group followed the same procedure but used the Happiness Analyzer after the first measurement for two weeks to track their subjective well-being in detail. Both studies showed that using the Happiness Analyzer increased participants' subjective well-being by paying more attention to their individual subjective well-being at least in the short-term.

In the third example (chapter 7), we evaluated whether a difference arises if participants reconstruct their everyday life and everyday feelings about the previous day on the evening of the same day or on the next day to evaluate how strict the timing of the data collection has to be. Based on a data collection with the Happiness Analyzer, we show that there is no difference if participants rate their happiness on the evening of the same day or on the next day and can suggest giving participants free choice when they want to fill in the diary to reduce response burden and increase participation rates.

In the fourth example (chapter 8), we evaluated whether nationally representative datasets can be generated by using the Happiness Analyzer. For this purpose, we reviewed the participation rates of two waves of the innovation sample of the German Socio-Economic Panel, during which the Happiness Analyzer was offered to participants for free use after they finished a household interview. In the first wave, participants did not receive any reward for using the Happiness Analyzer, whereas in the second wave, participants received a 50 Euro Amazon voucher if they used the Happiness Analyzer intensely. The participation rates increased from 2% to over 30% in the second wave, indicating that representative datasets can be generated with the Happiness Analyzer if people own a smartphone and are highly motivated. Nevertheless, a nationally representative dataset can only be collected by using a quota sampling approach.

In the last chapter (chapter 9), we conclude and discuss our findings, discuss limitations and suggest topics for further research, make suggestions for decision makers and end with an epilogue.

This dissertation provides a tool to collect better data on people's subjective well-being to help researchers and decision makers to better understand in which situations a certain group has differences in their reported subjective well-being compared to another group. Representative databases with data collected with the Happiness Analyzer might help to better inform decision makers to help them to increase their chances to make efficient decisions to improve subjective well-being and to learn continuously from a better evaluation of these decisions.

2 Measuring Happiness – A Practical Review

ABSTRACT

In times of increasing depression rates, happiness has gained interest as a goal for individuals and society instead of merely increasing gross domestic product. Unfortunately, happiness research remains a young discipline; thus, the definition of the term happiness is unclear across various disciplines, and many different measurement techniques have been developed and used thus far. This book chapter reviews different happiness definitions and ultimately selects the one used by the World Database of Happiness to then review which measurements are used and how frequently and to then evaluate their psychometric quality by reviewing published research thus far. In the end, the chapter presents a practical guideline of what a researcher should be aware of when measuring happiness.

Keywords: Happiness, Measurements, World Database of Happiness, Happiness Psychometrics

INTRODUCTION

Psychological diseases such as burnout and depression are on the rise these days. Accordingly, the World Health Organisation (WHO) forecasts that in 2030, depression will be the most common disease in high-income countries (Allianz & RWI, 2011; Mathers & Loncar, 2006). But this is not a problem that individuals must address on their own; rather, it is also of tremendous relevance for the economy. Indeed, psychological diseases already cause yearly economic costs of at least 7 billion euro, as calculated, for example, for the German population (DGPPN, 2013). The main reasons for high massive expenditures are the direct costs of therapy and indirect costs caused by general productivity loss (Allianz & RWI, 2011).

Beneficial effects of happiness

By comparison, people who live a happy and fulfilling life exhibit various positive characteristics. They are less likely to get sick, and they have a better immune system (Pressman & Cohen, 2005). Moreover, happy people tend to live longer (Danner, Snowdon, & Friesen, 2001; Diener & Chan, 2011), and states with happier citizens have lower suicide rates (Koivumaa-Honkanen, Honkanen, Koskenvuo, & Kaprio, 2003). Additionally, happiness is a crucial factor for job and general satisfaction (Judge & Watanabe, 1993), and in turn, higher job satisfaction predicts lower job turnover rates (Clark, Georgellis, & Sanfey, 1998; Frijters, 2000). Happy people also put more effort into their work and thus work harder (Judge, Thoresen, Bono, & Patton, 2001). Overall, it is thus not surprising that numerous surveys show a positive relationship between people's happiness and their productivity in different contexts (Cropanzano & Wright, 1999; Haas & Janssen, 2012; Harter, Schmidt, Asplund, & Kilham, 2010; Isen, Daubman, & Nowicki, 1987; Lyubomirsky, King & Diener, 2005; Oishi, 2012; Oswald, Proto & Sgroi, 2009; Wright & Cropanzano, 2004; Wright & Staw, 1999). Concerning the economy, happy people are associated with increased health, effort and innovative actions, which ultimately leads to better long-term economic welfare. But the list of benefits that happy people may bring continues. In fact, such people are more sociable (George, 1991), more engaged in prosocial behaviours (Cunningham, Steinberg, & Grev, 1980; Isen, 1970), more likely to volunteer more often (Thoits & Hewitt, 2001), more likely to donate (Priller & Schupp, 2011) and more likely to give more money to charities (Aknin, Sandstrom, Dunn, & Norton, 2011). Therefore, happy people influence not only economic factors positively but also social progress.

Against this background, it is completely rational and understandable that some nations have worked on implementing (e.g., Great Britain, England's Prime Minister David Cameron: Cameron, 2006; Stratton, 2010; White, 2007; France, Former President Nicolas Sarcozy: Jolly, 2009; Stiglitz, Sen, & Fitoussi, 2009) or have clearly announced and prioritized (Bhutan, King Jigme Singye Wangchuck: Pfaff, 2011; Priesner 1999) a more intense

focus on happiness when measuring economic performance and social progress. Accordingly, the European Commission, European Parliament, Club of Rome, OECD, and WWF discussed in 2007 how to improve progress and conditions of societies differently from merely focusing on economic factors (Commission of the European Communities, 2009). Many researchers have also intensely discussed this topic in the scientific community (Diener, 2000, 2012; Diener & Seligman, 2004; Di Tella & MacCulloch, 2008; Dolan & White, 2007; Frey & Stutzer, 2002; Kahneman & Krueger, 2006; Kahneman, Krueger, Schkade, Schwarz & Stone, 2004a; MacKerron, 2012).

It can be concluded that pursuing a happier society, that is, achieving a higher level of happiness for everyone (Veenhoven, 2010), seems to be worthwhile. But the following question remains: what can we do to reach this goal? To answer this question, we need to investigate the important factors and their interrelations that determine happiness. However, to be able to do so, some premises need to be met: (i) We need to know what we mean by happiness. Thus, we need a clear definition of this construct. (ii) We need measures that capture the defined concept of happiness as valid and as feasible as possible. Consequently, we need to investigate existing measures in terms of their (psychometric) quality and their applicability in various situations (e.g., research questions; populations). By doing so, we can determine the best way to assess happiness depending on the current context. The current book chapter aims to contribute to meeting these 2 premises in future studies.

DEFINITION OF HAPPINESS

For a long time, scholars have immensely engaged with the topic of happiness and the pursuit thereof. Ancient philosophy was concerned with the question of what is a good life, which was typically considered a morally good life denoted with the term happiness. For instance, Aristotle described striving for happiness as the most important of all goals and as the goal of life itself, as articulated in the following quotation: "Happiness is the meaning and the purpose of life, the whole aim and end of human existence" (as cited in: Bacon, Brophy, Mguni, Mulgan, & Shandro, 2010, p. 10). Other thinkers of ancient times, such as the Indian intellectual Dhammapada or philosophers from Confucianism, Taoism, and Buddhism (Judge & Kammeyer-Mueller, 2011; Lu, 2001), were also concerned with this question. In the Middle Ages, Thomas Aquinas stated that happiness was "the ultimate goal of the rational being" (Judge & Kammeyer-Mueller, 2011, p. 31) and therewith underlined the importance of striving for happiness. Finally, the American Declaration of Independence (1776) names the pursuit of happiness as one of the unalienable rights besides life and liberty and thus as one of the ultimate rights and goals of every human being. This follows the idea of Jeremy Bentham, who stated in his

doctrine: "Create all the happiness you are able to create; remove all the misery you are able to remove" (as cited in: Layard, 2005, p. 235).

Divergent use of the word

In sum, talking about happiness is not new at all, but the meaning of the word might have changed somewhat. However, research in this field remains very young and has particularly expanded since the 1990s (MacKerron, 2012; OECD, 2013). This is also reflected in the relatively recent launch of the "Journal of Happiness Studies", which has published papers on happiness since 2000 (Journal of Happiness Studies, 2017). As the discipline is so young, final agreement about the relevant terminology and definitions is currently lacking. In 2003, Easterlin posited that for him, happiness could be equated with utility, well-being, life satisfaction, and welfare. Other researchers have added additional terms that have often been used synonymously with happiness, such as "pleasure, life satisfaction, positive emotions, a meaningful life, or a feeling of contentment" (Diener, Scollon, & Lucas, 2003, p. 188). In their paper, Diener, Scollon, & Lucas (2003) use happiness and subjective well-being (SWB) interchangeably, and in accordance such usage, Seligman & Csikszentmihalyi (2000) stated that the term SWB is actually just "a more scientific-sounding term for what people usually mean by happiness" (p. 9; also cp. Diener, 2000, p. 24).

Need for a clear definition

Overall, this inconsistency in terminology can only cause confusion. To be clear in meaning in this book chapter, we will exclusively rely on the term happiness throughout to be consistent with the general tone of the entire book. In addition, we prefer the term happiness because we perceive it to be more easy going and understandable for all readers. Concerning the abovementioned challenge in definition, it must be said that there is no consensus between researchers in their different disciplines for a common definition of happiness (cp. Lu, 2001; Veenhoven, 1984, 2010). Here, a definition for happiness is presented with the aim of (i) integrating the most common definition but also (ii) differentiating the adopted definition from definitions that are relatively vague and probably too broad to capture happiness alone. This definition serves as a basis for the following selection and review of happiness measures. By choosing such a clear concept of happiness, we can assure that the measure selection contains only measures that really fit this definition.

What happiness means from our point of view

In general, research and survey literature has often emphasized two aspects related to happiness (Busseri & Sadava, 2011; Clark & Senik, 2011; Diener, 2000; Diener, Scollon, & Lucas, 2003; Diener, Suh, Lucas, & Smith, 1999; Dolan & White, 2007; Lucas, Diener,

& Suh, 1996; OECD, 2013; Stiglitz et al., 2009): (1) the emotional or affective aspect ("a person's feelings or emotional states, typically measured with reference to a particular point in time": OECD, 2013, p. 10) and (2) the rational, cognitive or evaluative aspect ("a reflective assessment on a person's life or some specific aspect of it": OECD, 2013, p. 10). On the one hand, some happiness definitions especially concentrate on the emotional aspect, as with the one of Bradburn (1969, p. 9), who referred to happiness as the "resultant of the individual's position on two independent dimensions – one of positive affect and the other of negative affect." Another well-known definition is the one of Goldings, who stated in 1954 (p. 31) that happiness for him "embraces feelings of elation, contentment, satisfaction, and pleasure at the positive pole and feelings of depression, discontent, and unpleasure at the negative pole." Further, affect-focused happiness definitions can also be found in Flügel (1925), Fordyce (1977) and Wessman & Ricks (1966). On the other hand, happiness definitions pay particularly attention to the evaluative aspect. Lemon, Bengtson, & Peterson (1972, p. 513), for example, referred to happiness as "the degree to which one is presently content or pleased with his general life situation," whereas Tatarkiewicz (1966, p. 1) merely briefly stated that happiness can be equated with "satisfaction with one's life as a whole." Another, evaluation-focused happiness definition can, for instance, be found in Michalos (1980). In addition to these either affect- or evaluation-focused happiness definitions, some definitions do not have a clear emphasis and combine both aspects instead. One exemplary and well-noticed definition comes from Diener, who wrote in 2000 (p. 34) that happiness for him means "people's cognitive and affective evaluations of their lives" (adapted versions can be found in Diener, 2012; Diener, Scollon, & Lucas, 2003; Diener, Suh, Lucas, & Smith, 1999). Another frequently cited definition of happiness was launched by the OECD (2013, p. 29), which considers happiness to refer to "Good mental states, including all of the various evaluations, positive and negative, that people make of their lives, and the affective reactions of people to their experiences." Further happiness definitions have also been proposed by Busseri & Sadava (2011), Dolan & White (2007) and Sumner (1996).

In sum, all the suggested happiness definitions deal with either feelings or cognitions or combine them both. But none of them assumes the affective and cognitive aspect as components of or views on happiness. In contrast to these previous definitions, our happiness definition does exactly this. Although our approach differs from previous ones in this manner, it is nevertheless generally aligned with the vast majority of literature using an affective and/or cognitive aspect in the definition of happiness (see above for single definitions). Thus, we define *Overall Happiness* as "the overall enjoyment of one's life as-a-whole" (Veenhoven, 2010, p. 611; cp. Veenhoven, 1994, 1997, 1984, 1991, 2008).

This general evaluation is then "based on both affective and cognitive appraisals of life" (Veenhoven, 2010, p. 611, cp. Veenhoven, 1984) or a "dual evaluation system" (Veenhoven, 2000, p. 14). The *Affective Happiness Component* of this system is meant to

evaluate "the degree to which the various affects a person experiences are pleasant; in other words: how well he usually feels" (Veenhoven, 1991, p. 10; cp. Veenhoven, 1984, 2010). The *Cognitive Happiness Component* of this system is then meant to evaluate "the degree to which an individual perceives his aspirations to have been met. In other words: to what extent one perceives oneself to have got what one wants in life" (Veenhoven, 1991, p. 10; cp. Veenhoven, 1984, 2010). Although this dual evaluation system composes the overall happiness evaluation, the latter should be considered separately in surveys. Given this idea, the construct of happiness should in sum be considered "a kind of trinity" (Veenhoven, 1984, p. 28). This approach makes sense when considering the following examples (derived from Veenhoven, 1984, p. 32), in which the calculation of overall happiness using only individuals' affective and cognitive judgments is rather unclear: (i) someone is more or less dissatisfied with what he/she has achieved in life but nevertheless feels tremendously good; (ii) someone obtained everything he/she wanted but nevertheless feels downhearted. Although research results suggest that affective aspects usually influence overall life evaluations more than cognitive ones (Schwarz & Strack, 1991; Veenhoven, 1997, 2000, 2010), we do not know the exact weighting of the factors. Besides this content-related reason, pragmatics play a role when favouring an additional overall happiness evaluation in surveys, as most researchers use overall happiness indicators in their studies (Veenhoven, 1984).

What happiness does not mean from our point of view

To create a clear definition of our happiness construct, it does make sense to define not only what happiness is but also what happiness is not in our understanding. We already fulfilled the first aspect in discussing what we exactly understand by the term happiness. To meet the second aspect, we first collected conditions that are regularly associated with the word happiness today and arranged them in a 2x2 matrix (Veenhoven, 2000, 2008, 2010; see table 1). As table 2.1 shows, happiness in our understanding is something that is judged in "the eye of the beholder" (Veenhoven, 2010, p. 608) and that concerns actual life (not only pre-conditions for a happy life).

Table 2.1: Conditions regularly associated with the word happiness today, classified into a 2x2 matrix.

	Outside the Person	Inside the Person
Possibilities	Liveability of the Environment	Life-Ability
Outcomes	Utility of Life	*Happiness*

Notes: Adapted from Veenhoven (2010, p. 608). Closer explanation of the terms used (p. 608): *Liveability of the Environment* = "good living conditions"; *Life-Ability* = extent to which the person is "equipped to cope with the problems of life"; *Utility of Life* = "a good life must be good for something more than itself", e.g., for "ecological preservation or cultural development"; *Happiness* = as we understand and defined it above.

Similarly, table 2.2 shows the relation of our happiness definition with other kinds of satisfaction that can be expressed by persons. In accordance with this visualization, happiness in our understanding concerns life evaluations that are not momentary and thus fleeting – but rather enduring (Veenhoven, 1997). Additionally, our happiness concept entails an evaluation focusing on overall life, not single life aspects, such as work and marriage (Veenhoven, 1984, 1997). Yet, studies have investigated the contribution of life domains to overall happiness. For example, Van Praag, Frijters, & Ferrer-i-Carbonell (2003) found that finance, health, and job satisfaction influence overall happiness in individuals to the highest extent. However, remarkably, the authors included only 6 life domains in their analysis. Consequently, they may not cover all relevant life domains. Accordingly, Dolan & White (2007) have criticized that how all the different life domains relatively contribute to overall happiness remains unclear today. But even if a researcher considered all important life domains, he still would not be able to calculate a precise overall happiness score because the importance weighting of every life domain for overall happiness has been shown to be highly individual (Diener, Lucas, Oishi, & Suh, 2002; Diener, Scollon, & Lucas, 2003). Thus, it currently remains unclear (i) which domains should be of relevance for overall happiness and (ii) how an overall happiness score can be gained from domain evaluation judgments. Life domains are therefore unsuitable as indicators for overall happiness in a precise happiness definition. Nevertheless, they can deliver valuable insights for researchers who are especially interested in particular life domains. These indicators could then even prove more meaningful than global judgments of happiness in such cases (see also Diener, Scollon, & Lucas, 2003 for this opinion).

Table 2.2: Various kinds of satisfaction that can be expressed by persons, classified into a 2x2 matrix.

	Passing	Enduring
Part of Life	Pleasure	Part Happiness
Life as a Whole	Peak Experience	*Happiness*

Note: Adapted from Veenhoven (2010, p. 609). Closer explanation of the terms used (p. 609): *Pleasure* = "can be sensoric, such as a glass of good wine, or mental, such as the reading of this text"; *Part Happiness* = "can concern a domain of life, such as working-life, and an aspect of life, such as its variety"; *Peak Experience* = "intense and oceanic" experience, also known as "enlightenment"; *Happiness* = as we understand and defined it above.

Another topic that needs to be discussed and distinguished from our happiness concept is *eudaimonia*. The term was originally created by Aristotle, who is today considered "the father" (Bruni, 2010, p. 391) of the eudaimonian happiness approach. According to him, eudaimonia can be equated with happiness. Similarly, "happiness is the final, or ultimate, end of life: [It] is the 'highest good' for the human being" (Bruni, 2010, p. 392). It is characterized as "something like flourishing human living, a kind of living that is

active, inclusive of all that has intrinsic value, and complete, meaning lacking in nothing that would make it richer or better" (Nussbaum, 2005, p. 171). Consequently, happiness can be reached by practicing virtues not in an instrumental way but in an intrinsically motivated way, where virtues are internalized and thus perceived as important and good to follow (*Aristotle's happiness paradox*; Bruni, 2010). Against this philosophical background, some researchers have suggested that not only affective and cognitive aspects but also eudaimonian aspects should be considered when defining happiness (e.g., Clark & Senik, 2011; Diener et al., 2010; Huppert et al., 2009; OECD, 2013; Ryan & Deci, 2001). Such aspects are meant to add individual judgments about the perceived degree "of meaning and purpose in life, or [of] good psychological functioning" (OECD, 2013, p. 10).

In general, the importance of considering the concept of eudaimonia in a definition of happiness should definitely be discussed in the future. However, to date, little research has confirmed its relevance for a definition of happiness, in addition to the affective and cognitive components. Instead, current research literature indicates that eudaimonia should be considered a moderating or influencing factor with respect to actual happiness rather than a clear component of happiness itself. For example, the OECD (2013) admits that the eudaimonian view on happiness brings a "more instrumental focus" (p. 32) with it than the perspective on affective and cognitive components. Further evidence for this point of view can be derived from investigations conducted in the context of *Self-Determination Theory* (SDT; Ryan & Deci 2000). According to this theory, three factors that are associated with self-realization or eudaimonia (autonomy, competence, and relatedness) basically contribute to an individual's degree of happiness (Ryan & Deci 2000, 2001). In addition to this content-related argumentation, further findings on the reliability and validity of eudaimonic measures are required (Dolan, Peasgood, & White, 2006; OECD, 2013) to be able to guarantee high psychometric quality when assessing eudaimonia in individuals.

In sum, no definition of happiness that is generally accepted currently exists. Thus, a concept that is as precise as possible and that fits with most common literature on affective and/or cognitive aspects of happiness is presented here. In our view, Overall Happiness can be equated with "the overall enjoyment of one's life as-a-whole" (Veenhoven, 2010, p. 611; cp. Veenhoven, 1994, 1997, 1984, 1991, 2008). Further, the Affective Happiness Component evaluates "the degree to which the various affects a person experiences are pleasant; in other words: how well he usually feels" (Veenhoven, 1991, p. 10; cp. Veenhoven, 1984, 2010). By contrast, the Cognitive Happiness Component covers "the degree to which an individual perceives his aspirations to have been met. In other words: to what extent one perceives oneself to have got what one wants in life" (Veenhoven, 1991, p. 10; cp. Veenhoven, 1984, 2010). These definitions now serve as a

basis for our selection of appropriate happiness measures to investigate them in terms of quality and applicability.

MEASURES OF HAPPINESS

The World Database of Happiness (WDOH; Veenhoven, 2016a) constitutes a library that fairly exhaustingly collects publications on happiness. In addition, it offers distributional and correlational findings that are calculated by the author and his co-workers themselves. In the context of this article, the WDOH is particularly helpful because it also offers a collection of happiness measures that are based on the abovementioned happiness definition (Veenhoven, 2016b). Currently[1], 2,118 measures are listed, and most are self-reports on single questions (1,516 measures, equalling 71.58%).

Classification

All accepted measures in the WDOH are classified by the (i) kind of happiness addressed, (ii) time frame, (iii) measure technique, and (iv) scaling. Each classification category is described in the following, based on Veenhoven (2015). Illustrative item examples are also given.

The kind of happiness addressed

As stated above, we assume that three components should be measured to capture happiness: Overall Happiness, the Affective Happiness Component and the Cognitive Happiness Component. In accordance with this view, all measures are assigned to one of these categories. For measures either that are ambiguous or that can definitely be classified to various categories, another fourth category is available (*Mixed Measures*).

Examples:

- Overall Happiness: "How do you feel about your life as a whole...?"
  - Rating: 7-point scale: 1: terrible – 2: unhappy – 3: mostly dissatisfied – 4: mixed – 5: mostly satisfied – 6: pleased – 7: delighted
  - Reference: Andrews & Withey, 1976
- Affective Happiness Component: "How is your mood these days...?"
  - Rating: 4-point scale: 1: not good almost all the time – 4: very good all the time
  - Reference: Levy & Guttman, 1975
- Cognitive Happiness Component: "How do you feel about what you are accomplishing in life...?"

1 As assessed on January 31st, 2017.

  - Rating: 7-point scale: 1: completely dissatisfied – 2: very dissatisfied – 3: dissatisfied – 4: satisfied-dissatisfied – 5: satisfied – 6: very satisfied – 7: completely satisfied
  - Reference: Buttel & Martinson, 1977
- Mixed Measures: "How many days in the previous week did you feel happy?"
  - Rating: 8-point scale: 0: none – 7: all
  - Reference: Simon & Nath, 2004

Time Frame

This category expresses the period of happiness addressed. The following time frames are included in the WDOH:

- Momentary, Now
- Last Instant
- Last Hour
- Last Part of the Day
- Last Day
- Yesterday
- Last Week
- Last Month, Last Few Weeks
- Last Quarter
- Last Year
- Last Years
- Over Lifetime
- Currently (Presently, Today, These Days)
- Generally
- Hitherto
- Since Event
- Various Time Frames (in Case of Mixed Measures)
- Time Frame Unspecified
- Time Frame Not Reported

Examples:

- Momentary, Now: "How are you feeling now...?"
  - Rating: 5-point scale: 1: very poor – 2: poor – 3: neither good nor poor – 4: good – 5: very good
  - Reference: Ventegodt, 1995
- Yesterday: "Overall, how happy did you feel yesterday?"
  - Rating: 11-point scale: 0: not at all – 10: completely
  - Reference: Office for National Statistics, 2012

- Last Year: "Generally, how happy have you been with your personal life during the past 12 months?"
  - Rating: 5-point scale: 1: unhappy most of the time – 2: sometimes fairly unhappy – 3: generally satisfied, pleased – 4: very happy most of the time – 5: extremely happy
  - Reference: Else-Quest, Hyde, & DeLamater, 2005
- Currently (Presently, Today, These Days): "All things considered, how would you describe yourself these days? Would you say you are...?"
  - Rating: 3-point scale: 1: not too happy – 2: fairly happy – 3: very happy
  - Reference: Kantor, Milton, & Ernst, 1978

Measurement technique

Here, the WDOH mainly distinguishes between self-reports on single or multiple closed questions, open questions or ego-documents (e.g., diaries) and ratings by others, such as clinicians, peers, own family or teachers.

Examples:

- Self-Report: "In thinking over the past year, indicate how elated or depressed, happy or unhappy you have felt in the last year?"
  - Rating: 10-point scale: 1: Utter depression and gloom. Completely down. All is black and leaden. Wish it were all over. – 2: Tremendously depressed. Feeling terrible, really miserable, "just awful". – 3: Depressed and feeling very low. Definitely "blue". – 4: Spirits low and somewhat "blue". – 5: Feeling a little bit low. Just so-so. – 6: Feeling pretty good, "OK". – 7: Feeling very good and cheerful. – 8: Elated and in high spirits. – 9: Very elated and in very high spirits. Tremendous delight and buoyancy. – 10: Complete elation, rapturous joy and soaring ecstasy.
  - Reference: Constantinople, 1967
- Rating by Others: "Overall how does your child usually feel?"
  - Rating: 7-point smiley scale: from sad face to happy face
  - Reference: Holder, Coleman, & Wallace, 2010

Scaling

Answers to questions can be given on different scales. The WDOH differentiates between four different scale-type categories. Verbal scales have each response option labelled, whereas numerical scales have only extremes defined. Graphical scales can be a scale with smiley faces, a ladder, a mountain or a thermometer, and the last category, miscellaneous scales, includes scales that cannot be classified among the former ones, such as the percentage of time being happy.

Examples:

- Verbal Scales: "How do you feel about your life as a whole...?"

  - Rating: 7-point scale: 1: terrible – 2: unhappy – 3: mostly dissatisfied – 4: mixed – 5: mostly satisfied – 6: pleased – 7: delighted
  - Reference: Andrews & Withey, 1976
- Numerical Scales: "How is your mood these days...?"
  - Rating: 4-point scale: 1: not good almost all the time – 4: very good all the time
  - Reference: Levy & Guttman, 1975
- Graphical Scales: "Overall how does your child usually feel?"
  - Rating: 7-point smiley scale: from sad face to happy face
  - Reference: Holder, Coleman, & Wallace, 2010
- Miscellaneous Scales:
  A: "What percentage of time that you were awake today did you feel happy?"
  B: "What percentage did you feel unhappy?"
  C: "What percentage did you feel neither happy nor unhappy?"
  - Rating: Percentages should in total sum up to 100%
  - Reference: Kammann & Flett, 1983

Table 2.3 reports how often each sub-category is listed in the WDOH in relation to specific population groups.

Psychometric Considerations

In happiness research, there has been a long-lasting debate about the extent to which the happiness construct can be reliably and validly measured (cp. Kahneman, 1994; MacKerron, 2012; Veenhoven, 1984, 2010). But research in this field today generally underlines the point of view that happiness can be reliably and validly measured (e.g., Diener et al., 1999; Diener, Suh, & Oishi, 1997; Judge & Kammeyer-Mueller, 2011; Kahneman, 1994; Kahneman & Krueger, 2006; MacKerron, 2012; Veenhoven, 2010). Test-retest reliability, for example, has been demonstrated over the course of various time frames (Bradburn, 1969; Bradburn & Caplovitz, 1965; Diener & Larsen, 1984; Diener et al., 2010; Fujita & Diener, 2005; Krueger & Schkade, 2008; Lepper, 1998; Lucas & Donnellan, 2012; Michalos & Kahlke, 2010; Watson, Clark, & Tellegen, 1988). In addition, various findings show the validity of happiness measures. For instance, positive correlations between self-rated happiness scores and happiness ratings delivered by significant others (such as friends and family: Lepper, 1998; Sandvik, Diener, & Seidlitz, 1993; Schneider & Schimmack, 2009), happiness ratings given by interviewers (Pavot & Diener, 1993), the frequency of smiling when individuals interact in a social context (Fernández-Dols & Ruiz-Belda, 1995), more left than right superior frontal cortex activity (Urry et al., 2004) and lower cortisol rates and lower heart rates (Steptoe, Wardle, & Marmot, 2005) demonstrate the convergent validity of happiness measures. Further, the discriminant validity of happiness measures has been shown by the low correlation between self-rated happiness

Table 2.3: Number of available studies for all classification categories, as listed in the WDOH*.

WDOH Category	Frequency of Studies
(i) Kind of Happiness Addressed	
Overall Happiness	459
Affective Happiness Component	114
Cognitive Happiness Component	1,243
Mixed Measures	311
(ii) Time Frame	
Momentary, Now	47
Last Instant	50
Last Hour	8
Last Part of the Day	7
Last Day	41
Yesterday	51
Last Week	100
Last Month, Last Few Weeks	115
Last Quarter	14
Last Year	46
Last Years	5
Over Lifetime	3
Currently (Presently, Today, these Days)	532
Generally	310
Hitherto	46
Since Event	9
Various Time Frames (in Case of Mixed Measures)	75
Time Frame Unspecified	498
Time Frame Not Reported	171
(iii) Measure Technique	
Self-Report	1,982
Rating by Others	141
(iv) Scaling	
Verbal Scales	1,288
Numerical Scales	556
Graphical Scales	139
Miscellaneous Scales	145

* As assessed on January 31st, 2017.

scores and optimism (Lucas, Diener, & Suh, 1996). Further results showing a positive correlation between self-rated happiness scores and specific life events (e.g., marriage) but a negative correlation between self-rated happiness scores and unemployment (Diener, 2012; Diener, Lucas & Scollon, 2006; Winkelmann & Winkelmann, 1998) demonstrate the criteria validity of happiness measures. Such evidence is further supported by positive correlations between self-rated happiness scores and level of income (Sacks, Stevenson & Wolfers, 2010), life circumstances (e.g., health status and social contact; Dolan, Peasgood & White, 2008) and daily activities (Kahneman & Krueger, 2006). In addition, low non-response rates for happiness indicators demonstrate the face validity of happiness measures (Kahneman & Krueger, 2006; Rässler & Riphahn, 2006).

Nevertheless, the WDOH has indicated that many measures are currently available to assess individuals' happiness. The high number of measures implies quality differences in terms of psychometrics (e.g., reliability, validity) and applicability in different contexts. The strengths and weaknesses of these happiness measures are introduced in the following.

Differences in time frames

A few years ago, Dolan, Peasgood, and White (2006) noted that many happiness measures hardly use exact time frames but instead ask about the respondent's life in general. Nevertheless, given the elaborated classification of happiness measures in the WDOH in combination with corresponding practical research activities, this statement needs to be rejected. In fact, the WDOH presently[2] lists numerous measures that use different frames. For example, 531 measures incorporate wording such as "currently (today, these days, presently)", 115 measures use wording such as "last month, last few weeks", 100 measures refer to the "last week", and 51 measures refer to "yesterday". Against this background, the question arises how respondents react to various time frames and then how they ultimately form their happiness judgments. With respect to the time frame currently (today, these days, presently), for instance, how past and future happiness and associated factors influence happiness questions containing this time frame remains unclear (MacKerron, 2012). Research conducted by Watson and colleagues (1988) has shown that people tend to answer similar questions under different time frame conditions slightly differently. In particular, the mean scores of the positive and negative affect subscales of their happiness measure increased with a lengthened time frame, indicating growing positive and negative affect. Given that respondents were to estimate the extent of a specific type of affect in, e.g., the last few weeks, the results seem reasonable. Indeed, as the time frame increases, the probability of experiencing a specific type of affect also rises.

2 As assessed on January 31st, 2017.

But what are the implications of the use of different time frames in happiness measures for psychometric quality? Generally, research on this relevant topic seems to be scarce (MacKerron, 2012), especially concerning validity issues. However, some important findings are available. Watson et al. (1988) found internal consistency to be unaffected by six variants of time frame. In contrast, test-retest reliability tended to increase with a larger time frame. According to the authors, the higher stability with larger time frames resulted from the respondents' approach of aggregating the types of affects they experienced on several occasions. By contrast, actual affect is more susceptible to change (Pavot & Diener, 1993); thus the stability is lower when it is assessed again at a later point in time. From a theoretical perspective, Pressman & Cohen (2005) argued that shorter time frames induce individuals to assess state happiness, whereas longer time frames induce them to assess trait happiness. As a main characteristic of traits consist of their relative stability over time, this statement provides a further explanation for the higher test-retest reliability when more extended time frames are used.

In sum, different time frames in similar questions seem to evoke varying response patterns. However, respondents' answers on not only questions with larger time frames but also on questions with shorter time frames show significant test-retest reliability (Watson et al., 1988). Thus, it currently seems reasonable to utilize happiness measures with various time frames.

Differences in number of items: Single-item measures vs. multiple-item measures

According to the WDOH, the majority of current[3] happiness measures use just a single item (71.58% of 2,118 measures in total). This finding is in line with the claim of several authors that such measures seem to be the standard in (large-scale) surveys (Clark & Senik, 2011; Diener et al., 1999; Huppert et al., 2009). Accordingly, Dolan and colleagues (2006) have underlined the quality of such measures in stating that a single item already makes it possible to identify differences in happiness among "those who are employed versus unemployed, single versus living with a partner, those who live in a state with good versus poor quality of governance, and so on" (p. 70). Test-retest reliability usually lies between .40 and .74 (Krueger & Schkade, 2008; Lepper, 1998; Lucas & Donnellan, 2012; Michalos & Kahlke, 2010). The correlation of two frequently used single-item measures, for instance, is .75 (Bjørnskov, 2010), indicating a degree of convergent validity. Nevertheless, measures that combine multiple items to capture happiness have better validity (Lucas et al., 1996) and test-retest reliability (range = .50-.83: Diener et al., 2010; Krueger & Schkade, 2008; Lepper, 1998; Lucas et al., 1996; Michalos & Kahlke, 2010). Additionally, correlations between single-item measures and multiple-item measures are not perfect (i.e., not equal to 1.0; Diener, Nickerson, Lucas, & Sandvik, 2002; Pavot

3 As assessed on January 31st, 2017.

& Diener, 1993), indicating that both kinds of measures do not completely capture the same construct. Consequently, it is not surprising that discussion about how to deal with the standard procedure of using only single items when measuring happiness is ongoing (Clark & Senik, 2011).

A first challenge that needs to be addressed when using single-item measures is methodological. Because respondents might understand the question differently (MacKerron, 2012) and because preceding items are prone to influence answers to the relevant happiness question (Huppert et al., 2009; Schwarz & Strack, 1999), there is a high probability of measurement error. In contrast, multiple-item measures contribute to reducing this type of error (MacKerron, 2012; Schneider & Schimmack, 2009), and thus, they enhance reliability by aggregating responses to numerous items (Krueger & Schkade, 2008). Higher reliability should thus positively influence validity. Schneider & Schimmack (2009) found evidence for this statement: in their happiness study, correlations between self-ratings and ratings by others were shown to be higher when multiple items were used instead of single items.

The second challenge with single-item happiness measures relates to theoretical concerns. In 2012, MacKerron wondered: "Is SWB reducible to a single dimension, and thus is it meaningful to ask – as single-item SWB questions often do – for a global evaluation of happiness, wellbeing, or satisfaction with life?" (p. 8). Following the abovementioned definition of happiness that assumes the existence of two distinct components (affective vs. cognitive), this question needs to be answered with "no". The reason is that such a two-dimensional construct cannot be covered by asking only a single question. To actually get information on both components, it is necessary to assess them separately. Consequently, using a multiple-item measure broadens the "breadth of coverage" (Diener et al., 2003, p. 208) of the happiness construct and thus captures it more validly. A few years ago, similar concerns arose in terms of covering affect completely. Today, the approach of investigating negative and positive affect separately from each other, as they seem to be distinct parts of overall affect, is widely accepted (e.g., Bradburn, 1969; Busseri & Sadava, 2011; Diener et al., 2010).

In sum, from a psychometric and a theoretical point of view, multiple-item measures for assessing happiness are to be favourable to single-item measures. Nevertheless, the psychometric quality of single-item measures is nevertheless sufficient for them to be used in surveys without a guilty conscience (Diener et al., 2003; Larsen, Diener, & Emmons, 1985). The applicability section below will discuss profound reasons why it may be sometimes beneficial to use a less reliable but quicker assessment of respondents' happiness and thus to use single-item happiness measures.

Differences in measurement technique: Self-report vs. rating by others?

Overall, happiness measures today are primarily based on self-reports (Diener, 2012). This finding is also supported by the WDOH, which indicates that there are currently 1,982 self-report measures but only 141 measures containing ratings by others[4]. However, experts have repeatedly noted the many weaknesses of self-reports (Diener, Suh, & Oishi, 1997). Indeed, they are affected by "information that is salient at the time of the judgment" (Diener et al., 2003, p. 196), as most happiness questions are rather complex to respond to and thus facilitate the use of heuristics (Schwarz & Strack, 1991). One type of information that seems to be used quite often to generate happiness judgements is current mood (Diener et al., 2003; Kahneman & Krueger, 2006; Schwarz, 1987; Schwarz & Strack, 1991), which can be influenced by the immediate context, such as the weather (Schwarz & Clore, 1983), the pleasantness of the room (Schwarz, Strack, Kommer, & Wagner, 1987) and the success of one's national team when watching a soccer championship game (Schwarz et al., 1987). By contrast, contradictory evidence indicates that temporary moods influence happiness ratings only marginally, especially in comparison with longer-term influences (Eid & Diener,1999). The influence of current mood on happiness judgments is also dependent on culture (Suh, Diener, Oishi, & Triandis, 1998). Indeed, respondents of individualistic-oriented cultures tend to particularly focus on their emotions as a reference for their happiness judgments, whereas respondents of collectivistic-oriented cultures use their emotions and cognition equally. Thus, the latter individuals are also oriented towards what significant others might think of the appropriateness of their current life. Besides possible mood influences, researchers have also noted the high potential for social desirability bias in happiness self-reports (Diener 2000; Sandvik et al., 1993; Schwarz & Strack, 1991), and this phenomenon has been found to be dependent on the situation. For instance, in his meta-analytic approach, Smith (1979) showed that respondents reported higher well-being when interviewed face-to-face vs. when assessed via mail survey. Consequently, avoiding personal contact in rating happiness can result in more honest judgments. Yet, some researchers have also obtained happiness judgments that were not contaminated by social desirability (Diener, Sandvik, Pavot, & Gallagher, 1991) – or at least not affected by social desirability to a substantial degree (Lepper, 1998). Another bias that can arise with self-reports results from placement or order effects, where the preceding questions may influence the following happiness question, e.g., by evoking specific feelings (Bjørnskov, 2010; Schwarz & Strack, 1991). Consequently, happiness questions should be asked at the very beginning of surveys to minimize this bias (Graham, 2005). A further challenge for happiness self-reports concerns individual response styles. For instance, individuals may favour high numbers in rating scales (Diener et al., 2003).

4 As assessed on January 31st, 2017.

In sum, to date, research has provided inconsistent information about the potential of various biases in self-reports. Nevertheless, the results suggest that cautious should be used with self-reports, as they are probably not free of unwanted influences. On the other hand, research evidence strictly supports the use of self-reports in surveys. For instance, based on test-retest reliability, there is substantial stability in happiness ratings from self-reports (r = .40-.83: Diener et al., 2010; Krueger & Schkade, 2008; Lepper, 1998; Lucas et al., 1996; Lucas & Donnellan, 2012; Michalos & Kahlke, 2010). Furthermore, in calculating the relative percentage of measurement errors in happiness judgements, Ehrhardt and colleagues (2000) found that a bias existed, but this bias diminished over the course of repeated surveys because of habituation processes and constant salient information that was used to make every judgment (Diener et al., 2003). Thus, particularly when asked several times over a longer time frame, people are capable of producing relatively stable self-reports. Accordingly, high correlations between self-ratings and ratings by others (Lepper, 1998; Pavot & Diener, 1993; Sandvik et al., 1993; Schneider & Schimmack, 2009) indicate that self-ratings cannot be so influenced by current mood or other circumstances (Lepper, 1998) to render them useless in happiness research. These findings have led several researchers to conclude that happiness can be measured by individual self-reports without a guilty conscience (Sandvik et al., 1993; Veenhoven, 2010). However, the above-described high correlations between self-ratings and ratings by others also underline the validity of ratings by others. For instance, it has been shown that raters who do not know the person to judge are able to precisely define his/her primary emotion at a specific point in time (Diener, Suh, Lucas and Smith, 1999). Nevertheless, it should be noted that significant others, such as family members or close friends, are not the person (or the "target") themselves who is actually judged. Thus, such a judgement "likely relies upon what the target has spoken, on how the target behaves, and on the expressions made by the target. This judgment made by a close significant other is likely a cumulative estimate of the verbal messages heard and the nonverbal expressions observed. A significant other likely has an abundant amount of information upon which to rely for such a judgment [...]" (Lepper, 1998, p. 368). However, as happiness is defined as "an attitude towards one's life" (Veenhoven, 1984, p.39), scholars have criticized that such an abstract construct can hardly be observed by others and that self-ratings instead should be favoured (Veenhoven, 1984). At this point, it can be argued that close friends of the target can judge the target's happiness appropriately because close friends usually talk to each other honestly and mutually share their current feelings. Furthermore, exhibited non-verbal emotional signals may deliver important emotional information to significant others. Such signals are difficult to hide (Sandvik et al., 1993) and are thus less prone to conscious distortion by the respondent. Additionally, external ratings by experts or clinicians have the vast advantage over self-reports and ratings by significant others in that they are the most objective of these alternatives (Diener et al., 2003).

Overall, although self-reports seem to deliver meaningful information about the respondent's happiness, it may be beneficial to include additional measures with their own strengths to get a more detailed understanding of the happiness construct (Diener, 2000, 2012; Diener et al., 2003; Diener et al., 1997; Krueger & Schkade, 2008; Pavot & Diener, 1993; Sandvik et al., 1993). These additional measures can then compensate for some weaknesses of self-reports (Diener et al., 1997). Such a multi-measurement approach seems to be path breaking since the use of such an approach has already led to deviations from previous study results obtained from self-reports (Diener, 2012). In such an approach, however, the above-discussed ratings by others do not represent the only happiness measures that can be combined with self-reports. Indeed, biological or physiological indicators, such as specific brain activation patterns (Urry et al., 2004), cortisol and heart rates (Steptoe et al., 2005), immunological parameters (Miller, Chen, & Cole, 2009) and electrophysiological measures (Diener et al., 1997) have shown to be potentially useful in complementing classical self-report happiness measures, as have recordings of facial emotional expressions (Harker & Keltner, 2001). The results of these alternative happiness measures highly converge with those of self-reports.

Differences in scaling: What scale should be chosen?

In happiness research, it is common to utilize a large variety of rating scales, such as rating scales with 3 to 11 points or more or face scales, to capture the construct of interest. Because of the use of different scales, it is difficult to make comparisons between happiness results from various investigations. For instance, it is well-known that the use of differing scales can lead to diverging answers (Lim, 2008). What are the critical points in creating response scales? What kind of scale should be chosen from a psychometric point of view?

The first issue concerns whether to use an uneven number of scale levels (containing a midpoint) or an even number of scales levels (forcing participants to take a position). Current research on this topic clearly indicates that the first option is preferred over the second one (Berk, 1979; Nowlis, Kahn & Dhar, 2002; Weijters, Cabooter & Schillewaert, 2010), as adding a midpoint to rating scales improves the reliability and validity of the judgments obtained, whereas omitting a midpoint forces participants to randomly choose one of the scale points that is closest to the midpoint (Krosnick & Presser, 2010). Thus, when happiness is measured, an uneven number of scale points is recommended.

The second challenge concerns the choice of the right number of answering options in total. In general, if there are many options available, participants are more likely to find a match between their attitude and one of the given answering options (Krosnick & Fabrigar, 1997). This point of view is supported by the finding that reliability rises with an increasing number of answering options (Diener, Lucas, Schimmack, & Helliwell, 2009). But one needs to be cautious, as using too many options can cause ambiguity and feel-

ings of being overwhelmed, which then can lead to enhanced satisficing (Krosnick & Fabrigar, 1997). According to Miller's research in 1956, which was accompanied by the statement of the "magical number seven, plus or minus two" (p. 81), 5-9 scale points should be perfect. In fact, current research underlines this claim in demonstrating that 5- and 7-point scales resulted in the higher reliability and validity scores than other scales (Krosnick, 1999a; Krosnick & Fabrigar, 1997). Some authors even clearly favour a 7-point scale, since it is commonly used in previous studies, it has high sensitivity and reliability, and people are able to classify their opinion precisely (which is more difficult with more scale levels, such as 9 or 11; Berk, 1979; Finstad, 2010; Menold & Tausch, 2015; Rohrmann, 1978; Weijters et al., 2010; Weng, 2004). In sum, research in this field indicates that a 5- or 7-point rating scale should be applied whenever possible.

Another critical issue in scale construction concerns the use of full labelling vs. partial labelling of individual scale points. Concerning this topic, researchers clearly favour the first alternative because, on the one hand, respondents admit to preferring it (Krosnick & Fabrigar, 1997) and, on the other hand, they obtain more reliable and valid results (Krosnick, 1999a, b; Krosnick & Fabrigar, 1997; Pasek & Krosnick, 2010). The primary reason for this finding seems to be that an explanation of each scale point facilitates respondents' understanding (Krosnick, 1999b; Pasek & Krosnick, 2010). Thus, defining the verbal label for every scale point is crucial, and researchers need to take care in selecting appropriate and clear labels that at least cover equally distributed intervals along the whole scale (Krosnick, 1999a, b; Krosnick & Fabrigar, 1997). In this way, researchers can obtain benefits from using fully labelled scales.

Another issue concerns the question whether to add numerical labels to already existing verbal labels. Research on this topic shows that people respond to scales containing the same verbal labels but diverging numerical labels differently (0 to 10 vs. -5 to +5; Krosnick & Fabrigar, 1997; Schwarz, Knäuper, Hippler, Noelle-Neumann, & Clark, 1991), indicating that respondents use numerical labels "to disambiguate the meaning of scale labels, resulting in different interpretations and, accordingly, different subjective scale anchors" (Schwarz et al., 1991, p. 570). Consequently, answers to the same questions obtained by using comparable verbal scale labels but differing numerical labels are difficult to compare. Furthermore, which results ultimately reflect respondents' true attitude remains unclear. Besides the distortion of answering patterns, another disadvantage of additional numerical labels concerns the increased cognitive burden for respondents (Pasek & Krosnick, 2010), which may lead to faster fatigue and a greater number of errors. In sum, research clearly suggests that numerical labels be omitted from rating scales.

Acquiescence caused by the use of agree-disagree rating scales is a common, well-known effect that also needs to be considered, as it can lead to distorted response patterns (Krosnick, 1999a; Saris, Revilla, Krosnick, & Shaeffer, 2010). Krosnick (1999a) listed possible reasons for this phenomenon, such as people's tendency to act politely

rather than oppositely in interpersonal situations or to defer to people of higher status. According to his and many other authors' investigations, the best method to eliminate acquiescence is to avoid agree-disagree rating scales when possible and to offer construct-specific response choices instead (e.g., "completely unhappy" – "completely happy"). Scales comprising such answering options have been shown to provide higher quality responses (Saris et al., 2010).

A final issue concerns graphical scales, such as the face scale, which is regularly used in studies of happiness (according to the WDOH, 32 different face scale measures[5]). Such face scales may particularly bring advantages over classical rating scales, as uneducated respondents can use them (Veenhoven, 1984), especially children who are still illiterate, and as they focus on the respondent's happiness evaluation based on their experiences (Veenhoven, 1984). By contrast, respondents perhaps pay too much attention to their emotions, and thus, face scales are meant to exclusively measure the affective happiness component rather than overall happiness (Veenhoven, 1984). Furthermore, to date, validation studies remain pending. Hence, what such scales actually measure and with what level of quality remains unclear. Thus, researchers should be cautious in applying them, as their effects on respondents remain uncertain.

Overall, on the basis of current research, the following recommendations for scale construction can be given: make use of (i) an uneven number of scale levels (with a midpoint) with at most (ii) 5 or 7 points and (iii) full scale level labelling. In contrast, omit (iv) additional numerical labels and (v) agree-disagree rating scales whenever possible. Be cautious with (vi) graphical scales, such as the face scale, as research on their psychometric quality is still pending.

Differences in study construction: Cross-sectional vs. longitudinal design?

According to the WDOH, cross-sectional data on happiness constitutes the vast majority of such data in comparison with longitudinal data (1,599 measures vs. 108 measures[6]). This finding is in line with the feelings of researchers claiming that cross-sectional studies compose the standard in happiness research despite their various shortcomings (Diener, 2012; Diener et al., 1999). Criticized points in particularly are that by exclusively relying on cross-sectional designs, causal directions remain unclear (Diener, 2012; Diener et al., 1999; Graham, 2005), and happiness ratings "can be disproportionately influenced by brief peak moods" (Pressman & Cohen, 2005, p. 963). To solve the latter problem, longitudinal designs can offer an aggregated estimate over the course of several points in time (Diener et al., 2003; Pressman & Cohen, 2005). By doing so, especially measurement errors can be avoided or at least reduced (Graham, 2005). Further advantages of

5 As assessed on January 31st, 2017.
6 As assessed on January 31st, 2017.

applying a longitudinal design are that it allows a much deeper investigation of the underlying processes (cp. Pressman & Cohen, 2005) and the correction of unobserved personality traits (Graham, 2005).

An exemplary method that can measure "daily moods over an extended period" (Sandvik et al., 1993, p. 321) is the experience sampling method (ESM: Csikszentmihalyi & Larsen, 2014; Csikszentmihalyi, Larson, & Prescott, 1977; Hektner, Schmidt, & Csikszentmihalyi, 2007) also known as ecological momentary assessment (EMA: Stone & Shiffman, 1994; Stone, Shiffman & DeVries, 1999). In investigations applying this method, researchers contact their respondents via a beeper (or nowadays via mobile devices; MacKerron, 2012) several times a day over the course of a longer time period and ask them to complete questions on what they currently do, where, with whom, and how they feel meanwhile. A recent study indicates that using their smartphone, respondents needed only 30 seconds to accomplish a single notification (Hendriks, Ludwigs, & Veenhoven, 2016). In this way, the ESM/EMA constitutes a method that can be applied in longitudinal panel designs that aim at surveying respondents in especially short time intervals (multi-moment assessment) in order to obtain additional time-use information. Besides the abovementioned benefits of a longitudinal design, the ESM/EMA offer more valid results than those obtained in cross-sectional studies with retrospective happiness assessments because memory on happiness should be more precise for a relatively brief time period than for a longer time period that needs to be retrieved (Sandvik et al., 1993). In addition, respondents may be more honest because admitting that a particular day was bad, as doing so is not as grave as claiming that their whole life is a mess (Sandvik et al., 1993). The ESM/EMA is additionally beneficial for assessing happiness in people who have problems with their memory because they can no longer possess based on their hippocampus. Such people exclusively live in the present and are thus still able to judge their current feelings (cp. Diener et al., 1997).

In sum, the still often used cross-sectional designs need to be extended by more sophisticated longitudinal designs (Diener, 2012; Diener et al., 1999; Frey & Stutzer, 2002; Graham, 2005). Such use of longitudinal designs seems to be especially urgent since recent investigations applying longitudinal designs have revealed deviating results from those obtained by regular cross-sectional design usage (Diener, 2012). As an example, the ESM/EMA offers a beneficial multi-moment assessment method to capture happiness in the long term.

To complete this psychometric section, table 2.4 summarizes all recommendations for high quality happiness measure construction, derived from the previous elaborated literature review and discussion. The next section addresses the numerous challenges that can arise when the highest psychometric standard is implemented in real investigations.

Table 2.4: Summary of recommendations about how to construct high quality happiness measures, as described in section 3.2 in detail.

Topic (Referring Section)	Recommendation(s)
Time Frames	Use the time frame that fits your research question, context and/or population best.
Single-Item Measure vs. Multiple-Item Measure	Prefer multiple-item happiness measures over single-item happiness measures.
Self-Report vs. Rating by Others	Complement self-reports of happiness with further methods with their own strengths, such as ratings by others, biological and physiological indicators or recordings of facial emotional expressions.
Scaling	Make use of: (i) an uneven number of scale levels (containing a midpoint) with at most (ii) 5 or 7 points and (iii) full scale level labelling. In contrast, omit (iv) additional numerical labels and (v) agree-disagree rating scales whenever possible. Be cautious with (vi) graphical scales, such as the face scale.
Cross-Sectional vs. Longitudinal Design	Implement longitudinal designs whenever possible.

Measure Applicability

Researchers who decide to incorporate happiness measures in their investigations generally use measures with the highest quality. Nevertheless, specific research questions, contexts or populations of interest may cause researchers to think about alternative, more feasible measures. The following section discusses circumstances that can justify the use of happiness measures that do not fulfil highest criteria for psychometric quality. Nevertheless, possible approaches that may help researchers revise the truthfulness of one's own reasons for not attaching to the highest psychometric standards are also presented.

Differences in time frames

Above we showed that from a psychometric viewpoint, there is currently no reason to avoid using a specific time frame when constructing happiness measures. Consequently, the time frame that the researcher uses strongly depends on the question of interest and the population surveyed. Imagine, for example, an investigation in a company that needs to be conducted to derive significant interventions for the sake of achieving higher well-being and productivity among employees in the long term. In this case, the status quo is the main interest, and thus, a relatively short time frame is useful. In contrast, a further example illustrates that for another context, a longer time frame could be beneficial. Indeed, national (panel) surveys are often conducted at regular time intervals of about one year or more (such as the Eurobarometer Survey: EC, 2016; Gallup World Poll: Gesis, 2012; Socio-Economic Panel [SOEP]: DIW, 2016; World Values Survey: WVS Association, 2016), and they generally strive to determine the nation's happiness over the

course of several years or even decades. Thus, covering the entire time interval from one survey to another is of tremendous relevance, and determining average happiness in the previous year, for example, does make sense such these circumstances. Furthermore, the importance of time frame in creating happiness measures is demonstrated in the following example: Some people no longer have a hippocampus, leading to cognitive impairment for past memories (cp. Diener et al., 1997). Such people are exclusively able to judge their feelings with respect to the present. In addition, children have difficulties in remembering past experiences, owing to maturity (Friedman & Kemp, 1998; Hayne & Imuta, 2011). Consequently, for both target groups, responses of higher validity can be obtained by utilizing relatively short time frames.

Differences in number of items: Single-item measures vs. multiple-item measures

From a psychometric perspective, we showed above that multiple-item measures are better than single-item measures even though the latter can still be used without losing too much quality. In fact, there may be specific conditions under which the use of single-item measures may be preferable because of practicability issues. For example, large-scale (panel) surveys aim at depicting respondents' views on a huge variety of topics. Consequently, the capacity to evaluate every topic remains limited, meaning that only a few items can be used, as the burden on respondent should be kept at an acceptable level (cp. Diener et al., 2003; Lucas & Donnellan, 2012). Thus, this context may justify the use of lower quality single-item measures instead of multiple-item measures. Another scenario that may lead to the use of single-item measures as a favourable approach arises as soon as observations are required in very short time intervals (such as multi-moment assessments). Then, multiple-item measures might cause tremendously high drop-out rates because respondents associate too much effort with the corresponding investigation. Hence, in this case, single-item measures could be the preferable option. A further critical issue can arise if the survey aims to investigate people with cognitive problems, such as attentional deficits. For this population, it might be beneficial to keep the investigation as brief as possible in order to guarantee the highest validity attainable. Furthermore, imagine a survey target population that currently exercises everyday life and that is to be surveyed spontaneously during, for instance, walking through a pedestrian zone. Such a situation frequently evolves when market research is conducted among the general population. A crucial issue is that people often have numerous responsibilities in the moment and are thus not willing to interrupt them in order to complete a questionnaire of 250 items. Consequently, keeping the survey as short as possible by only using single-item measures for every construct should help obtain reasonable participation rates.

In sum, there are several understandable reasons for not choosing the highest psychometric standard and instead using single-item measures only. But where is the proof

that all these assumptions reflect reality? Hence, the validity of these reasons should be considered. In this way, researchers can verify whether response rates actually diminish, whether the perceived burden for respondents actually rises and whether cognitive disabled persons actually provide answers of lower quality. In the best case, all concerns can be proved as unfounded, and thus, there are no longer reasons to further omit multiple-item measures as more valid measures of the happiness construct. One opportunity to verify the abovementioned assumptions consists in simply splitting the sample, where half completes the single-item version and the other half completes the longer, multiple-item version. Another possibility for longitudinal designs, such as multi-moment assessment or classical national-wide panel surveys, can be to generally use the single-item version but to sometimes use the more intense, multiple-item version. With these two suggestions, researchers can revise their reasons for choosing single-item measures and maybe change their view.

Differences in measurement techniques: Self-report vs. ratings by others?

Above, we noted that from a psychometric view, self-reports do not deliver information of the highest quality possible because of likely answering distortions. Thus, the use of complementary measures is recommended to obtain more valid results. But despite the weaknesses of self-reports, we also stated that self-reports can capture meaningful information on happiness. Thus, depending on context and population, using only self-reports or only ratings by others can be justifiable. As previously mentioned, especially large-scale (panel) surveys need to take care of respondents' burden and thus limit the number of items used to cover the constructs of interest. In addition, monetary aspects play a role, as funding made available for such studies is definite. Overall, both arguments indicate that particularly for large-scale studies, it can be more feasible to use only self-reports for happiness instead of adding further, complementary measures, such as ratings by others. By contrast, for extremely mentally disabled persons, it can make sense to generally omit self-reports and apply alternative measures to assess individuals' level of happiness.

As similarly stated above, all these arguments support comprehensible reasons for avoiding the use of complementary happiness measures. Nevertheless, the extent to which these reasons can be considered well-grounded remains unclear. Consequently, these assumptions need to be verified in terms of their correctness: Is the respondent burden actually increased if complementary happiness measures are added to the survey? Is money definitely too restricted as to allow the application of additional measures? Are mentally disabled persons really not able to provide meaningful information about their current happiness? The above suggested methods may help to shed light on these questions.

Differences in scaling: What scale should be chosen?

It was generally noted that using variously constructed scales can lead to differences in respondents' answers. Thus, if possible, researchers should follow the recommendations above to fulfil the highest psychometric standards. However, there are exceptional situations in which relying on commonly used but lesser quality measures can be beneficial. Some surveys have been conducted for many years and thus have a tradition that cannot be denied. Particularly with respect to comparability issues, keeping the current scaling constant may be a better alternative. Consequently, spontaneously changing the scales used can even cause disadvantages, such as worse comparability with previous investigations. Concerning graphical scales, such as the face scale, it has been claimed that psychometric research in this field is not available yet. Nevertheless, for assessing children and mentally disabled persons, such scales may be the best option. Namely, using alternative rating scales may result in answers of much lower quality, as the respondents are not capable of answering such questions. Thus, using a face scale under such circumstances can definitely make sense. To conclude, it nevertheless needs to be posited that under the premises of a normal population and a completely new survey, researchers should consider the recommendations here concerning scale construction. In this way, the highest quality standards can be achieved.

Differences in study construction: Cross-sectional vs. longitudinal design?

From a psychometric point of view, longitudinal designs offer many advantages over cross-sectional designs and should be implemented, if possible. But the context and investigated population are also important, and they can shift the preference towards a cross-sectional design. Imagine, for example, a representative, longitudinal survey conducted among refugees. Such a target group is tremendously difficult to investigate over a longer time period, as many refugees change their living place frequently. Consequently, they may not be able to participate in subsequent surveys and may thus drop out of the panel. Thus, for the group of refugees, a detailed, cross-sectional investigation instead of a longitudinal survey, which has considerable potential to suffer from a high drop-out rate, may be beneficial. However, in the long-term, solutions should be found to facilitate longitudinal investigations on such challenging populations. As will be shown in the following paragraph, applying newly developed, digital technologies implement complex longitudinal designs may be a promising strategy for future studies.

Another critical point in applying longitudinal designs consists in their often extraordinarily high (time and monetary) costs for both respondents and researchers. Consider, for example, the ESM/EMA that was introduced above to depict multi-moment assessment, a specific method used in longitudinal designs. When this method is implemented, respondents are asked to complete questions on their happiness, several times a day over the course of a longer time period. In the original study, answers were recorded via

paper-pencil usage; thus, respondents needed to carry this equipment with them all the time (Csikszentmihalyi et al., 1977). Consequently, researchers are required to make ESM sheets and pencils available, and in the end, they need to manually transfer the entire data sets to the PC. Overall, it is not surprising that the ESM/EMA is repeatedly described as a rather time consuming and costly method to capture happiness during everyday life (Dolan & White, 2007; Kahneman, Krueger, Schkade, Schwarz, & Stone, 2004b; Krueger & Schkade, 2008). It was also criticized that the ESM/EMA is able to show only excerpts of everyday life instead of depicting the whole day and thus allowing precise time use information (Kahneman et al., 2004b). Because of these shortcomings, Kahneman and colleagues (2004) developed the *Day Reconstruction Method* (DRM), where respondents are regularly asked to reconstruct their prior day in single episodes, for instance, by filling in a diary (e.g., 8-9 o'clock breakfast, 9-12 o'clock work). Thus, they indicate which activities they undertook in which time period, where they did, with whom, and how happy they were doing so. In addition, the DRM was initially primarily implemented via paper-pencil format (Kahneman et al., 2004b). Although it could be found that DRM data highly converges with ESM/EMA data (Kahneman & Krueger, 2006; Kahneman et al., 2004b), it is not necessary to abandon the ESM/EMA completely. On the contrary, this method especially needs to become more efficient for researchers to apply and for respondents to use. However, the DRM would also benefit from such an approach. As mentioned before, most DRMs are still recorded via paper and pencil. For this purpose, the current digitalization process brings new possibilities for applying both methods in a more efficient way than before. In recent times, diverse researchers launched mobile phone applications containing ESM/EMA (Killingsworth & Gilbert, 2010[7]; MacKerron, 2011[8]) and additionally DRM (Hendriks, Ludwigs, & Veenhoven, 2016). In this way, the study of Hendriks et al. (2016), for example, shows the enormous cost and time efficiency obtained by conducting the study on a digital basis: Over the course of two weeks, more than 16,000 happiness ratings and 4,000 diaries could be achieved with only a single researcher running the whole study and total costs of about 500€ (excluded are costs of the researcher). The burden for every respondent in terms of time was about 2 hours a week at maximum, and all data were automatically saved and analysed via app without the need to manually code and evaluate the data. In sum, it becomes obvious what opportunities can arise when researchers dare to embrace new digital technologies in forthcoming investigations. Thus, it is promising that newly invented technologies can facilitate the implementation of longitudinal and especially multi-moment assessment designs.

7 Free app download: http://www.trackyourhappiness.com (lastly retrieved on: 2017-01-31)
8 Free app download: http://www.mappiness.org.uk (lastly retrieved on: 2017-01-31)

SUMMARY AND PROSPECTS

At the beginning of this book chapter, we showed that it is worthwhile to research happiness since it brings a large variety of benefits for every individual and for society. Concurrently, we need to find out more about this construct to achieve the ambitious goal of making people happier in general. In particular, appropriate measures are needed to capture the happiness of various people in different situations as validly as possible. Based on the WDOH, the status quo of existing happiness measures was noted, with a particular focus on their strengths and weaknesses. Researchers should consider these findings in selecting their happiness measures in future investigations. Particularly the continuing digitalization process will enable happiness researchers to easier include multi-moment and multi-method assessments in their studies. If they do so, results will be obtained that were impossible to obtain a few years ago. However, considerable progress needs to be made in this field with respect to the following questions: How can we guarantee representative samples especially with respect to extremely young and elder respondents? Do respondents change their behaviour over the course of the investigation, owing to repeated monitoring (cp. Dolan & White, 2007)? There are also certainly many other questions one can think of. Answering these questions to guarantee high quality measures by using new technologies should be an important goal of forthcoming studies.

To conclude, we can state that despite initial concerns about the question whether happiness could be measured satisfactorily, we can now state that considerable progress has been made in this field. However, researchers should be cautious about keeping up to date with the newest findings on the psychometrics of the happiness measures. If they discover weaknesses in their measures and if it is feasible to revise them, they should dare to do so to guarantee the highest quality standards. In particular, openness towards new developments owing to digitalization processes is required, as arising methods in this field are promising for further enhancing currently available happiness measures.

3 The Happiness Analyzer – a new technique for measuring subjective well-being

ABSTRACT

In 2013 the OECD published a guideline that represents the gold standard for measuring subjective well-being in greater detail to collect data in the quality needed as a basis for efficient decisions to improve subjective well-being and the evaluation of those decisions to enable continuous learning. Unfortunately, most existing studies nevertheless do not measure subjective well-being according to this standard, as traditional methods (e.g., paper and pencil or personal interviews) require considerable resources (from both researchers and participants) to capture i) people's subjective well-being at multiple points in time using general questionnaires; ii) people's everyday life and everyday feelings; iii) people's specific feelings in the moment; and iv) a combination of subjective and objective well-being measurements. To resolve this issue, we developed an app as a mobile assessment tool, the "Happiness Analyzer" which is outlined in this paper.

Keywords: happiness research; subjective well-being; multi-moment assessment; experience sampling method; day reconstruction method; mobile application

INTRODUCTION

Rise of SWB research

In recent years, happiness research has become a prominent topic in academics, politics and the media. Many people have begun to realise that after a certain level, more consumption and higher material well-being do not make people happier (Clark et al. 2008). Thus, it is important to conduct research to understand who is happy and how happy they are and particularly to understand what makes different people happy and what does not. To accomplish this, substantial correlative data regarding people's "subjective well-being" (SWB), which is used in this paper as a synonym for happiness, has been collected for many years by, for example, the Gallup World Poll and the European Social Survey. This collection of data represents a generally positive development, but as researchers and practitioners have analysed these data, they have realised that better data are required to understand the underlying mechanisms of SWB (e.g., the precise effect a certain intervention has on different life domains and happiness evaluations), particularly for studies at the micro (individual) and meso (city or corporation) levels. Thus, the OECD (2013) published guidelines that represent the gold standard for measuring SWB in greater detail.

Call for better measurement

According to the OECD, SWB is a construct consisting of three elements: i) life evaluation – a reflective, cognitive judgement of a person's life or specific parts of it; ii) affect – a person's positive and negative emotions and feelings; iii) eudaimonia – according to Aristotle's 2000-year-old construct, a person's judgement of his or her life life in terms of meaning and purpose in life (for more details on these definitions, see OECD 2013).

To measure these different elements, the OECD suggests using six different modules: i) a core module about happiness and life-satisfaction with a single question; ii) an affect module with multiple specific questions; iii) a life evaluation module with multiple specific questions; iv) an eudaimonic well-being module with multiple specific questions; v) a domain evaluation module with multiple specific questions about satisfaction in specific life domains (e.g., health); and vi) an experienced well-being module. The guidelines recommend using the gold standard experience sampling method (ESM; Csikszentmihalyi and Hunter 2003), in which people record how they feel and what they are doing, with whom and where at specific moments in time in affective time-use diaries, or the day reconstruction method (DRM; Kahneman et al. 2004), in which people reconstruct their day in episodes (e.g., breakfast from 8-9am) and rate how happy they felt during those episodes. In addition to these modules, the OECD generally advocates for more longitudinal studies instead of cross-sectional studies and for linking SWB data

to objective data, including location data, economic variables or biological markers such as heart rate variability, face emotion recognition and others.

The requirements can be ordered according to an onion model, as displayed in Figure 3.1, consisting of three layers within a frame. i) General Measurement: A general SWB questionnaire including OECD modules one to five that is designed to obtain a cognitive measurement. ii) Activity-based Measurement: Daily life and daily affective experience measurement employing affective time-use diaries via techniques such as the DRM to comprehensively capture people's time use to obtain more contextual information. iii) Experience Sampling: Affective measurement in the moment using, for example, the ESM. iv) Objective Markers: Around the subjective layers is an objective frame integrating other objective markers to increase validity, for example, location data, economic variables, and biological markers such as heart rate variability or face emotion recognition as noted above. Unlike the OECD, we separate the DRM and ESM into two different layers. We agree that both methods primarily measure experienced well-being, but the DRM provides far more contextual information about a person's life and activities

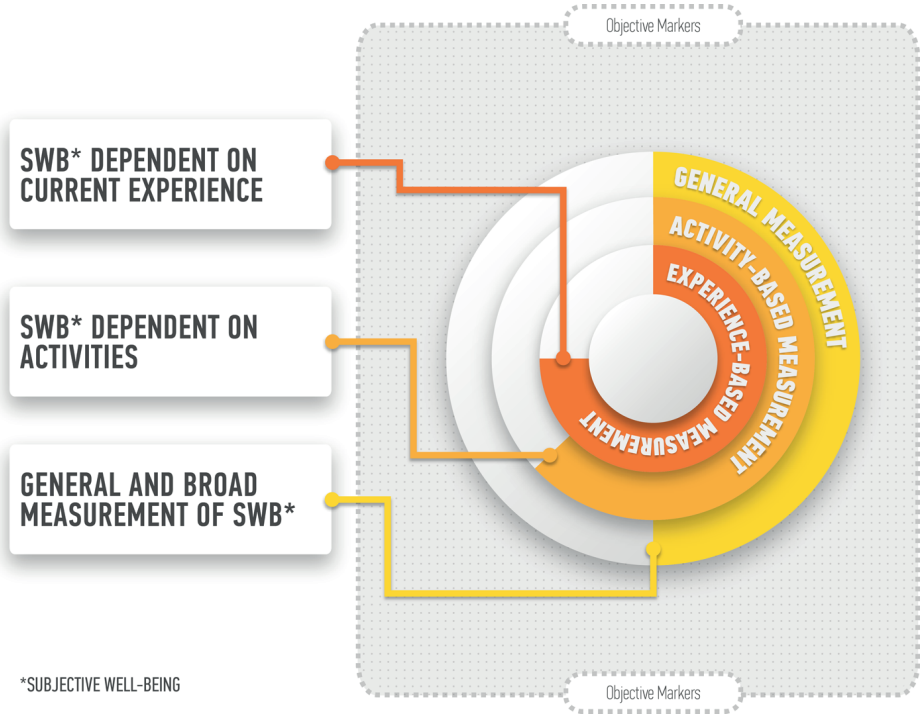

Figure 3.1: The onion model for measuring subjective well-being: The figure displays the Onion Model which is our approach to fulfil the OECD gold standard for measuring subjective well-being. More details are explained in the text.

than the ESM because the DRM collects information about the entire 24 hours in a day, rather than simply a few moments. Thus, the DRM helps in obtaining a more detailed understanding of the underlying mechanisms of SWB.

Unfortunately, most existing studies do not measure SWB according to this standard because considerable resources are needed (from both researchers and participants) to capture the following information: i) people's SWB at multiple points in time using general questionnaires; ii) people's everyday life and everyday life feelings; iii) people's direct feelings in the moment; and iv) a combination of subjective and objective well-being measurements such as people's subjective ratings of SWB and their objective stress level indicated by, for example, heart rate variability.

Table 3.1 displays a set of questions for the subjective well-being layers that are in line with the OECD guidelines.

Table 3.1: Measures of subjective well-being according the OECD gold standard

Measure	Items	Scale Range
Happiness Core (HC)	Taking all things together, how happy would you say you are?	0: Extremely unhappy 10: Extremely happy
Life Satisfaction Core (LC)	All things considered, how satisfied are you with your life as a whole nowadays?	0: Extremely dissatisfied 10: Extremely satisfied
Scale of Positive and Negative Experience (SPANE)	How often did the interviewed person experience the following emotions in the last two weeks: 1: Negative 2: Unpleasant 3: Good 4: Bad 5: Happy 6: Afraid 7: Pleasant 8: Contended 9: Sad 10: Angry 11: Joyful 12: Positive	0: never 7: always
Satisfaction With Life Scale (SWLS)	Indicate your agreement which each item: 1: In most ways, my life is close to my ideal 2: The conditions of my life are excellent 3: I am satisfied with my life 4: So far, I have gotten the important things I want in life 5: If I could live my life over, I would change almost nothing.	1: Strongly disagree 7: Strongly agree
Flourishing Scale (FS)	Indicate your agreement with each item: 1: I lead a purposeful and meaningful life 2: My social relationships are supportive and rewarding 3: I am engaged and interested in my daily activities 4: I actively contribute to the happiness and well-being of others 5: I am competent and capable in the activities that are important to me 6: I am a good person and live a good life	1: Strongly disagree 7: Strongly agree

Table 3.1 (*continued*)

Measure	Items	Scale Range
	7: I am optimistic about my future 8: People respect me	
Domain Evaluation Questionnaire (DEQ)	The following questions ask you how satisfied you feel about specific aspects in your life: 1: Standard of Living 2: Health 3: Productivity 4: Personal relationships 5: Safety 6: Community 7: Personal Security 8: Free time 9: Environment 10: Job	0: Not at all satisfied 10: Completely satisfied
Day Reconstruction Method (DRM)	What did you do in this period? Where have you been in this period? Who was with you in this period? How did you feel during this episode?	0: Unhappy 10: Happy
Experience Sampling Method (ESM)	How do you feel right now? What are you doing right now? Where are you right now? Who is with you right now?	0: Unhappy 10: Happy

Current e-tools

Due to the rapid proliferation of the Internet and smartphones in particular over the past years, it is possible to collect more data with less resources. In 2009, when the OECD began its work on the guidelines and the shortcomings were already apparent in the academic world, Killingsworth and Gilbert published an iPhone application that used the ESM to track people's happiness (Killingsworth and Gilbert 2010; www.trackyourhappiness.com). The app first performed a general measurement using an app survey and then contacted people a few times per day with brief notifications asking them to rate their happiness on a scale and to explain what they were doing, with whom and where. A similar app was developed by MacKerron in 2011 to construct a happiness map of Great Britain. Both apps reward participants by providing them with graphical feedback in the form of a happiness profile. In addition, both apps use GPS to track participants' location as an objective marker (MacKerron 2012; www.mappiness.org.uk). Both apps were highly successful; for example, there have been over 60,000 downloads of mAPPiness.

A major restriction of these apps, however, is that they work only on iPhones. Given that not everyone can afford an iPhone, it is difficult to build representative samples using these apps. Additionally, the apps do not capture layer two (the activity-based measurement) and only partially capture the frame (objective markers); thus, they do

not solve problem two (comprehensive data about people's everyday life and everyday life feelings) and only partially address problem four (combination of subjective and objective well-being measurements) through the integration of a GPS module.

In 2012, Veenhoven, Bakker and Oerlemans developed a different approach, the "Gelukswijzer" or "Happiness Indicator" (www.gelukswijzer.nl; Bakker et al. 2016). The Happiness Indicator is a webpage that offers people the opportunity to report their happiness score as often as they wish, and participants are reminded via email to track their happiness. In return, participants receive a happiness profile that illustrates their happiness trend. Additionally, the authors integrated an applied form of the DRM. Because the Happiness Indicator is a webpage and not an iPhone app, more people – especially older people – can use it, and a more representative sample can therefore be collected. Over 80,000 people have now signed up to track their happiness. Unfortunately, analyses of the data collected via this instrument indicated that very few participants reported their happiness score at least once per day and that only a few participants provided their DRM responses daily over a longer period of time, likely because it is not feasible for users to reply to e-mail notifications on a regular basis (Bakker et al. 2016). In addition, the web-based tool does not offer the option of using the ESM. Overall, the Happiness Indicator solves problem one and generally solves problem two, but it is not capable of solving problems three or four. Table 3.2 summarises the features of the various tools.

Table 3.2: Comparison of the features of recent happiness tracking tools and the Happiness Analyzer.

Features	Track Your Happiness	mAPPiness	Happiness Indicator	Happiness Analyzer
Survey SWB using general questionnaires at multiple points in time	✓	✓	✓	✓
Survey people's everyday life and everyday life feelings using affective time-use diaries such as the DRM	•	•	(✓)	✓
Survey people's direct feelings in the moment using, for example, the ESM	✓	✓	•	✓
Combining other objective and subjective indicators to obtain a more valid measurement	(✓)	(✓)	•	(✓)
Multiplatform (Web, Android, iOS) for representative sampling	•	•	•	✓

✓ = has the feature; (✓) = feature partially implemented; • = does not have the feature.

To overcome the aforementioned problems, we developed the "Happiness Analyzer" based on the OECD guidelines and the onion model to solve problems one, two and three using a multiplatform tool. For problem four, thus far we have conducted evaluations and applied a GPS module, and we are currently experimenting with additional

sensors and markers, which will be explained in greater detail at the end of this paper. In part two, we outline the method in more detail. In part three, we explain and discuss the main evaluation results of the most recent evaluation study. In part four, we conclude and outline future prospects.

METHOD: THE HAPPINESS ANALYZER

In this section we will outline the method by first summarizing all functionalities of the Happiness Analyzer and then conclude with some screenshots and a screencast of the Happiness Analyzer in figure 3.2.

Functionalities

General SWB measurement
The survey tool administers the questionnaires included in the OECD modules and collects demographics. To measure SWB according to the onion model, the basic version uses a SWB module that requires between six and eight minutes of the participants' time (see table 1). This survey as well as the DRM (with e-mail-reminders) can also be done on a website if a participant would prefer to answer the survey and use the DRM on a bigger screen. The ESM can just be used on a smartphone.

Activity-based measurement
To measure people's activities and assess their time use and how happy they are during their activities (Happiness-Time-Use), the DRM developed by Nobel prize laureate Daniel Kahneman is applied in the app. At 9pm, participants receive a notification to reconstruct the past day. Additionally, the app offers participants the option to make notes about the day, take a picture or record an audio note after they have reconstructed their day. This feature is not mandatory and can be excluded with minimal efforts by the developers if the third party researcher does not want to collect this kind of data.

Experience-based measurement
To measure people's happiness in a more affective and detailed manner, the tool can notify participants at specific or random moments. In the standard setting four notifications are sent out randomly to participants between 8am and 8pm with a minimum time gap of 2.5 hours between two notifications asking them to record i) how happy they are, ii) what they are doing, iii) with whom, and iv) where. The frequency of notifications and the time-frame can be adapted with minimal efforts by the developers depending on the research questions and the wishes of the third party researcher. Additionally,

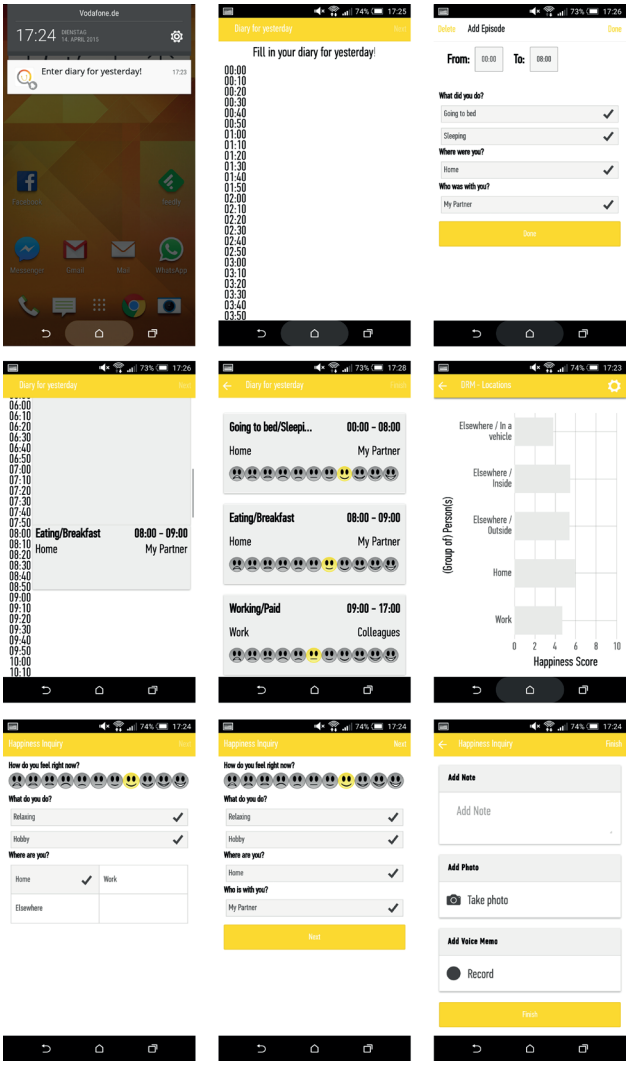

Figure 3.2: The Happiness Analyzer: The figure displays the Happiness Analyzer. In the first screenshot on the left top you see a notification to use the DRM. The second screenshot on the top middle position displays how participants can select an episode of their previous day. The third screenshot on the top right displays the screen that pops-up when an episode is selected. The participant can adapt the episode time, define what he did, where and with whom by selecting from different answering options. As displayed in the fourth screenshot on the middle left the participant can then continue to select more episodes until the whole day is reconstructed. Then the participant can rate how happy he felt during the different episodes, which is displayed in the screenshot in the middle. All results of the DRM and the ESM are displayed in bar-charts to show how happy the participant felt on which day, during which activities, at which locations and in which social setting. The last three screenshots display the ESM where a participant has to rate how happy he feels right now, what he is doing, where and with whom by selecting from different answering options. If the participant wants to add a note, a picture or an audio note he can do this afterwards, which is displayed on the last screenshot on the bottom right position. A screencast can be viewed at https://vimeo.com/135966012.

specific questions or a higher frequency of questions can be asked at a specific point in time. After responding to these questions, participants have the option to make notes about the described moments, take a picture or record an audio note. This feature is not mandatory and can be excluded with minimal efforts by the developers if the third party researcher does not want to collect this kind of data. Event triggered notifications are not supported in the version presented in this paper.

Biological markers

The survey tool can notify people at given points in time when they should carry out, e.g., a saliva probe or collect other biological markers. Regarding saliva, it is possible to, for example, analyse people's cortisol or testosterone levels, which appear to be good indicators of well-being.

Graphical feedback

All the collected data are displayed in real time to the participants in a visual format. This should help raise the participants' awareness about what makes them happy and what does not. Additionally, this feedback should motivate people to participate longer in a study to build their individual happiness profile. Participants get graphical feedback in bar charts, separated for ESM and DRM, on how happy they were on which day, how happy they are during which activities, how happy they are at which locations and how happy they are in which social setting. Some examples are shown in figure 2. If researchers are concerned that this feedback feature could be a reason for biases (e.g. that people act different or become happier because of their increased awareness) the feature can be excluded with minimal efforts by the developers if the third party researcher does not want to include this feature.

Design and flexibility

To ensure high response rates and large samples, the design and name of the app can be adapted for every individual study.

Multiplatform/Multidevice

To ensure that studies conducted using the Happiness Analyzer are as accessible as possible, the questionnaires and the DRM (with e-mail-reminders) can be answered in the browser on desktop and mobile devices. The ESM can only be used when using a mobile device. In addition, the Happiness Analyzer runs as a native app on iOS and Android. Thus, participants can be directly notified about available questionnaires using smartphone notifications.

Cross-device sync

The Happiness Analyzer can be used on multiple devices in parallel. When responding to a questionnaire on one device, the data are directly synced to other connected devices (Internet connection required).

Offline support

Once installed, the native smartphone app does not require a continuous Internet connection. The participant can be notified about all the questionnaires and respond to them without a data connection. An Internet connection is required only at the end of the study to sync the entered data with the server.

Smartphone sensors

The Happiness Analyzer has access to sensors built into smartphones, such as camera, microphone and GPS. The corresponding data can be retrieved in addition to the information manually entered by the participants while responding to a questionnaire.

Localization

The app has localization support, i.e., the language, date format, and so on adapt to the device language.

Data export

The data can be exported in many different formats with minimal efforts of the developers and used in various analytical software packages, such as STATA, SPSS, and others. This also includes an export of the qualitative data entered as notes, photos or audio notes at the end of the DRM or ESM.

User identification

To guarantee the anonymity of the study participants, i.e., to ensure no personal information collected via the Happiness Analyzer is linked to the actual person who submitted the data, participants are identified by a cryptographically secure randomly generated usercode that is created on the first use of the app. Thus, any association of the identity used in the app to the actual user of the app, which might be possible using a user-provided identity such as an email address or username, is impossible.

As a second option for user identification, a list of random usercodes can be generated before the study begins, and these generated codes can be transferred to the study participants who then use the codes to log in to the app.

Data storage

The study data are stored on servers that are directly managed by the Happiness Research Organisation (HRO; www.happiness-research.org), an independent German research institute, and located in Germany. To maintain control of the data and of their physical location and to remain independent of any third party, the HRO chose not to use cloud-based data storage. The information stored for a participant can be deleted at the request of the participant or the third party researcher.

EVALUATION

In 2013, we had the idea to develop the Happiness Analyzer. As of 2017 we have ran various evaluation studies and applied the tool in different versions for various research studies (as an example see Hendriks, Ludwigs & Veenhoven, 2016). In this section, we want to outline the focus group feedbacks of our last detailed evaluation study with psychology students at the University in Düsseldorf.

To improve the Happiness Analyzer, the app was evaluated in a research study with psychology students at the University in Düsseldorf in November & December 2016 (112 users). The participants were required to use the app for two concurrent weeks, during which they completed the general SWB questionnaire, answered some demographic questions at the beginning of the app and then responded once per day to a DRM notification and 4 times per day to an ESM notification. In addition, we tracked the participants' location via GPS. As a reward, the participants' tracking results were displayed graphically in the app to increase their awareness of what makes them happy and what does not. All the app's features performed well during this study, and we collected the data according to the onion model (see table 1). The most important feedbacks that we collected in focus groups at the university in Düsseldorf after the study are discussed below.

The participants' first major critique was that the response options in the DRM needed to be optimised. Over half the participants said that the options were not sufficiently detailed. Precise response options are required to understand people's everyday life and everyday life feelings in a quantitative manner. To achieve this goal, we will continue to analyse the response options in the free fields, endeavour to adapt the response options to offer greater detail according to specific samples and studies and conduct qualitative research to complete our list of possible activities.

The participants' second major critique was that the app's artificial intelligence (AI) was insufficient. The participants wanted the app to learn from and adapt the response op-

tions, e.g., when a participant selects "commuting", the app should pre-select the option "in a vehicle" in the "where" field. We have already developed an idea for an AI that will be able to suggest popular choices, for example, based on the user's location according to GPS tracking. Each of these steps must be discussed carefully because it is important to avoid the risk of biasing the ratings and choices and to prevent participants from feeling that the app is tracking them too closely. Instead, participants should continue to perceive the app as a private diary that collects data anonymously to build their happiness profile.

The third aspect that requires improvement is the sense of "fun" participants experience when using the app. Even though the drop-out rates were low, and approximately one-quarter of the participants reported that they could imagine using the app for a month or permanently, only one-quarter clearly stated that they fully enjoyed using the app. Although this may have been caused by the intensity of the app used in this study, the aspect of "fun" must be improved. The participants explained that they generally appreciated receiving a notification that asked how happy they were at a specific moment in time. The participants particularly enjoyed maintaining a daily diary. They said that they liked analysing their day in the form of a digital diary, as they occasionally did this with the help of social media such as Facebook or Twitter; however, they preferred to perform the diary activity privately and anonymously. Another "fun function" could be a "decision helper" that would, for example, help a user decide whether to go to the cinema with his best friend or with his girlfriend by analysing the happiness ratings of his past activities.

In sum, the app features all performed well in the evaluation, and we were able to produce detailed datasets according to the onion model; however, some future improvements are required to reduce user burden.

CONCLUSION AND FUTURE PROSPECTS

Using the Happiness Analyzer, it is possible to measure SWB according the OECD gold standard because the app does not require substantial resources (from researchers or participants) to capture the following: i) people's SWB at multiple points in time via general questionnaires; ii) data about people's everyday life and everyday feelings; iii) people's direct feelings in the moment; and iv) a combination of subjective and objective well-being measurements.

In the future, we will endeavour to reduce user burden, optimise the app's usability and enhance the tool's flexibility for adaptation to different studies. Additionally, we plan

to build more connections to objective data resources to link the SWB data with other objective well-being data to optimise the frame around the three subjective layers of the onion model. To accomplish this, we plan to develop and test face emotion recognition modules, voice emotion recognition modules, and a smartwatch application to connect the app to smartwatch sensors, as well as different connections to other devices including heart rate variability sensors and smart home devices, such as smart scales, and to other databases.

With the current tool and these future improvements, we hope to offer researchers the ability to conduct studies matching the OECD gold standard with less resources. By this we hope to support researchers to have better data to better understand SWB in order to increase the probability of developing more efficient interventions to increase people's well-being.

4 Evaluation Objectives

INTRODUCTION

This chapter aims to summarize the introduction (chapter 1), review (chapter 2) and method (chapter 3) part and introduce the evaluation objectives of chapters 5, 6, 7 and 8. The different research questions and study plans are outlined.

SUMMARY OF CHAPTER 1, 2 AND 3

In chapter 1 we outlined the importance of collecting better data on subjective well-being with three examples and gave an overview of the goal and structure of this dissertation.

In chapter 2 we reviewed different subjective well-being assessment methods, discussed for which scenarios which assessment method is most efficient and concluded that modern technologies could help to collect better data on people's subjective well-being with manageable efforts.

In chapter 3 we outlined our approach, the development process and features of the Happiness Analyzer App and summarized the user feedback in our most recent evaluation study.

In conclusion, we outlined the problem of too few studies measuring subjective well-being according to the OECD gold standard and our solution using modern technologies.

EVALUATION OBJECTIVES

After we developed the Happiness Analyzer, we ran a first study and started to discuss the tool with different well-being researchers in various personal meetings, academic seminars and conferences such as the conferences of the International Society for Quality of Life Studies. Based on these discussions, we realized that it is not enough to simply do an evaluation study to investigate the users' experience and technical challenges.

Therefore, we started to collect topics out of these discussions, summarized them and shared them afterwards with those scholars to then start to investigate them in detail.

The first main question that came up was if the tool can collect better data to better understand subjective well-being. In order to answer this question, we collected data on domestic migrants' and locals' subjective well-being using the Happiness Analyzer and investigated if the tool helps us to better understand the differences between domestic migrants and locals and can explain more of the variance. This study is outlined in chapter 5.

Another question that came up was if participating in a Happiness Analyzer data collection has beneficial effects for the participant or in other words if the tool has a beneficial effect. In order to answer this question, we collected data in two studies in 2013 and 2016 on young German adults' subjective well-being in two different ways: The one group simply answered subjective well-being questionnaires three times in four weeks and the other group additionally used the Happiness Analyzer for two weeks after filling in the first subjective well-being questionnaire. This study is outlined in chapter 6.

A more technical question that came up was if it matters if the Day Reconstruction Method is used in the evening of the same day or on the next day. This question became relevant because most users in our previous Happiness Analyzer studies reported that they would prefer to do a diary on the evening of the same day instead of on the next day. Thus, we collected data using the Happiness Analyzer with free choice of filling in the diary on the evening of the same day or on the next day on a representative German sample collected in cooperation with the German Socio-Economic Panel. This data was analyzed comparing differences in the subjective well-being ratings in the diary episodes to answer the research question. This study is outlined in chapter 7.

The key question that came up after some successful pilot studies was if it is possible to collect a representative dataset using the Happiness Analyzer App. In order to answer this question, we collected data in cooperation with the German Socio-Economic Panel in 2015 and 2016. After a household interview, participants saw a video about the Happiness Analyzer and were asked if they want to use it for a week. In 2015 the participants could not win anything by participating. In 2016 we offered a 50 Euro Amazon-Voucher for the participants if they used the Happiness Analyzer successfully for one week. In

2017 we analyzed which demographic groups participated to investigate the research question. This study is outlined in chapter 8.

In conclusion, we took the main questions that we collected from discussing the tool with different well-being scholars, designed research studies to investigate these questions and answered them in four different research papers. Of course, there are many other questions that are worth being investigated but are not part of this dissertation (some examples are discussed in chapter 9). Figure 4.1 summarizes chapter 4.

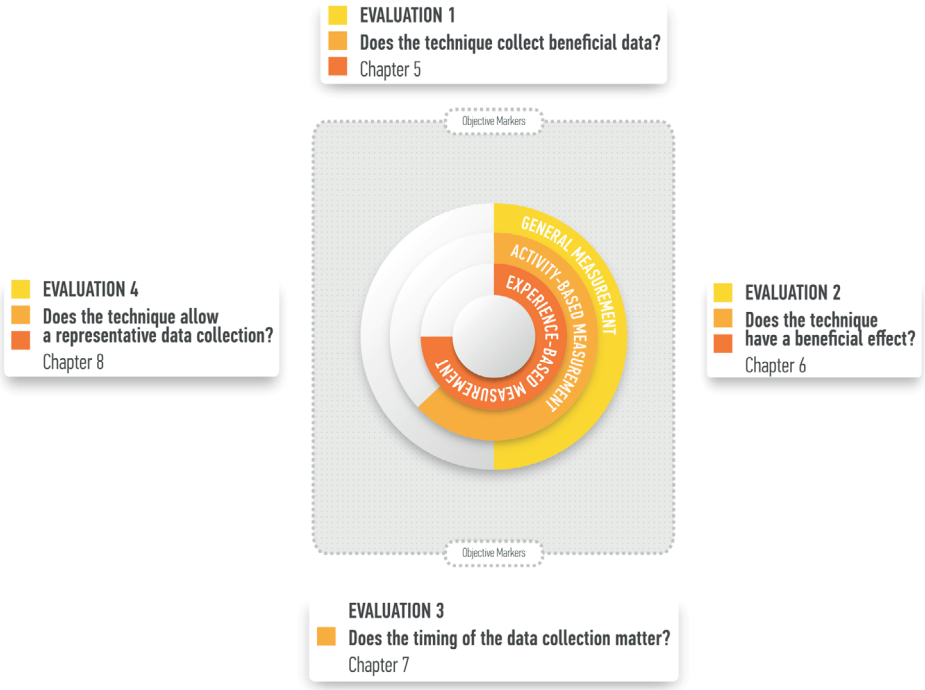

Figure 4.1: Displayed is the onion model and the key questions of the four evaluation studies. The color code next to the evaluation boxes displays which data sources respectively layers of the onion model the evaluation study looks at. Evaluation study 1, 2 and 4 investigate data sources from all subjective layers in a triangulation approach. Evaluation study 3 only focuses on the data collected with the Day Reconstruction Method respectively activity-based measurements. No study investigates the relationship between the subjective and objective markers.

5 Why are Locals Happier than Internal Migrants? The Role of Daily Life

ABSTRACT

Several survey studies have found that internal migrants report lower levels of happiness than locals, even after accounting for socio-economic factors. Traditional global self-ratings reveal that the migrant–local happiness-gap is also present in the data we present. The reasons for the migrant–local happiness-gap are as yet unclear. This paper aims to open this 'black box' by exploring the role of daily activities among a population that has generally been overlooked despite their high migration frequency: young adults. An innovative smartphone application is used that combines two techniques for multiple moment assessment: the experience sampling method and the day reconstruction method. Based on the application data, we examine whether internal migrants spend their time differently than locals and in which situations they feel noticeably less happy than locals. The data reveal that internal migrants distribute less time to happiness-producing activities such as active leisure, social drinking/parties, and activities outside home/work/transit. Internal migrants feel less happy than locals when spending time with friends and while eating. Possible explanations focusing on the role of social capital are discussed. Further analyses reveal that daily life experiences greatly enhance the explanation of the migrant–local happiness-gap. This paper demonstrates the potential value of real-time data and phone applications in solving happiness puzzles.

Keywords: Internal migration, Residential mobility, Happiness, Subjective well-being, Experience sampling method, Day reconstruction method

INTRODUCTION

Research from multiple countries has found that internal migrants are unable to reach similar levels of happiness as locals (Aksel et al. 2007; Cheng et al. 2014). Some studies even find migrants to be less happy than people in the place of settlement (locals) and people in the place of origin (stayers), which holds even after controlling for a range of socio-economic factors (Appleton and Song 2008; Knight and Gunatilaka 2010). A specific, well-researched case of internal migration is that of German internal migrants after reunification. Those who moved post-reunification from the former East Germany to the former West Germany have become happier after migration but have not reached similar happiness as former West-Germans over time, whereas West-to-East migrants have become unhappier but remain having higher happiness than East-Germans over time when accounting for socio-demographic factors (Frijters et al. 2004; Fuchs-Schündeln and Schündeln 2009; Melzer 2011). These findings raise three questions: why does the migrant–local happiness-gap occur? Does the lower general happiness reported by migrants in surveys also translate in less happiness in daily life? What can be done to reduce the gap? This paper makes a two-fold contribution. The first contribution is theoretical: the current paper progresses current knowledge on internal migrants by opening the 'black box' on factors causing the migrant–local happiness-gap. This is achieved by evaluating the migrant–local happiness-gap from a new perspective: the role of daily life, which allows us to address the three questions that were raised in the first paragraph. The second contribution is methodological: a pioneering smartphone application is introduced that cost-effectively incorporates two leading research methods to zoom in on daily life: the experience sampling method (ESM) and the day reconstruction method (DRM). Although the application as developed is focused on happiness, the concept of the application is applicable and valuable across many research fields. This paper specifically focuses on internal migrants to intensify inquiry among scholars and policy makers into this population. Internal migration has remained an underdeveloped theme, especially when compared to international migration. Yet, migrating within a country greatly exceeds the number of international migrants and is a crucial life event, as it largely disrupts and destabilises the pattern of daily life (Molloy et al. 2011). Additionally, this paper focuses on a population that has been overlooked in the literature on the happiness of internal migrants: young adults. Ironically, young adults are those who migrate most frequently as they face major life changes such as starting a job, study, family, or moving in with a partner. They are also especially sensitive to life-disruptions because compared to their older peers, they have less self-esteem and life experience, and they are more dependent, less emotionally stable, and more vulnerable to peer pressure (Rosenberg 1965). Hence, this population in particular needs support in making the most of migrating.

What can multiple moment assessments add to general surveys? Global self-reports, based on aggregated data retrieved from memory, are not always exhaustive in explaining sociological, economic, or psychological puzzles (Csikszentmihalyi and Hunter 2003; Kahneman et al. 2004). Hektner, Schmidt, and Csikszentmihalyi suggest that zooming in on people's daily lives can open black boxes that remain closed when using general surveys. Zooming in on daily life provides insight into what one does and how one feels throughout the day. Examples of studies benefiting from such data are Knabe et al. (2010), who use it to understand the happiness gap between the employed and the unemployed, and Kahneman et al. (2006), who use it to clarify the weak relationship between income and happiness. Current migration literature has merely used ratings of overall happiness derived from general questionnaires. There is reason to believe that examining daily life in greater detail contributes to explaining the migrant–local happiness-gap. Upon moving, migrants need to build a new social network, create new life patterns, and engage in new activities, organisations, and teams. This substantially impacts daily life in two fashions. First, it can result in a different distribution of time as one's daily routine is disrupted and certain activities are voluntarily or involuntarily initiated or discontinued (e.g., one's job, volunteering, sports). Second, it can lead to less enjoyment of particular daily life activities; for instance, going out with a person one has recently met may be less enjoyable than going out with a long-time friend. Why incorporate both the ESM and the DRM? Both methods capture a representative sample of individuals' actions, thoughts and feelings throughout the day, across contexts, close to their actual occurrence. In the ESM, respondents are asked to report their present feelings and actions at short notice after receiving each of several signals distributed throughout the day. The DRM asks respondents to complete diaries of the previous day in which the feelings experienced during each performed activity are reported. The unique methodologies imply unique strengths and weaknesses, making co-existence desirable (an extensive discussion is presented in Sect. 3). Despite the value of combining these methods, researchers typically choose between these two methods to reduce costs and to increase people's willingness to participate (Diener and Tay 2013). Therefore, a pioneering smartphone application is introduced that cost-effectively combines these methods.

Young adults in the Düsseldorf area (Germany) used the application for 2 weeks. The multiple moment assessments obtained four sources of information for every period of the day:

1. How the person is feeling
2. What the person is doing
3. Where the person is
4. Who is with the person

The multiple moment assessments began after completing a baseline questionnaire. The baseline questionnaire includes six global measures of subjective well-being, of which four specifically measure global happiness. For all six measures, migrants report lower happiness/subjective well-being than locals. Moving to the data obtained from DRM and ESM, two steps are taken to detect whether migrants and locals also have dissimilar daily life experiences. First, we examine whether migrants distribute less time to happiness-promoting activities, places, and people. Second, we examine whether migrants feel unhappier in certain places, during certain activities, and with certain people. In a subsequent analysis, daily life experiences are moved to the position of explanatory factors to discover the degree to which daily life experiences explain the migrant–local happiness gap present in global self-ratings. Data on daily life experiences obtained by the smartphone application are shown to have substantial explanatory power, indicating that the two sought-for research contributions are achieved.

The remainder of the paper is organized as follows. Section 2 reviews the literature that relates internal migration to happiness. Section 3 discusses the value and co-existence of the DRM and the ESM. Section 4 introduces the smartphone application used in this paper. Section 5 describes the experimental set-up, and Section 6 presents the results. The final section concludes, discusses, and provides policy implications.

INTERNAL MIGRATION AND HAPPINESS

Internal migration is usually defined as the geographical relocation of people over a substantial distance within national borders. The impact of migration on people's lives goes beyond mere 'geographically relocation'. One must adapt to another physical, social, and cultural environment, develop new life patterns, and build new social contacts. It leads to challenges such as finding new friends, communities, hobbies, and sports (teams), implying that stress typically accompanies migration. It offers migrants great opportunities but also substantial risks. The migration literature shows that migrants may not get the best out of migration. Studies in Australia and Turkey have found that 30-50 percent of migrants regret their decision to migrate, implying that it is difficult to overcome the hardships that come with migration (Stimson and Minnery 1998; Aksel et al. 2007; Fozdar and Torezani 2008). Although our study does not focus on pre-migration versus postmigration happiness development, it is relevant to review studies testing whether migrants gained in happiness by migrating. Recent studies in Finland (Ek et al. 2008), Australia (Kettlewell 2010), and the UK (Nowok et al. 2013) conclude that most people fail to obtain greater happiness by internal migration. These findings imply that migrants do not commonly reach similar levels of happiness as locals, as they do not gain in happiness at all. The gap is shown to be especially relevant for recent migrants,

although it is still present among long-term migrants (Knight and Gunatilaka 2010; Cheng et al. 2014).

The key question is why migrants cannot get the best out of migration and how this failure translates into lower levels of happiness in daily life. The social capital literature shows that most migrants face difficulty connecting to others, which makes building new networks an underestimated challenge (Portes 2000; Putnam 2000). Hardships arise because: (1) migrants have little initial social capital via which potential friends can be met; (2) an internal migrant often becomes a member of the out-group because they are seen as 'different', for instance, due to a different accent; and (3) migrants face impaired self-esteem as they are still adapting to the new environment and feel 'pressured' to build new social capital. In daily life, these hardships can translate into feelings of inferiority when being around others, mostly because of being less in one's 'comfort zone', having friendships of lower quality, and lacking social support, and/or spending less time with good friends (e.g., one is less invited to social events). The broaden-and-build theory (Fredrickson 2001) suggests that migrants will spend less time on physically and mentally effortful activities. Namely, internal migrants typically give in on mental health due to the stress that comes with migration (Chen 2011); experiencing negative feelings impairs openness to experiences and reduces activeness because one has less mental energy to distribute. Lower mental health results in less time allocated to effortful events as active leisure, meeting people, and out-of-home activities. Moreover, migrants may not have access to their preferred activities in the place of settlement, forcing them to discontinue certain activities. Relatedly, migrants may lack the motivation to rebuild in the new place what they had built up in the old place over the years (e.g., one's reputation on a sports team).

Taken together, it suggests that migrants are less engaged in effortful leisure activities. This can reduce happiness, as activity theory suggests that active, social, and challenging activities typically result in higher happiness (Rodriguez et al. 2008). As a by-product of less social capital, migrants may enjoy social or effortful leisure activities less as they cannot always perform these activities with their long-time friends but have to settle with less close friends. Relatedly, they occasionally need to resort to less desired activities as they have no friends, team, or community with whom they can perform the desired activity (Putnam 2000). The time composition of everyday activities as eating, sleeping, and housekeeping is less of a choice, which implies that time distributed to these activities is not so much affected by migrating. Activities where no social and energy-restrictions are involved are also hypothesised to be less affected; examples include in-home activities and individual activities such as reading. In conclusion, it is hypothesised that lower happiness of migrants is particularly caused by less enjoyment of and time composition for social and effortful activities, but to a lesser degree by individual and daily activities.

THE VALUE AND CO-EXISTENCE OF MULTIPLE-MOMENT ASSESSMENTS

Global self-reports used in prior literature have only allowed the examination of the happiness-gap from a 'helicopter-view', whereas tracking daily experiences provides the opportunity for a more detailed look at the migrant–local happiness-gap. The advantages are more diverse. Primarily, global self-reports (e.g., satisfaction with life-as-a-whole) rely on distorted information derived from memory. People face hardships in retrieving all the relevant data from their memory and to subsequently aggregate these data because one's answer depends strongly on recent and vivid experiences (Diener et al. 2013). The use of memory and the need for aggregation are minimised for assessments on a momentary (ESM) or daily (DRM) basis; it makes these methods less vulnerable to recall- and aggregation bias, implying a better representation of actual experience. Second, episodic assessments complement global self-reports because the latter typically encloses a strong cognitive component of happiness, whereas the former encloses more 'pure' affective evaluations. An advantage of the ESM is that it promotes ecological validity, as participants provide self-reports in the environment where they truly experience the feelings. A final contribution over traditional cross-sectional studies is that the panel element allows for more robust individual assessments of happiness as biases such as the mood bias are cancelled out on the individual level. The principal limitation of the ESM and DRM are their high costs. A second limitation is that assessments of momentary happiness (i.e., mood) are aggregated to an assessment of long-term happiness, whereas in actuality, the predictors are not exactly the same. A third limitation is that overall scores are biased as participants selectively present information, e.g., participants are reluctant to report that they have been making love (Diener and Tay 2013). Accordingly, global self-reports and episodic assessments are supplementary in assessing happiness.

The key advantages of the DRM over the ESM are: (1) the obtained data cover the complete day, and (2) completing the diaries does not interrupt the flow of daily life, making the DRM less burdensome than the ESM (Diener and Tay 2013). The key advantages of the ESM over the DRM are: (1) superior ecological validity, as participants provide self-reports in the environment and at the moment of truly experiencing the feelings, and (2) it does not face the issue of the DRM that some participants divide the day into more episodes than others, resulting in potential divergent responses between two groups. Their co-existence also gives the opportunity to test for convergent validity of daily life assessments. This is valuable because the psychometric support for the ESM and especially the DRM is limited at this point in time. In conclusion, clear grounds for a multimethod design are present.

THE SMARTPHONE APPLICATION

In studies applying the DRM and ESM, experimenters typically handed participants a PDA that had to be used during the experimental period. The use of smartphones has several benefits. First, smartphones promote further ecological validity as participants complete the assessments on a mobile phone to which they are accustomed and which they do not only use for experimental purposes, resulting in lower awareness of 'being followed'. Second, the burden on participants is lower because they do not have to carry around an extra device. This can raise the rates of response and participation. Third, it reduces costs, as no digital devices need to be provided to participants, which implies that there are no costs for purchasing, repairing, and retrieving the equipment. Finally, applications allow for the establishment of 'fuzzy' samples consisting of people who voluntarily and regularly indicate their feelings. With this information, the impact of shocks can be revealed (e.g., the impact of a natural disaster or one's national team winning a soccer match) and data on momentary assessments can be easily collected cross-nationally. The use of smartphone applications has become a more attractive option as many people now own smartphones. We created a cost- and time-effective happiness application that allows participants to complete both daily happiness diaries and momentary assessments on their personal mobile phones. This is not the first smartphone application developed to examine episodic happiness. Killingsworth and Gilbert (2010) successfully used a happiness application for experience sampling, and MacKerron (2011) developed a happiness app to study the influence of the local environment (e.g., the weather) on episodic happiness. The application we introduce has two main advantages over these applications. First, the DRM is included on top of the ESM. Second, this application is downloadable for Android users, whereas earlier apps focused on iPhone users. Using Android decreases the self-selection problem because approximately 65 percent of smartphone users have Android, whereas iPhone users represent only 25 percent of smartphone users (Kantar World Panel 2014).

METHODS

Sample and Procedure

In the recruitment procedure, individuals were informed that the study involves a tracking of happiness and time-use and that the study incorporates questions about demographics. Individuals were additionally informed of the confidentiality of answers, the possibility to skip a question in any case, and the possibility to opt out of the study at any moment. Interested individuals were asked to list their e-mail address. In exchange for course credit, 123 young adults (17–30 years old) studying psychology in the Düs-

seldorf area were recruited for participation. Concurrently, 75 young adults living in the Düsseldorf area were recruited via word-of-mouth communication to increase the heterogeneity of the sample (e.g., to include non-students). To prevent dropout, a lottery was announced in which those who completed at least two-third of the diaries and experience samples could win a 250 euro Amazon.com voucher.

The multi-method study consists of three parts: (1) an online questionnaire (2) momentary assessments, and (3) day reconstructions. A brief questionnaire, including a range of measures of happiness and demographic questions, was purposively designed to obtain a baseline-estimate of a person's happiness. All participants completed the baseline questionnaire on the same evening and were afterwards informed how to download the application onto their mobile phones. After downloading the application, a tutorial with instructions on how to use the application was provided to raise conscientiousness and boost response accuracy. The episodic assessments began for all participants on November, 7, 2013 (Thursday) and lasted for two weeks until November, 21, 2013 (Thursday). Each day, six signals were distributed throughout daytime covering the entire waking day; two consecutive signals were always more than an hour apart. When signaled, participants were asked to indicate (1) how they were feeling, (2) what they were doing, (3) where they were, and (4) who was with them. The activities they could choose from were equal to those of the HappinessIndicator (Oerlemans et al. 2014) and in line with categorisations made by Kahneman et al. (2004). Every morning, participants were asked to complete a diary of the previous day in which they answer the same four questions.

Participants are included in the analysis when the signal response rate was over 66 percent in both the Experience Sampling Data (56 out of 84) and the daily diaries (10 out of 15). These minima have been successfully surpassed by 109 locals and 41 migrants, indicating an attrition rate of 24 percent. These 150 participants yielded a total of 11,455 momentary assessments and 1,918 day reconstructions. The participants' response rates in the ESM and DRM are, respectively, 91 and 85 % and there were no noteworthy differences in compliance between migrants and locals. One to two minutes for completing an experience sample is supported in literature as being desirable to make the interruptions brief and less intrusive (Hektner et al. 2007). In our study, this was considerable less with an average of about 30 seconds. The time to complete a diary was typically in the range of 5–10 min.

Variables

Dependent Variables

To facilitate robust assessments about happiness and subjective well-being, the OECD published guidelines on measuring subjective well-being (OECD 2013) in which they

propose including measures for (1) overall happiness (2) sub-elements of happiness (life satisfaction and hedonic affect) and (3) elements of broader subjective well-being (eudaimonic well-being and domain evaluations). The baseline questionnaire follows the proposed structure by including one overall happiness measure, three measures focusing on sub-elements of happiness, and two measures focusing on broader subjective well-being, although it should be noted that we are mainly interested in the four measures focusing on happiness. Overall happiness is our primary interest and assessed by the question: 'Taking all things together, how happy would you say you are?'. It is answered on a numerical 11-point scale ranging from extremely unhappy to extremely happy. Life satisfaction embodies the cognitive side of happiness. The life satisfaction measure is 'All things considered, how satisfied are you with your life as a whole nowadays?' ranging from 0 (extremely dissatisfied) to 10 (extremely satisfied). Additionally, a frequently used multiple-item scale to measure life satisfaction is included: Diener's Satisfaction With Life Scale (SWLS; Diener et al. 1985). The original 10-item Affect Balance Scale is included to assess the affective side of happiness (Bradburn 1969). The first measure of broader subjective well-being is psychological functioning (also referred to as eudaimonia or flourishing) and is measured by the Flourishing Scale (Diener et al. 2010). This scale incorporates eight items (e.g., I lead a purposeful and meaningful life) and is rated on a seven-point scale (1 = strongly disagree; 7 = strongly agree). Next, participants evaluated satisfaction with ten life domains (e.g., satisfaction with financial status or satisfaction with health status; for the other items, see OECD 2013) on a 10-point scale (0 = not at all satisfied; 10 = completely satisfied). As recommended by the OECD, every participant started the baseline questionnaire by rating one's overall happiness and finished with rating domain satisfactions to minimize spill-over effects in the measures we are most interested in.

Prior studies applying the DRM or ESM have either used an affect balance scale (multiitem) or a single question on current happiness to assess affective feelings in daily life. We opted for the latter option, operationalized by the questions 'How do you feel?' in the ESM and 'How did you feel' in the DRM and answered on an 11-item numerical scale ranging from 'unhappy' to 'happy', for three reasons. First, it is less of a burden on participants. Second, the single-item measure is strongly correlated with the multi-item measure, implying that the results are largely robust to the measure used (Knabe et al. 2010). Third, there is an ongoing debate about the validity of aggregating specific emotions. When aggregating, there is no agreement on how specific feelings should be weighed against each other (White and Dolan 2009). For instance, joy may have a stronger effect than relaxation. By rating the 1-item question, the respondent himself weights which emotions and thoughts are most important for his overall feelings during an activity. Additionally, positive and negative affect do not lie on a single dimension in the affective system, which makes it hazardous to aggregate them (Cacioppo and Bernston 1994).

Independent Variables

An important reason behind the scant research on internal migrants is that there is no clear boundary between an internal migrant and a local. Two recent studies have utilised a question asking about the spatial distance (measured in kilometres) one has moved (Kettlewell 2010; Nowok et al. 2013). This measure may not be optimal as the 'perceived' distance of migration may depend on factors such as one's transportation resources and cultural differences. Measures of cultural distance and travel-time from the place of origin encompass elements that are related to happiness (cultural differences) and other confounding elements (transportation opportunities often serve as a proxy for income), leading to an inherent bias. Therefore, we preferred to use an objective measure: spatial distance. The exact question we used is: 'What is the approximate distance, from the city or town where you lived in before you moved, to the city or town in which you are currently living?'. The five answer-options were 'More than a 1,000 km', '100–1,000 km', '25–100 km', '\25 km' and 'I always lived here'. Due to the limited sample size and to maintain a parsimonious model, we chose to merge these categories into a single 'migrant'-category and a single 'locals'-category. Nowok et al. find that those who moved over more than 100 km (long-distance movers) have lower happiness than those who move smaller distances (25–100 km). Kettlewell defined migrants who moved more than 20 km as migrants. We wished to distinguish between those who truly experience the need to build new social lives from those whose daily lives are less affected. Therefore, we categorized those who migrated over 100 kilometers as migrants and other participants as non-migrants.

Control Variables

To minimize self-selection bias, the following factors that are typically associated with happiness are included as covariates in the empirical analysis: age, gender, having a partner, immigrant status (including 1st and 2nd generation immigrants are classified as immigrants), household income, employment status, and whether one has a chronic condition as a proxy for health. Additionally, personality is incorporated as measured by the 'TIPI-scale' (Gosling et al. 2003). More specific controls for the purposes of our study are also included. These are household situation (as it is strongly correlated to migration), being a student (as a substantial portion of the participants are students), and being born in East Germany (those born in East Germany are typically unhappier).

Analytical Strategy

The ESM-data do not only account for momentary happiness but also yield an assessment of a participant's general happiness level by taking the mean of the separate momentary responses (Kahneman 1999). The same applies to the DRM, although one cannot simply aggregate all happiness-assessments as activities have dissimilar durations. We follow

previous studies in applying the duration-weighted method of aggregation (e.g., White and Dolan 2009), having the following formula:

$$Total\ Daily\ Happiness\ By\ DRM_i = \sum_j (Episode\ Duration_{ij} * Feeling\ During\ Episode_{ij})$$

in which *i* represents an individual and *j* an episode. This formula is not optimal as it is unlikely that people assign a similar weight to each episode; however, because there is little knowledge concerning the importance of particular episodes in daily life, this method is acknowledged to be best practice. Happiness is commonly treated as being cardinal in subjective well-being literature; we follow this approach as the results of cardinal models are more intuitive and easier to interpret than estimates from ordinal probit models. In addition, cardinal and ordinal analyses of life satisfaction yield, in general, similar results (Ferrer-i-Carbonell and Frijters 2004). The presence of the migrant–local happiness-gap is examined in two steps. First, the presence of differences in happiness between migrants and locals is checked for each of the six global self-reports and for the aggregated scores of the ESM and the DRM. Next, a MANCOVA is applied, which takes the positive correlation between the dependent variables into account. To avoid multicollinearity in the MANCOVA, the rule-of-thumb of Maxwell (2001) is followed, which indicates that dependent variables should be low to moderately correlated with a lower boundary of 0.30 and an upper-boundary of 0.70. In Table 5.1, the correlations in bold are very close to or exceed the upper-boundary, which implies that the MANCOVA is likely to suffer from multicollinearity when including all variables. Possible solutions to avoid multicollinearity are (1) the formation of a composite variable, or (2) dropping variables that cause the multicollinearity problem. We opted for the latter option because there is no clear basis for aggregating the measures into an overall measure of subjective well-being, making it hazardous to form a composite variable (OECD 2013). Additionally, the central focus of this paper is the happiness-gap. Therefore, the multivariate effect is based on four elements of happiness that are typically distinguished: overall happiness (as measured by the overall happiness measure), daily happiness (by the aggregated score of the day reconstructions), the cognitive element of happiness (by the SWLS), and the affective element of happiness (by the affect balance scale).

To explore the role of daily life, we will proceed by applying two perspectives to zoom in on the data obtained by experience sampling and happiness-diaries. A trait-like perspective is applied to detect possible time-composition effects. The following question is answered from this perspective: do migrants and locals allocate a different amount of time to certain activities, people, and places? Next, a state-like perspective is applied to answer the question: do particular activities, interaction partners, and locations lead to different happiness for migrants and locals? Knabe et al. (2010) refer to differences in happiness for the same type of activity as the 'saddening' effect. To enable more robust, meaningful, and interpretable results, the twenty-one specific activities from which respondents could select when completing their diaries and experience samples (e.g., exercising and eating) are aggregated into five overarching categories derived from the time-and-leisure composition literature (Dardis et al. 1994). These categories are social leisure/entertainment, active leisure, passive leisure, daily activities, and work-related activities. Next, we aim to show the incremental value of daily life experiences in explaining the migrant–local happiness-gap. Three OLS-regression models are developed for each of the six global self-ratings, which are the explained variables. Model 1 includes migration as an independent variable and also controls for individual and household characteristics to minimize selection-bias. Model 2 adds the time-composition differences as independent variables to examine the effect of different time-composition on the migrant–local happiness-gap. Model 3 presents a full model by including activities in which a clear saddening effect is present as independent variables.

RESULTS

Descriptive Statistics

Migration-specific questions in the baseline-questionnaire reveal that 63 % of the migrants in the sample moved 3 years ago, half of the migrants moved without knowing any friends or family in the host place, 58 % plan to live 3 more years in the study area, and 67 % moved for study purposes. More general sample descriptors are listed in Table 2. We will examine two models to analyze the happiness-gap, one including and one excluding the household situation. The reason for this division is that the household situation may have a potential mediating effect on the happiness outcome of migration because a different household situation is a logical consequence of migrating. The correlation matrix (Table 5.1) shows the convergent validity among the data derived from the DRM and the ESM, as the correlation between these methods is significant, sizeable, and larger in magnitude than their correlations with global happiness measures.

Table 5.1: Correlation-matrix of subjective well-being measures.

	Pearson Correlation coefficient (r)							
	ESM	DRM	Overall happiness	1-item life satisfaction	Affect Balance	SWLS	Eudaimonia	Domains
ESM	x							
DRM	**0.80**	x						
Overall happiness	0.45	0.34	x					
1-item life satisfaction	0.42	0.34	0.64	x				
Affect Balance	0.46	0.33	0.48	0.39	x			
SWLS	0.46	0.38	0.53	**0.68**	0.40	x		
Eudaimonia	0.42	0.36	0.44	0.49	0.43	**0.67**	x	
Domains	0.45	0.41	0.53	0.61	0.44	**0.74**	**0.69**	x

Note: all correlations are significant at the 1% level.

The Migrant–Local Happiness-Gap

Model A The degree to which the gap is present for each dependent variable when not controlling for household situation, but after controlling for all other covariates included in Table 5.2, is shown in model A of Table 5.3. Univariate results show that locals consistently reported significantly higher scores (at the 10 %-level) over the six measures. The gaps are somewhat larger for the four global self-ratings of happiness than for eudaimonic wellbeing and domain satisfactions. The practical importance of the findings is at least as important as the p-values in our study because we combine a limited sample size with a wide range of covariates to minimize selection-bias. The partial eta squares indicate that the effect sizes are modest, which can be considered as substantial in the context of happiness. Episodic assessments are typically found to be more rigid (e.g., Knabe et al. 2010), which is also the case in our data. A marginally significant gap is observed for the DRM-data, whereas the ESM-data do not display a statistically significant difference. Using the Bonferroni procedure, the MANCOVA reveals a significant difference between migrants and locals, and a mediocre effect of migration, on the four combined happiness variables ($F(4, 130) = 2.49$; $p = 0.046$; Wilks' Lambda = 0.929; gq2 = 0.071).

Model B The household situation is additionally controlled for in model B of Table 5.3. The results of model B are highly similar to those of model A. Non-reported results reveal that the low impact of the household situation can be explained by the weak and nonsignificant relationship between household situation and happiness. The practical and statistical significance of the combined variables is also largely similar to model A ($F(4, 127) = 2.31$ $p = 0.061$; Wilks' Lambda = 0.932; gq2 = 0.068). The results suggest that the happiness-gap is not completely driven by differences in demographics and income and that over 6 % of the multivariate variance of the dependent variables is associated with the migrant-dummy. The episodic assessments in model B do not reveal significant differences between migrants and non-migrants. Nonetheless, substantial happiness-gaps in the aggregated DRM-data and ESM-data remain present, which make it interesting to analyze the DRM-data and ESM-data in more detail to reveal which specific daily experiences cause these gaps.

Table 5.2: Descriptive characteristics of the sample

	Full sample 150	**Migrants** 41	**Locals** 109	
Variable	Mean (*SD*) / %	Mean (*SD*) / %	Mean (*SD*) / %	F statistic
Age in years	21.7 (3.10)	22.7 (3.19)	21.4 (2.99)	**5.99***
Gender (% male)	18%	10%	21%	
Has a partner	50%	54%	49%	
(Psychology) Students	73%	71%	73%	
Personality				
Extraversion	4.08 (0.66)	4.13 (0.52)	4.06 (0.71)	
Conscientiousness	3.99 (0.63)	4.00 (0.58)	3.98 (0.65)	
Openness	4.18 (0.78)	4.18 (0.73)	4.18 (0.80)	
Agreeableness	4.52 (0.80)	4.62 (0.80)	4.48 (0.80)	
Emotional stability	3.83 (0.73)	3.82 (0.78)	3.83 (0.72)	
Chronic condition	21%	27%	19%	
Immigrant	14%	17%	13%	
Born in East-Germany	4%	10%	2%	**4.96***
Monthly Income				**6.75***
Below modal	69%	88%	62%	
Modal (€2.500 net)	19%	7%	23%	
Above modal	11%	5%	13%	
Having a job (% yes)	44%	44%	44%	
Household situation				**6.75***
At parents' home	41%	22%	48%	
Alone	21%	29%	17%	
With partner	19%	17%	20%	
Flat-sharing with others	19%	32%	15%	

Note: Significance levels at the 10%, 5%, and 1% level are denoted by $^{+}$, *, and ** respectively

Table 5.3: Univariate differences in happiness and subjective well-being.

Model A – *all covariates, except for household situation*		Locals *109*		Migrants *41*		Mean difference	ANCOVA test *(df = 134)*	Effect size
		Adj. Mean	SE	Adj. Mean	SE	Locals-Migrants	*F* -statistic	$\eta\rho^2$
Overall happiness	(0-10)	6.56	0.15	5.96	0.25	*0. 60*	**4.09***	.030
1-item life satisfaction	(0-10)	6.60	0.19	5.97	0.32	*0.63*	**2.75** $^{+}$	.020
Affect Balance	(-5; 5)	+0.91	0.16	+0.02	0.28	*0.89*	**7.22***	.051
5-item SWLS	(1-7)	5.05	0.09	4.64	0.16	*0.41*	**4.74 ***	.034
Eudaimonic well-being	(1-7)	5.43	0.08	5.12	0.13	*0.31*	**3.78** $^{+}$	.027
Domain satisfactions	(0-10)	6.77	0.13	6.31	0.21	*0.46*	**3.29** $^{+}$	.024
Day reconstruction	(0-10)	6.67	0.08	6.37	0.13	*0.30*	**3.41**$^{+}$	.025
Experience sampling	(0-10)	6.53	0.08	6.32	0.14	*0.19*	**1.57**	.012
Model B - *All covariates included*		Locals *109*		Migrants *41*		Mean difference	ANCOVA test *(df = 131)*	Effect size
		Adj. Mean	SE	Adj. Mean	SE	Locals-Migrants	*F* -statistic	$\eta\rho^2$
Overall happiness	(0-10)	6.57	0.15	5.93	0.26	*0.64*	**4.19***	.031
1-item life satisfaction	(0-10)	6.58	0.19	6.01	0.32	*0.57*	**2.19**	.016
Affect Balance	(-5; 5)	+0.91	0.17	+0.02	0.29	*0.93*	**6.82****	.049
5-item SWLS	(1-7)	5.06	0.09	4.63	0.16	*0.43*	**4.87***	.036
Eudaimonic well-being	(1-7)	5.43	0.08	5.12	0.14	*0.31*	**3.49+**	.026
Domain satisfactions	(0-10)	6.78	0.13	6.30	0.22	*0.48*	**3.28+**	.024
Day reconstruction	(0-10)	6.66	0.08	6.40	0.14	*0.26*	**2.42**	.018
Experience sampling	(0-10)	6.52	0.08	6.34	0.14	*0.18*	**1.15**	.009

Note: Bonferroni adjustment for multiple dependent variables. Significance levels at the 10%, 5%, and 1% level are denoted by $^{+}$, *, and ** respectively.

Time-Composition of Daily Activities

In this section, a trait-like perspective is applied to zoom in on the time-allocation of migrants and locals. All covariates are included, which provides a substantial degree of certainty that the time-composition differences caused by migrating are filtered. Using the DRM-method, Table 5.4 shows that locals spend significantly more time on active leisure activities. Interestingly, being engaged in active leisure is typically recognised as promoting happiness. When looking at specific activities, it emerges that locals spend significantly more time on social drinking/parties, exercising, and in places other than home/ work/transport. In contrast, migrants spend more time on the computer, possibly to communicate with those not living nearby. We rule out the possibility that socio-demographic differences (e.g. differences in financial resources) account for these differences by including socio-demographic controls when examining the time-distribution. Unreported OLS regressions, controlling for the full set of covariates and the migration-dummy, show that greater time-allocation (in hours) to exercising, social drinking/ parties, and places other than home/work/transport, are associated with higher scores on the overall happiness measure in our sample (B's are, respectively, 0.28, 0.35, and 0.02). In contrast, spending time on the computer is associated with lower scores on happiness (B = -0.05). In sum, we can conclude that time composition seems to play a role in the happiness-gap.

Table 5.4: Happiness from a trait-like perspective (DRM-data)

	Time-allocation		
	Locals	Migrants	
	109	*41*	*ANCOVA*
	Mean time a day (Hours:minutes)		*F*-statistic
Social leisure/entertainment	**1:22**	**1:19**	
Social drinks and partying	0:19	0:05	7.66**
Visiting cinema, theatre, or sports	0:08	0:11	
Talking	0:34	0:40	
Shopping	0:07	0:09	
Institutional event (e.g. church-event)	0:14	0:14	
Active Leisure	**0:55**	**0:32**	4.70*
Intimacy/sex	0:08	0:07	
Exercising	0:24	0:12	3.68^{+}
Hobbies	0:23	0:13	
Passive Leisure	**2:50**	**2:48**	
Reading	0:08	0:07	
Watching TV	1:28	1:27	
Listening to music	0:05	0:08	
On the Computer	0:18	0:39	4.18*
Resting	0:51	0:27	
Other leisure	*0:48*	*0:38*	
Daily activities	**4:09**	**4:30**	
Eating	1:09	1.22	
Taking care of others	0:12	0:20	
Cooking	0:13	0:09	
In transit	1:40	1:42	
Getting up and ready	0:55	1:06	
Work-related activities	**5:57**	**6:14**	
Housekeeping	0:37	0:40	
Working	1:09	1:07	
Studying	4:11	4:27	
Total	**16:16**	**16:09**	
Interaction partners			
Partner	2:09	1:51	
Friends	3:43	4:00	
Direct family	1:40	1:56	
Alone	7:16	7:10	
Colleagues	1:21	1:21	
Total	**16:09**	**16:18**	
Locations			
Elsewhere out of house	6:20	5:20	3.35^{+}
At home	7:51	8:26	
At work	1:05	1:30	
In public transport/vehicle	0:55	1:03	
Total	**16:11**	**16:19**	

Note: Mean time a day corrected for differences in all covariates listed in table 1. Significance levels at the 10%, 5%, and 1% level are denoted by $^{+}$, *, and ** respectively.

Enjoyment of Daily Activities

Table 5.5 applies a state-like perspective to detect whether migrants and locals feel dissimilar happiness during certain episodes. It appears that migrants feel as good as locals during activities such as working and being with colleagues. In contrast, locals report significantly higher happiness scores in both the ESM and the DRM compared to migrants while eating and being with friends. Sizeable gaps are also present in other activities such as intimacy/sex, but no strong inferences can be drawn as the gaps are not significant due to the limited sample size. The happiness-gap of 'eating' is analyzed in greater detail in Table 5.6. A 'deeper' analysis of the eating category reveals that the happiness-gap especially occurs due to locals' greater enjoyment of eating with significant others, which is in line with the fact that migrants generally enjoy being with friends less. Interestingly, we observed in Table 4 that locals do not spend more time on eating and with friends. These results suggest that the saddening effect adds unique value in explaining the happiness-gap. On the contrary, time distribution to some other activities is in line with the happiness derived from those activities; that is, locals report both substantially higher happiness and time spent on activities outside home/work/transit and on exercising. Additionally, migrants spend more time on the computer and appear to enjoy that time more.

Table 5.5: Happiness from a state-like perspective in the ESM and the DRM

	DRM			ESM		
	Locals	Migrants		Locals	Migrants	
	109	*41*	*ANCOVA*	*109*	*41*	*ANCOVA*
	Current Happiness	Current Happiness	*F*-statistic	Current Happiness	Current Happiness	*F*-statistic
Social leisure/entertainment	**7.87**	**7.71**		**7.42**	**7.17**	
Social drinks and partying	8.32	7.96		7.65	7.80	
Going to cinema/theatre/sports	8.09	8.26		7.71	7.47	
Talking	7.79	7.60		7.03	7.15	
Shopping	7.59	7.42		7.72	7.05	
Institutional events (e.g. church event)	7.07	7.36		7.12	7.05	
Active Leisure	**7.74**	**7.48**		**7.20**	**7.26**	
Intimacy/sex	8.37	6.84		7.74	7.35	
Exercising	7.72	7.15	2.72^{+}	7.32	7.08	
Hobbies	7.24	7.57		6.97	6.90	
Passive leisure	**6.72**	**6.92**		**6.52**	**6.34**	
Reading	7.58	5.48	7.71*	6.81	6.52	
Watching TV	7.31	7.18		6.82	6.79	
Listening to music	7.00	7.31		6.99	6.50	
On the Computer	6.40	6.97	2.93^{+}	6.52	6.59	
Resting	5.97	6.41		5.95	5.84	
Other leisure	*7.63*	*7.27*		*7.15*	*7.22*	
Daily activities	**6.47**	**6.27**		**6.53**	**6.34**	
Eating	7.44	6.95	5.89*	7.04	6.72	3.15^{+}
Taking care of others	6.96	5.68		7.76	7.39	
Cooking	6.68	5.97		6.87	6.52	
In transit	6.29	5.94		6.40	6.33	
Getting up and ready	5.65	5.47		6.08	6.00	
Work-related activities	**6.18**	**6.01**		**6.10**	**5.94**	
Housekeeping	6.31	5.97		6.47	6.00	2.96^{+}
Working	6.25	6.26		6.14	6.29	
Studying	6.10	5.99		6.02	5.79	
Interaction partners						
Partner	7.45	7.16		6.90	7.10	
Friends	7.28	6.82	5.76*	7.08	6.70	4.43*
Direct family	7.19	6.77		6.86	6.47	3.22^{+}
Alone	6.16	6.06		6.11	6.00	
Colleagues	6.14	6.17		6.02	6.20	
Locations						
Elsewhere out of house	6.90	6.63		6.76	6.45	3.30^{+}
At home	6.54	6.32		6.39	6.28	
At work	6.21	6.11		6.07	6.00	
In public transport/vehicle	5.96	5.83		6.04	6.35	

Note: Results are mean-adjusted for differences in all covariates listed in table 1. The five overarching categories are weighted for frequency of performing the underlying specific activities. Significance levels at the 10%, 5%, and 1% level are denoted by $^{+}$, *, and ** respectively.

Table 5.6: Unraveling the happiness-difference of eating (DRM).

	Frequencies		Momentary happiness		
Interaction partners	*Locals*	*Migrants*	*Locals*	*Migrants*	*F-statistic*
Friends	12%	11%	7.73	7.23	2.86+
Partner	22%	16%	7.62	7.19	
Direct family	33%	41%	7.43	7.17	
Colleagues	1%	2%	-	-	
Alone	32%	30%	6.73	6.64	

Note: Percentages and means are adjusted for differences in all covariates listed in table 1. Significance levels at the 10%, 5%, and 1% level are denoted by +, *, and ** respectively.

The Incremental Value of Daily Life Experiences

The final model shows the incremental value of including time composition effects, and saddening effects that consistently and significantly differed between migrants and locals. Table 5.7 proceeds in stages. In the baseline model (model 1), the full set of covariates is added to minimize the chance that the more extensive models (model 2 and 3) pick up selection effects instead of adding unique explanatory power. In model 2, the time allocation of the five distinguished categories is incorporated and supplemented by the category 'other leisure', as it does not belong in a specific category. Finally, a "full" model is created (model 3) in which the consistently-found saddening effects of eating and being with friends are included. The value of model 2 and 3 is threefold. First, it allows us to examine the degree to which the inclusion of daily life experiences can reduce the coefficient of the migration dummy. Logically, the migration dummy in model 1 is largely similar to the migration gap reported in model B of Table 3 as the same covariates are included but a different statistical procedure is applied. When also including time allocation (model 2), the gap decreases for all dependent variables relative to the first model and becomes insignificant for four of the six variables. The gap decreases further when saddening effects are included (model 3), resulting in insignificant migration variables for all dependent variables. Although part of the gap remains present in model 3, the gap decreases substantially in comparison to the less comprehensive models. Hence, daily assessments are valuable contributions in explaining the happiness-gap.

Second, the decreased gap can be explained by examining the effect of the time spending categories and saddening effects on the dependent variables. Active leisure is positively related to happiness and partially explains the gap because locals spend more time on this happiness-producing activity. Feeling good while eating and being with friends is also associated with higher scores for the dependent variables. One's feelings with friends can be expected to have a stronger association with long-term happiness than one's feelings during eating as social networks are crucial for people. Interestingly, the reverse was consistently found over the dependent variables. Note, however, that the

Table 5.7: The incremental value of including daily life experiences.

	Overall happiness (0-10)	*1-item life satisfaction (0-10)*	*Affect Balance (-5; 5)*	*SWLS (1-7)*	*Eudaimonia (1-7)*	*Domains (0-10)*
	B	*B*	*B*	*B*	*B*	*B*
Model 1 – Controlling for personal characteristics (df=130)						
Internal migration	-0.64* (0.31)	-0.58^{+} (0.39)	-0.91** (0.34)	-0.40* (0.19)	-0.29^{+} (0.16)	-0.46^{+} (0.26)
R-squared	*0.201*	*0.218*	*0.248*	*0.268*	*0.101*	*0.205*
Adjusted R-squared	*0.084*	*0.104*	*0.138*	*0.161*	*-0.030*	*0.089*
Model 2 – Additionally controlling for time-allocation (df=124)						
Internal migration	-0.50 (0.32)	-0.38 (0.40)	-0.80* (0.35)	-0.26 (0.19)	-0.24* (0.17)	-0.31 (0.27)
Social leisure	0.10 (0.13)	0.14 (0.17)	0.11 (0.15)	0.07 (0.08)	0.12^{+} (0.07)	0.21^{+} (0.11)
Active leisure	0.37* (0.15)	0.53** (0.19)	0.23 (0.17)	0.25** (0.09)	0.08 (0.08)	0.23^{+} (0.13)
Passive leisure	-0.05 (0.07)	-0.08 (0.09)	-0.19 (0.08)	-0.08^{+} (0.05)	-0.11** (0.04)	-0.02 (0.06)
Other leisure	0.01 (0.14)	-0.01 (0.17)	-0.24 (0.15)	0.07 (0.08)	-0.02 (0.07)	0.04 (0.12)
Daily activities	0.06 (0.06)	0.11 (0.07)	-0.03 (0.07)	0.01 (0.04)	0.02 (0.03)	0.01 (0.05)
Work-related	-0.02 (0. 05)	-0.11^{+} (0.06)	-0.05 (0.05)	-0.04 (0.03)	-0.02 (0.03)	-0.09* (0.04)
R-squared	*0.250*	*0.295*	*0.289*	*0.336*	*0.183*	*0.268*
Adjusted R-squared	*0.099*	*0.153*	*0.146*	*0.203*	*0.019*	*0.129*
Model 3 – Additionally controlling for 'saddening' effects (df=122)						
Internal migration	-0.31 (0.32)	-0.18 (0.40)	-0.59 (0.36)	-0.19 (0.20)	-0.13 (0.17)	-0.11 (0.27)
Feeling during eating	0.39* (0.16)	0.48* (0.20)	0.29^{+} (0.17)	0.21* (0.10)	0.21** (0.08)	0.22^{+} (0.13)
Feeling with friends	0.10 (0.16)	0.02 (0.20)	0.25 (0.18)	0.00 (0.10)	0.07 (0.08)	0.31* (0.14)
R-squared	*0.297*	*0.333*	*0.327*	*0.363*	*0.246*	*0.332*
Adjusted R-squared	*0.141*	*0.186*	*0.178*	*0.222*	*0.079*	*0.185*
Observations	150	150	150	150	150	150

Note: OLS estimation. Standard errors in parentheses. Significance levels at the 10%, 5%, and 1% level are denoted by $^{+}$, *, and ** respectively

gap in eating is largely caused by enjoying eating with friends less. Finally, we examine whether each progressive model has incremental value in explaining general happiness and subjective well-being. For this purpose, adjusted R squared statistics are reported because they impose a penalty on additional parameters to a model. Models 2 and 3 have incremental value in explaining each dependent variable as the R^2 rises in any case from model 1 to model 2 and from model 2 to model 3. This implies that time composition and differences in momentary feelings are valuable explanatory factors, on top of individual and household characteristics, for explaining subjective well-being and happiness.

Robustness Check

The cross-sectional data cannot rule out the possibility that the migration effect is driven by migrants who are genetically unhappier than locals and therefore never obtain a similar happiness level. For multiple reasons, it is unlikely that the entire migration gap would be caused by genetics; German internal migrants after reunification appeared to be somewhat happier than non-migrants (Fuchs-Schündeln and Schündeln 2009), migrants are typically self-confident and optimistic individuals, which correlates with higher happiness (Knight and Gunatilaka 2010), and the gap we observe remains sizeable after controlling for personality. Still, a more formal test of possible selectivity would increase the leverage of the results. Therefore, we follow Bartram (2013) in applying a 2-stage treatment effects model. The goal of the first stage is to obtain an estimation of the probability that on migrates. For this purpose, an instrumental variable is needed to predict the probability that someone would migrate, in addition to factors that affect the decision to migrate but cannot be affected by the act of migration (age and gender in our study). The instrumental variable must be strongly related to the independent variable (i.e., the migration-decision) but unrelated to the dependent variable (i.e., happiness). Spatial distance in kilometers to the nearest university offering a psychology-program was chosen as an instrument. That is, most migrants indicated moving for the purpose of studying psychology (67 %), which implies that only those who lived far away from a university felt the need to migrate (see column 1 in Table 5.8). The distance to a psychology program is unrelated to happiness. The adoption of this instrument implies that only the student population of our sample was utilized in this analysis (N = 101). The second stage examines the effects of the independent variables on happiness. A residual for each observation is obtained in both stages. Estimates are biased in a standard OLS if the unobservables in stage 1 are correlated with the unobservables in the stage 2 model. The negative q in Table 8 indicates a negative correlation between the unobservables in the first stage and in the second stage. Thus, unobservables are negatively (positively) related to migration and positively (negatively) related to happiness. This suggests that, if at all, migrants are positively selected from the population in terms of happiness, implying an increased gap when further removing selection effects.

This is confirmed when looking at the coefficient sizes of a standard OLS (B = 0.42) and the two-stage model (B = 0.51). The insignificant k indicates that there is no ground for preferring this more advanced model over a basic OLSmodel.

Table 5.8: Two stage treatment model

First-stage				Second-stage		
Probability to migrate	B (SE)			Overall happiness	B (SE)	
Distance to psychology-program (instrument)	0.023	**	(0.00)	Internal migration	-0.51	(0.56)
Age	0.11	+	(0.06)			
Gender	0.02		(0.47)	λ	-0.10	(0.41)
Constant	-4.55		(1.36)	N	101	
				ρ	-0.07	
				σ	1.44	

Note: B's of full set of covariates not reported. Significance levels at the 10%, 5%, and 1% level are denoted by $^{+}$, *****, and ****** respectively

DISCUSSION

The key goals of this paper have been to present a widely applicable methodology that can accurately zoom in on daily life and to advance the migration literature by addressing why internal migrants are typically unhappier than locals from a new point of view: the role of daily life. Global self-reports reveal that a gap remains present among young adults after controlling for socio-demographic differences. These gaps are shown to be particularly caused by a different experience of daily life. One part of the explanation comes from a different distribution of time. The young adult migrants allocated less time to active leisure and activities outside of home/work/transit. Specifically, they spent significantly less time on exercising and social drinking/parties. This is unfortunate as these activities are all associated with high momentary happiness and high global self ratings of happiness. A viable explanation for the time composition differences can be drawn from the broaden-and-build theory (Fredrickson 2001). The stressful lives of migrants lead them to have less mental and physical energy to engage in happiness-promoting activities as they are typically effortful. A snowball develops in which less openness to, and energy for, happiness-promoting activities leads to less positive feelings, which subsequently leads again to less mental energy, and so on. Locals also spend more time on the computer, which is generally associated with lower happiness. However, migrants were somewhat happier than locals when spending time on the computer. A profound reason for higher enjoyment and more time allocated to the computer may be the use of social media on the computer as a tool to communicate with

their friends and family who do not live close by. Another part of the explanation for the migrant–local happiness-gap comes from lower happiness of migrants while being with friends and while eating. The lower enjoyment of eating is not caused by eating with friends less often, but by experiencing less enjoyment while eating with friends. The social capital literature provides two potential explanations for migrants' lower enjoyment when being or eating with friends. Locals have lived in a certain place for a long time and have therefore had the opportunity to be selective in choosing friends and maintaining relationships over time. Migrants have had limited opportunities to build new social networks and must therefore choose from a considerably smaller pool of people (Putnam 2000). Relatedly, migrants' ties with others are typically not as strong as the ties that locals have with their long-time friends, leading to feelings of low social support and lower levels of comfort while spending time with the 'new friends' (Portes 2000). Temporal comparison processes may augment the dissatisfaction with current friends because the less intimate new friends compare negatively to long-time friends from the place of origin. Both the time composition differences and the differences in momentary happiness while being with friends explain a substantial part of the migrant–local happiness-gap. The knowledge regarding the impact of internal migration on happiness is limited, especially compared to international migration. This paper reveals that migrants can benefit from aid in daily life decisions. From a policy perspective, migrants appear to benefit from help in building new social lives and engaging in positive experiences such as active leisure. This can be achieved by local governments offering help, information, or discounts to migrants for becoming engaged in new communities and social- and active leisure activities in the host region. National governments can reduce migration stress by simplifying moving processes, such as the process of registering in a new city. Methodologically, we show that the use of technologies that zoom in on daily life add to the explanation of global subjective well-being. Future research can solve other economic, sociological, and psychological puzzles by using cost-effective smartphone applications that zoom in on the phenomenon under interest to obtain detailed information that is difficult to derive from general surveys. The black boxes opened by this methodology can greatly benefit public policy making as it helps to clarify what causes particular phenomena, thereby offering a step forward in acting upon these problems and thereby improving happiness in society. Despite its contributions, this study also has limitations. The independent variable may not have been optimal as spatial distance may be different to perceived distance. The future robustness of findings on internal migration can be improved by the development of an index of perceived distance including factors as travelling time and culture distance and complementing the measure of spatial distance. Another limitation was our sample; it was quantitatively limited, directed to a specific region, and included mostly females. We encourage future studies to check whether our results are generalizable to other regions, countries, and popula-

tions. A limitation in the measurement of happiness was that participants completed all six global subjective well-being scales consecutively, which lowers the advantages of measuring multiple constructs as biases as the impact of mood and question-order may consistently occur in the baseline-questionnaire. A limitation of applying both the ESM and the DRM is that there are likely to be carry-over effects from one activity to the next. A final limitation was the cross-sectional research design. Although we managed to minimise the chance of selection biases by using a twostage model, a longitudinal design incorporating pre-migration data would have been preferable. Yet, longitudinal datasets incorporating migration-data are largely unavailable and therefore this study is not an exception in having to resort to cross-sectional data. In conclusion, by relating the introduced smartphone application to a baseline questionnaire, we show the incremental value that multiple moment assessments can bring to general surveys in advancing knowledge. In this paper, this technology has been valuable in explaining the migrant–local happiness-gap by revealing the role of daily life experiences. We hope human knowledge will be advanced by applying similar technology on a wider scale.

Acknowledgments: The authors thank Stephan Erdtmann for technological help with constructing the application and Martijn Burger and Thomas de Vroome for comments on earlier versions.

6 How Does More Attention to Subjective Well-Being Affect Subjective Well-Being?

ABSTRACT

There is an ongoing debate as to whether pursuing happiness is beneficial for people's subjective well-being (SWB). To address this question, we tested whether attention to SWB – measured by participation in SWB surveys – is related to experienced SWB in two longitudinal studies. The initial study was conducted from November to December 2013 ($N = 129$), and the replication study, three years later from November to December 2016 ($N = 120$). The studies include two groups: one group (the control group) answered three SWB surveys over 4 weeks, and the other group (the experimental group) followed the same procedure but additionally tracked their SWB in detail using the experience sampling method four times a day and the day reconstruction method once a day using a smartphone application for two weeks to heighten their attention to their SWB. Both studies show higher SWB scores at later measurements compared to the first ones.

Keywords: Attention, Subjective Well-Being (SWB), Experience Sampling Method (ESM), Day Reconstruction Method (DRM), Mobile Application

INTRODUCTION

In the present paper, we investigate the relationship between subjective well-being (SWB) and attention to SWB. According to the OECD guidelines on measuring SWB, SWB is defined as "*Good mental states, including all of the various evaluations, positive and negative, that people make of their lives and the affective reactions of people to their experiences*." (OECD, 2013, page 10). In their paper, Diener, Scollon, & Lucas (2003) use happiness and subjective well-being (SWB) interchangeably, and in accordance such usage, Seligman & Csikszentmihalyi (2000) stated that the term SWB is only "a more scientific-sounding term for what people usually mean by happiness" (p. 9; also cp. Diener, 2000, p. 24). In line with these authors we will use the term SWB and happiness interchangeably as well. According to Brown & Ryan (2003), the term attention describes a heightened sensitivity for experiences. In our study, the heightened sensitivity to SWB is inflicted by participation in SWB surveys.

Understanding how attention to SWB is related to actual SWB levels is important given the strong interest in the pursuit of happiness by media and the general public alike (e.g., marketing claims such as Coca Cola's "Open Happiness" and number one music hits such as "Happy" from Pharrell Williams). We define the pursuit of happiness as every action an individual takes with the intent to increase its SWB for example hiring a happiness coach or reading a book on how to become happier. This interest in the pursuit of happiness is reflected in an exponentially rising number of scientific publications on happiness as well, with more than 500 every year (Veenhoven 2015) since 2000. Hence, it seems crucial to understand the effect that efforts to pursue happiness have on people's SWB. This issue seems especially relevant, since research shows that individuals who are most concerned with their own SWB tend to have lower levels of SWB (Mauss, Tamir, Anderson & Savino, 2011). Presumably those who are concerned about their own SWB also pay attention to their SWB levels; however, whether attention on its own has a negative or positive effect remains unclear. Thus, an important research question that needs to be answered first before evaluating specific actions to pursue happiness concerns how higher attention to SWB affects SWB.

Findings showing that attention to subjective well-being may be detrimental

Studies on the "dark side of happiness" (Gruber, Mauss & Tamir, 2011) suggest that explicitly pursuing high SWB can paradoxically lead to lower levels of SWB. Most notably, Schooler, Ariely and Loewenstein (2003) showed in an experimental study that people who were told to try to become "as happy as possible" by listening to hedonically ambiguous music [music that is novel to the listener and from which the listener does not have expectations of how the music should make them feel] rated themselves unhappier afterwards than a control group that merely listened to the same music. On a similar

note, Schooler and colleagues (2003) asked participants before New Year's Eve what they expected from the evening, and participants who had high expectations for the evening rated their SWB and their evening lower the next day, while Mauss, Tamir, Anderson and Savino showed in 2011 that participants who valued SWB higher reported lower levels of SWB and had lower mental health when they had low life stress.

Findings showing that attention to subjective well-being may be beneficial

In combination with reflection, attention can result in more awareness, which is – in turn – associated with higher SWB (Brown & Ryan, 2003).

In this regard, Wismeijer, Van Assen, Sijtsma & Vingerhoets (2009) showed that people who are more aware of their emotions and who are better able to label these emotions are better at regulating their mood and report higher SWB. Similarly, Beaman's longitudinal research (2010) on retirees' SWB supports the mediating role of emotional awareness to SWB at least by facilitating positive emotions. Lyubomirsky and Layous (2013) outline evidence that happiness can be actively increased by focusing on activities (behaviours or thoughts) that can increase SWB. Higher attention to SWB can help to find out the right activities for the right person in the right situation to increase SWB. Bryant, Smart and King (2005) showed that positive reminiscence can increase experienced happiness by paying more attention to a happy moment in the past. In a study by Bakker et al. (2016), people were invited by public commercials to share their happiness levels and to complete happiness diaries on a website using e-mail-reminders, with the goal of increasing the awareness of their SWB levels. The authors found a significant increase of 0.14 points on a 0-10 happiness scale for monthly happiness if participants completed at least 10 diaries.

In sum, attention influences awareness and higher awareness is associated to higher SWB levels.

Findings showing that surveying people frequently may increase awareness

Two survey methods increase people's attention and may increase their awareness. The first method is the experience sampling method (ESM; Csikzentmihalyi & Hunter, 2003; Csikszentmihalyi & Larson, 2014), a method that signals participants in the moment and asks them to report how they feel right now and what, with whom and where they are doing such activities. The second method is the day reconstruction method (DRM; Kahneman et al., 2004; Diener and Tay, 2014), a happiness diary method that asks participants to reconstruct their day in episodes and to report how they felt in these episodes.

Scollon, Prieto & Diener (2009) summarized ESM reactivity effects in their review of this method. They stated that being asked many times a day with this method can alert participants to their more inner thoughts and potentially change their behaviours. Likewise, Litt, Cooney & Morse (1998) showed that alcoholics reported that they were more

aware of their addiction after an ESM study, while Thompson and colleagues (2011) investigated the relationship between attention to emotion and affect intensity in an ESM study, and found that participants who were prompted eight times a day for seven days showed both higher positive and negative affect because of the higher attention to their emotions.

Hence, some evidence suggests that ESM studies increase participants' attention and awareness, resulting in emotional changes. Such an increase is likely to be observed for not only ESM but also DRM studies, as seen in the study from Bakker and colleagues (2016). Indeed, as participants reconstruct their day, they may obtain a better sense of what was going on in their life in comparison with those who did not participate in such an activity. Accordingly, participation in the DRM could lead to higher attention and awareness.

Findings showing that surveying people frequently may not increase awareness

Although it seems clear that ESM and DRM can increase attention, it is not clear whether the ESM and DRM increase participants' awareness. Thus, an individual that answers surveys about its SWB has to reflect on the questions, ratings and findings instead of just paying attention to the questions, ratings and findings to become more aware of its SWB. Without reflection there is just attention to the SWB surveys but not awareness for the SWB. In line with this Conner-Christensen (2003) showed that not all individuals report their personal experience and reflect on it deeply to become more aware. Indeed, some participants merely report semantic knowledge or respond more behaviourally to the prompts. In line with this finding, Killingsworth & Gilbert (2010) showed in an ESM study that for 46.9% of the prompts, participants' minds were wandering, which indicates that they were not aware of the current situation.

Summary

On the one hand, evidence shows that explicitly pursuing higher levels of SWB and valuing SWB to a greater extent increases the chance that individuals are disappointed and thus have lower levels of SWB.

On the other hand, we expect that the ESM and DRM enhance attention and that greater attention in combination with reflection can help to increase awareness. Greater awareness is then associated with higher SWB, and hence, it should be beneficial to pay more attention to SWB.

Thus, given the literature so far (especially the study from Bakker et al. 2016), we do not expect a negative effect and even expect a potential increase in participants' SWB by paying more attention to their SWB when investigating a non-clinical sample.

A metaphor can help to understand this thought: A speedometer in a car should help drivers pay attention of the speed and thus minimize the risk of having an accident. If the driver has good driving skills and if the track is easy, he does not need to look at the speedometer too often because he has a good feeling of the speed and can focus on experiencing the track and focus on driving. Adapted to the old idea of a hedonimeter (Edgeworth, 1881), this driver would be in the 'green' (Thriving) zone and identify the merit of experiencing higher awareness rather than the merit of meta-awareness (Schooler & Mauss, 2010). Accordingly, more attention to SWB would not have a strong positive effect but also not a strong negative effect. If the driver has good driving skills but the track is difficult to traverse or vice versa, he or she needs to pay more attention to his or her speed and look at the speedometer more often to reach his or her goal. Such a driver would be in the 'yellow' (Struggeling) zone and need to pay a bit more attention to his or her happiness. Thus, he or she would need more meta-awareness and a bit less experience to find the right track or just realize that he or she is on the right track to be able to live in the green zone. Accordingly, more attention to SWB, would have a positive effect. If the driver has poor driving skills and if the track is difficult, it does not seem likely that looking at the speedometer all the time would be the right way to reach the goal. The driver would just focus on the speed, which is not his biggest problem, as he or she first needs to learn to drive (again). Such a driver would be in the 'red' (Suffering) zone and would, for example, suffer from depression. If he or she would focus on his or her (un)happiness every day instead of thinking about doing his or her clinical interventions right, he or she would not become happier. Accordingly, more attention to SWB would have a self-defeating effect. Thus, in a non-clinical sample without too many participants in the red zone we would expect a positive effect by paying more attention to SWB on participants' SWB.

This study

If paying more attention to one's SWB has a clear negative impact on SWB, then people should not pursue happiness directly by hiring a happiness coach or reading books about happiness; instead, they should focus more on other aspects of their life to increase their SWB. If paying more attention to SWB is beneficial people should track their happiness in more detail to raise their awareness and understand better which activities, in which situations would increase their SWB. Thus, to investigate the role of more attention to SWB in how and whether we should pursue happiness seems to be very important and is investigated in this study.

Unfortunately, to our knowledge no other study investigated the effect of attention to SWB in an experimental study. Bakker et al. (2016) only had a quasi-experimental approach investigating the effect of the frequency and length of use of a website to track happiness on the happiness ratings. In order to investigate their finding in more

detail we report a controlled longitudinal study that we ran in 2013 and a replication study that we ran in 2016 to investigate if paying more attention to SWB has a positive effect on a non-clinical sample.

METHOD

Initial study in 2013

Sample

Our sample for the initial study included psychology students of Heinrich-Heine University Düsseldorf who need to participate in experiments for course credits. All participants needed to be able to speak English or German. The participants in the experimental group needed to have an Android smartphone. The psychology students were recruited from presentations of the study in their classes. To control for participants' motivation to participate in the study, they were first asked whether they would participate in the group that would use the app. If they answered that they would like to participate but that they did not have an Android smartphone, they were informed about a different study in which they just needed to complete some online-questionnaires about their SWB (the control group). Hence, the sample was not truly randomly assigned; the difference between the two groups was not in their motivation, it was only the difference that participants in the experimental group owned an Android smartphone and that participants in the other group did not. As a reward, every participant who finished the entire study took part in a lottery for a 250 Euro Amazon voucher and got an individual SWB profile. Additionally, the psychology students in the experimental group got eight half participant hours (course credits), and the psychology students in the control group got four (they need to have 60 in total as part of their bachelor degree).

Overall, 188 people finished the first questionnaire (134 experimental group), 170 finished the second one (122 experimental group), and 156 finished the whole study (112 experimental group). Thus, the attrition rates were quite similar for both groups (16.5% experimental group; 18.5% control group). In this paper, we will just look at participants who finished all three baseline questionnaires, did at least 50% of the DRM and ESM modules if they were in the experimental group (as a manipulation check) and did not have changes in their life satisfaction that were higher than two standard deviations between the different measurements because such changes could be due to other life events and not just study participation. Thus, the final sample comprises 129 participants (90 experimental group), with a mean age of 21.95 (SD = 4.63) and 108 women (74 experimental group) and 21 men (16 experimental group). The descriptive characteristics of the sample can be seen in Table 6.1:

Table 6.1: Descriptive characteristics of the sample

Variable	Experimental Group 90	Control Group 39	
	Mean (*SD*) / %	Mean (*SD*) / %	*F* value
Age	21.24 (4.07)	23.56 (5.68)	6.88* 6.88*
Gender (% male)	17.8%	12.8%	0.49
Education			1.54
High School	84.4%	76.9%	
Vocational Training	7.8%	7.7%	
Bachelor or Higher	7.8%	15.4%	
Personality			
Extraversion	1.29 (2.82)	0.95 (2.98)	0.38
Conscientiousness	3.06 (2.21)	3.10 (2.25)	0.01
Openness	2.38 (2.14)	2.56 (1.97)	0.22
Agreeableness	2.26 (1.97)	2.69 (2.08)	1.29
Emotional stability	4.53 (1.52)	4.44 (1.52)	0.11
Chronic condition (yes)	8.9%	5.1%	0.53
Immigrant	4.4%	2.6%	0.26
Height	170.42 (8.53)	172.62 (7.23)	1.96
Weight	63.18 (11.94)	64.31 (12.34)	0.24
Alcoholic	1.1%	0.0%	0.68
Smoker	5.6%	7.7%	0.97
Religion			0.28
Atheist	21.1%	15.4%	
Catholic	32.2%	41.0%	
Protestant	33.3%	23.1%	
Others	13.3%	20.5%	
Monthly Income			1.55
Below modal	80.0%	74.4%	
Modal (€2.500 net)	15.6%	12.8%	
Above modal	4.4%	12.8%	
Having a job (% yes)	37.8%	43.6%	0.38
Household situation			0.34
At parents' home	41.1%	25.6%	
Alone	23.3%	33.3%	
With partner	34.4%	12.8%	
Flat-sharing	1.1%	28.2%	

Note: The table displays the exact descriptive characteristics of the full sample, experimental group and control group. As can be seen, the whole sample is comparably young, there are more women than men, and the participants are highly educated for their age, have low immigrant rates, are not overweight, live healthily according to alcohol and smoking habits, have a lower income and have a mixed household situation. Significant differences between the two groups are displayed with their *F*-values calculated by a MANOVA (* $p < .05$; ** $p < .01$). The questionnaire group is significantly older.

Materials

The study used three different material modules: i) the self-developed smartphone app; ii) the online questionnaires measuring participants' SWB (SWB Questionnaire; see Appendix A); and iii) some additional questionnaires (see Appendix D).

Smartphone application

Regarding the increasing attention to SWB, we developed a smartphone application that combined two widespread and commonly used scientific techniques in happiness research (Diener & Tay, 2014): i) ESM and ii) DRM. With both methods we try to increase the attention that participants pay to their SWB. Details about the app will be explained in the study procedure below.

SWB questionnaire

To measure the effect of increased attention for participant's SWB on their SWB, an online-questionnaire was programmed with the software EFS survey. This questionnaire was based on the OECD guidelines for measuring SWB (OECD, 2013) and consists of the following components: i) Overall SWB Scores; ii) Affective SWB Scores; iii) Cognitive SWB Scores.

Overall SWB measurements:

The first question was a *core measure of happiness* (HC) from the European Social Survey (ESS; 2013). Participants had to rate how happy they are on a Likert scale from 0 to 10. The test-retest reliability lies between .50 and .70 in prior research (Krueger & Schkade, 2008).

The second question was a *core measure of life satisfaction* (LC) from the European Social Survey (ESS; 2013). Participants had to rate how satisfied they are with their life on a Likert scale from 0 to 10. The test-retest reliability after two weeks also lies between .50 and .70 in prior research (Krueger & Schkade, 2008).

Affective SWB measurements:

The *Affect Balance Scale* (ABS; Bradburn, 1969) was integrated into our survey as a third measurement to assess positive and negative affect in more detail. Participants are asked to indicate whether they have felt a given emotion (5 positive and 5 negative ones) in the last two weeks (yes/no). If a negative emotion is rated with "yes," a -1 is coded, and if a positive emotion is rated with "yes," a +1 is coded. When these ratings are summed, the "Affect Balance" is calculated with a range between -5 and +5. Unfortunately, the only published test-retest study is still from Bradburn (1969), but it shows a test-retest reliability of .76 after three days, and the internal consistency is between .55 and .73 (Cronbach's α) for positive affect and .61 and .73 for negative affect in prior research.

Cognitive happiness measurements:

To measure participants' life satisfaction in more detail than in the core-question, the fourth measurement included is the *Satisfaction with Life Scale* (SWLS; Diener, Emmons, Larsen & Griffin, 1985). The SWLS is a short 5-item instrument on a 7-point Likert scale designed to measure global cognitive judgments of satisfaction with one's life. The scale usually requires only one minute to answer, and it is often used in research studies all over the world. The test-retest reliability is above .80 (Diener, Inglehart & Tay, 2013) in most studies, and the internal consistency Cronbach's α is about .80 or higher in prior research (Diener et al., 2013; Eid & Diener, 2004).

To measure the construct of eudaimonia, the *Flourishing Scale* (FS; Diener et al., 2009) was integrated into the survey as the fifth measurement. The Flourishing Scale is a brief 8-item summary measure of the respondent's self-perceived success in important areas such as relationships, self-esteem, purpose, and optimism. The scale provides a single psychological well-being score. Its test-retest reliability is about .80, and the internal consistency about .80 in prior research (Cronbach's α).

As a last measurement, we added the *Domain Evaluation Questionnaire* (DEQ; OECD, 2013), which is a module suggested by the OECD (2013) to measure satisfaction in 10 different domains of life (e.g., health) on an 11-point Likert scale. Because it is a very new scale, no evidence of test-retest reliability is available. However, because of the OECD's recommendation and the provision of additional information, it was implemented in the survey.

Additional questionnaires

A demographic questionnaire collecting some personal information was integrated at the end of the first survey. To control for personality effects, the Big Five short scale (TIPI; Gosling, Rentfrow & Swann, 2003) was used. For details, see Appendix D.

Study procedure

The whole study procedure lasted about six weeks. In the first week, the study was presented to evaluate different happiness measurements among psychology students at the Heinrich-Heine University Düsseldorf in their classes. First, participants who owned an Android phone signed up to a list with their email address in order to participate in the more intense measurement group (experimental group). Afterwards, participants without a smartphone signed up on a list with their email address in order to participate for the repeated measurement group (control group).

To ensure that as many of the participants who signed up to take part in the study were really going to participate, reminder emails for the start of the study were sent on Sunday November 3rd 2013, Wednesday November 6th 2013 and Thursday November 7th 2013 before the start of the study.

To control the effects of weekday (e.g., Helliwell & Wang, 2014) and time of day, the first online questionnaire had to be filled in by both groups on Thursday, November 7th 2013 between six pm and three am. Before starting the questionnaire, participants had to create their own individual participant code. The first online questionnaire consisted of the SWB questionnaire, the personal information questionnaire and the Big Five questionnaire outlined above. The entire online questionnaire took the participants about 13 minutes to finish.

After finishing the same online questionnaire as the control group, the experimental group received an explanation about how to download the app on the next page of the survey. For participants who had problems downloading the app, there was a service hotline (3 people needed help). Ultimately, every participant succeeded in downloading the app.

To ensure an easy download and installation process for the participants, the app was published in the Google Play Store for the duration of the study. After downloading the app, the participant needed to sign up with an individual participant code created by the participant. By doing so, participants could ensure that their data were anonymous and that it could not be tracked, so social desirability biases should have been reduced. All tracking information (e.g., GPS information) were automatically deleted.

To ensure that participants understand how to use the app, a tutorial explaining all different functions and screens was provided. It was not possible to skip this tutorial; thus, every participant had the same understanding of the app. As part of the tutorial, the app sent examples of the two different notifications that popped-up, such as a short message.

The first one was a notification for the ESM: when the participant clicked on the notification, the app asked the participant to merely rate his or her SWB on a smiley scale between 0 and 10 (Veenhoven, 2004). Next, the participant was asked to describe what he or she was doing in that moment using a series of multiple-choice questions. Sequent, the participant rated where he or she was in that moment and with whom. If an answer to a notification took the participant about 30 seconds after some experience, it was possible to give an answer for a notification in nearly every situation.

The other notification was for the DRM. In this regard, participants were asked to reconstruct their previous day by describing it in episodes (e.g., breakfast from 9-10, and so on). For the description of the activity, the same questions as in the experience sampling questionnaire were used. After finishing the reconstruction of the previous day in episodes, participants had to rate episodes on a smiley scale (Veenhoven, 2004).

On the home screen, participants could see their ratings on different graphs to ensure high attention to SWB. Thus, participants were able to see how happy they were during different activities, with different people or in different locations, how they used their

time and how happy they were on which day. All this information was based on the participant's answers in the ESM or DRM.

Although downloading the app and signing up with a participant code requires a data connection (wi-fi or cellular), the app itself worked both with and without an internet connection. If the participant completed the survey with no connection, the data were transferred to the server as soon as the user had an internet connection. Figure 6.1 displays the ESM and DRM of the self-developed smartphone application.

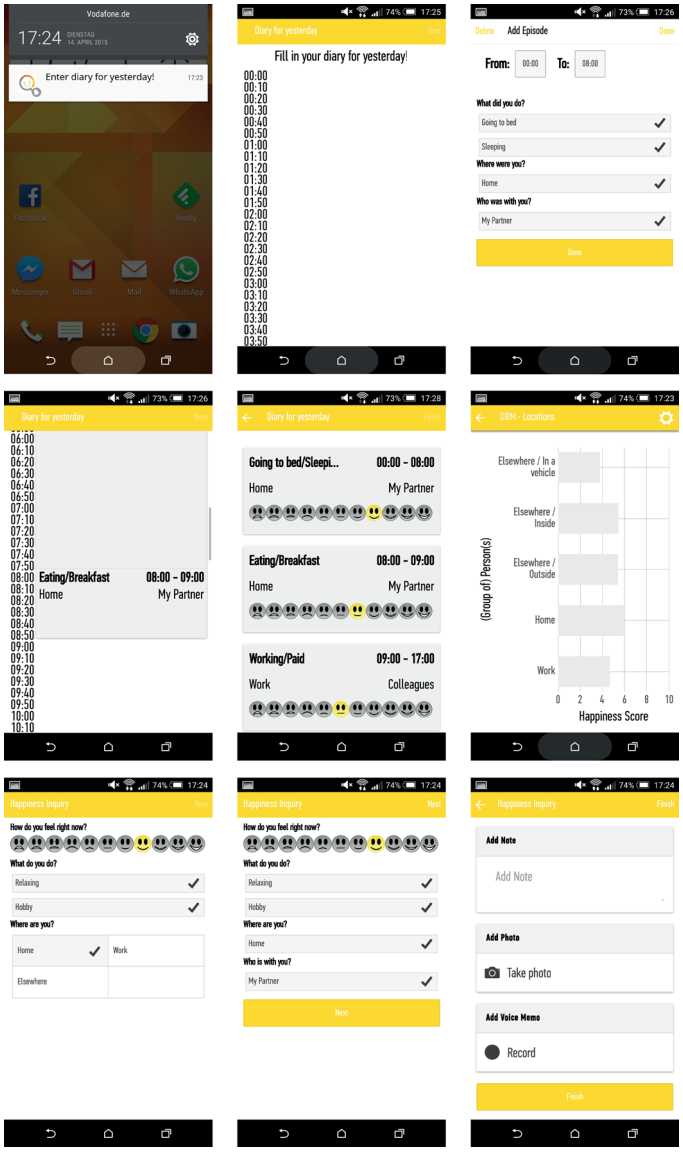

Figure 6.1: The figure displays screenshots of the self-developed smartphone application.

For two weeks, the experimental group used the smartphone application. The intensity of notifications was high to ensure that participants were really paying attention to their happiness and that enough data was collected to give them graphical feedback about their SWB. Accordingly, the participants got six ESM notifications per day and were asked to do the DRM every day. The effort was about 3-5 minutes for the ESM notifications and about 10-15 minutes for the DRM per day.

On Sunday, Wednesday and Thursday before the second questionnaire on Thursday, November 21st 2013, reminder mails were sent to the participants. The experimental group was told to ensure that their smartphones would be connected to the internet on Thursday to ensure that all the data could be transferred to the server. On Thursday afternoon at six pm, the app was deleted, so that it was not possible to use it anymore.

Subsequently, both groups were asked to complete just the SWB questionnaire a second time between six pm and three am. Hence, the online questionnaire took only about 6-8 minutes to complete. The participants again needed to sign in with their participant code for this purpose.

Two weeks later, both groups had to complete the SWB questionnaire a third time. Reminder mails were again sent on Sunday, Wednesday and Thursday, December 5th 2013, before the start of the third online questionnaire. This was the same for the second questionnaire for both groups.

When they finished the last SWB questionnaire, participants were asked to send an email to a previously unknown e-mail address (happinessprofile@web.de) to confirm that they finished the study. Only the participants that sent this mail were allowed to participate in the lottery to receive the 250 Euro Amazon voucher, to get a link to their individual SWB profile and to a get certificate to get their participant hours (course credits). In this way, it was possible for participants to remain anonymous, as they did not need to report their individual participant code.

After we exported their data, individual SWB profiles could be automatically generated. The participants who finished the study received a link to their profile in their "Thank You" mail. To ensure anonymity once again, participants had to sign up with their individual participant code again on this webpage to be able to access their individual happiness profile. This profile showed their data in the form of graphs, and the data were explained to them. In the "Thank You" e-mails, all participants in the app group were invited to attend a "mulled wine-evening" at Heinrich-Heine University. However, this event was organised not only to say thank you but also to evaluate the app and the study process among the participants. The lottery for the 250 Euro Amazon voucher also occurred on this evening. Additionally, in the "Thank You" e-mails, the psychology students were informed of the dates when they could get their signatures for their participant hours (course credits).

The whole study process took about eight weeks: two weeks to recruit participants, four weeks to conduct the study itself and two weeks to finish the study by sending the profiles, the "Thank You" e-mails, the non-responder questionnaires, the "mulled wine evening" and the signatures for the participant hours (course credits). Figure 6.2 summarizes the study procedure:

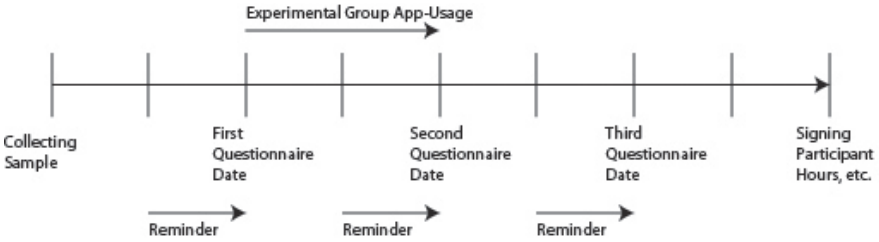

Figure 6.2: The figure displays the study procedure. Further details are explained in the text above.

Study design

For the study, a one-factorial repeated between-subjects design with two intensities was implemented. The single factor is an increase in attention for SWB, which should be higher for the experimental group and lower for the control group. To measure the SWB levels in both groups, six dependent variables were measured three times in four weeks in both groups: the happiness core question score (HC), the life satisfaction core question score (LC), the domain evaluation questionnaire score (DEQ), the flourishing scale score (FS), satisfaction with life scale score (SWLS) and the affect balance scale (ABS) score.

Replication study in 2016

Sample

The sample and sampling process was nearly the same as in 2013. The main difference was that we wanted to run the whole study on smartphones and that we wanted to have a fully randomized sampling process. Thus, the app was available for Android smartphones and iPhones this time, and we merely allowed psychology students to participate in the study if they owned either an Android smartphone or an iPhone. As the app could now also display the SWB questionnaires, participants signed up in class without knowing the group for which they would be selected. By waiting three years for the replication study, we ensured that none of the participants participated in the initial study.

In sum, 197 people finished the first questionnaire (105 in the experimental group), 152 finished the second one (79 in the experimental group), and 135 finished the whole study (72 in the experimental group). The attrition rates are the same in both groups but

are higher than in the first study because participants had to download an update for the app in the middle of the study (31.5% experimental group; 31.5% control group). Similar to the initial study sample, we merely assess participants who finished all three baseline questionnaires, did at least 50% of the DRM and ESM modules if they were in the experimental group and did not have changes in their life satisfaction that were more than two standard deviations between the different measurements because this could be due to other life events and not just from their study participation. Thus, the final sample comprises 120 participants (63 experimental group), with a mean age of 21.71 (*SD* = 5.10) and 101 women (52 experimental group) and 19 men (11 experimental group). The descriptive characteristics of the sample are summarized in Table 6.2:

Materials

Most materials remained the same for the replication study. The main differences are that now all questionnaires were displayed in the app and that there are no external online questionnaires to realize a fully randomized sampling process.

As we doubted that the ABS would be sensitive enough because it merely asks participants to give "yes" or "no" answers, we replaced it with the Scale of Positive and Negative Experience (SPANE; Diener et al. 2010; for details see Appendix B), which asks respondents about the strength of 12 different emotions in the last weeks on a 7-point Likert scale.

Additionally, we added the mindful attention awareness scale (MAAS; Brown & Ryan, 2003; for details, see Appendix C) to determine whether people merely pay more attention to their SWB in the experimental group or whether they really have a higher general awareness.

Study procedure

The general procedure was almost similar to that of the initial study in 2013. To ensure that the effects were not due to the context of Christmas becoming closer, we decided to start the study one week later than we did in 2013. In the first week (November 7th to November 13th 2016), we collected our sample by presenting the study in psychology classes. On Sunday, Wednesday and Thursday, we reminded participants about the start of the study by e-mail. On Thursday November 17th 2016, participants were able to download the app in the Google Play Store or the Apple App Store at six pm.

By downloading the app, participants were randomly assigned to the experimental or the control group. Subsequently, participants filled in the first questionnaire and were informed about the further study process afterwards. Participants in the experimental group did the tutorial afterwards and did the ESM and DRM modules for the next two weeks. Because participants gave feedback in 2013 that six ESM modules per day required too much effort, we decided to reduce the number to four ESM modules for this study.

Table 6.2: Descriptive characteristics of the sample

Variable	Experimental Group 63	ControlGroup 57	
	Mean (*SD*) / %	Mean (*SD*) / %	*F* value
Age	21.43 (4.26)	22.02 (5.90)	0.40
Gender (% male)	17.5%	14.0%	0.26
Education			1.11
High School	88.9%	84.2%	
Vocational Training	3.2%	0.0%	
Bachelor or Higher	7.9%	15.8%	
Personality			
Extraversion	1.44 (2.84)	0.96 (2.48)	0.96
Conscientiousness	3.14 (2.15)	3.23 (2.20)	0.05
Openness	2.52 (2.35)	2.40 (2.25)	0.08
Agreeableness	2.48 (1.93)	2.12 (2.04)	0.95
Emotional stability	4.56 (1.57)	4.79 (1.57)	0.66
Chronic condition (yes)	9.5%	8.8%	0.02
Immigrant	4.8%	5.3%	0.02
Height	171.16 (6.96)	169.49 (8.17)	1.46
Weight	63.09 (12.17)	60.70 (9.74)	1.39
Alcoholic	1.6%	0.0%	0.43
Smoker	6.3%	1.8%	2.93
Religion			0.24
Atheist	23.8%	29.8%	
Catholic	41.3%	36.8%	
Protestant	25.4%	24.6%	
Others	9.5%	8.8%	
Monthly Income			2.49
Below modal	85.7%	73.7%	
Modal (€2.500 net)	7.9%	14.0%	
Above modal	6.3%	12.3%	
Having a job (% yes)	61.9%	45.6%	3.23
Household situation			8.91**
At parents' home	28.6%	42.1%	
Alone	36.5%	10.5%	
With partner	12.7%	15.8%	
Flat-sharing	22.2%	31.6%	

Note: The table displays the exact descriptive characteristics of the full sample, the experimental group and the control group. As can be seen, the whole sample is comparably young, there are more women than men, participants are highly educated for their age, they have low immigrant rates, they are not overweight, they live healthily according to alcohol and smoking habits, they have a lower income, and they

a mixed household situation. Significant differences between the two groups are displayed with their *F*-values calculated by a MANOVA (* $p < .05$; ** $p < .01$). The household situation differs significantly.

On the Sunday, Wednesday and Thursday before the second questionnaire on Thursday December 1st between 6pm and 3am, the app reminded the participants to participate in the study additionally to the reminder e-mails. For the third questionnaire on Thursday, December 15th between 6pm and 3am, the same procedure was used. Participants were able to see their final profiles, including all questionnaire results right after they finished the third questionnaire.

The lottery voucher was sent to one participant on Friday December 16th through the app. In the last study week, the app reminded the participants of the dates when they could get their participant hours signed and when the mulled-wine evening was going to happen. Thus, we were able to shorten this process from two weeks to one week to be able to finish the study before the Christmas Holiday.

Study Design

The study design remained the same as that for the replication study. The only differences were that the SPANE replaced the ABS as a dependent variable and that we added the MAAS as a dependent variable.

RESULTS

Initial Study 2013

Between group results

To examine whether there is a significant difference in the change of the scores of the dependent variables between the control and experimental groups, a repeated-measures MANCOVA between them was conducted for all dependent variables controlling for all covariates. There was no significant effect between the two groups (Pillai's Trace: $F = 0.752$; $p = .697$; $\eta^2 = .009$).

Within group results

To investigate in general whether paying more attention to one's SWB affects one's SWB, the differences in the questionnaire scores in all three assessments (t1, t2, t3) were tested by repeated-measures MANCOVAs for every dependent variable for the control group and for the experimental group. Tables 6.3 and 6.4 display the results. The results for the experimental group show increased ratings for all scores (t1 vs. t3) except the ABS and significantly higher ones for DEQ (Pillai's Trace: $F = 5.165$; $p = .008$; $\eta^2 = .129$), FS (Pillai's Trace: $F = 3.493$; $p = .036$; $\eta^2 = .091$) and SWLS (Pillai's Trace: $F = 6.097$; $p = .004$;

$\eta^2 = .148$). The control group also shows higher ratings for all scores except HC and ABS, but none of the changes are significant. The results do not differ significantly if we do not control for all covariates.

Table 6.3: Results ANOVAs: Within results for the different questionnaires in the experimental group (*N*=90).

Measure	Score t 1	Score t 2	Score t 3	*Cohen's D (t2 – t1)*	*Cohen's D (t3 – t2)*	*Cohen's D (t3 – t1)*	*F* value	*p* value	η^2
HC	6.76	7.13	7.04	.240	-.055	.189	2.617	.080	.070
LC	6.81	6.91	7.00	.059	.056	.109	0.894	.414	.025
DEQ	6.75	6.98	7.00	.164	.011	.177	5.165**	.008	.129
FS	5.41	5.53	5.48	.146	-.068	.077	3.493*	.036	.091
SWLS	5.04	5.20	5.17	.147	-.028	.123	6.097**	.004	.148
ABS	0.77	0.82	0.73	.029	-.046	-.018	0.140	.870	.004

Note: HC = Happiness Core; LC = Life Satisfaction Core; DEQ = Domain Evaluation Questionnaire; FS = Flourishing Scale; SWLS = Satisfaction with Life Scale; ABS = Affect Balance Scale. Significance Levels of $p < .05$ and $p < .01$ are displayed by * and **.

Table 6.4: Results ANOVAS within for the different questionnaires in the control group (*N*=39).

Measure	Score t 1	Score t 2	Score t 3	*Cohen's* D (t2 – t1)	*Cohen's* D (t3 – t2)	*Cohen's* D (t3 – t1)	*F* Value	*p* value	η^2
HC	6.67	6.41	6.59	-.137	.100	-.047	0.371	.695	.038
LC	6.64	6.69	6.92	.025	.122	.140	0.910	.420	.087
DEQ	6.70	6.75	6.80	.029	.026	.055	0.404	.673	.041
FS	5.15	5.29	5.20	.151	-.097	.057	2.156	.143	.185
SWLS	4.74	4.83	4.91	.076	.061	.135	0.980	.393	.094
ABS	0.67	0.72	0.64	.027	-.041	-.014	0.102	.904	.011

Note: HC = Happiness Core; LC = Life Satisfaction Core; DEQ = Domain Evaluation Questionnaire; FS = Flourishing Scale; SWLS = Satisfaction with Life Scale; ABS = Affect Balance Scale. Significance Levels of $p < .05$ and $p < .01$ are displayed by * and **.

Replication Study 2016

Between group results

To investigate whether we find similar effects to the initial study in 2013, we ran the same calculations for the dataset of the replication study in 2016. The repeated measures MANCOVA did not show significant differences in the results between the two groups (Pillai's Trace: $F = 0.660$; $p = .785$; $\eta^2 = .008$).

Within group results

To investigate whether we find similar effects to the initial study in 2013, we ran the same calculations for the dataset of the replication study in 2016. The results are displayed in

tables 6.5 and 6.6. The experimental group shows increased ratings in all scores (t1 vs. t3) and significantly higher ones in the DEQ (Pillai's Trace: $F = 4.681$; $p = .014$; $\eta^2 = .179$). In the SPANE, we see significant variance (Pillai's Trace: $F = 6.834$; $p = .003$; $\eta^2 = .241$). First, there is a significant drop indicating lower ratings on positive emotions and then an increase indicating more positive emotions. The control group shows higher ratings in all scores (t1 vs. t3), except SPANE, and significantly higher ones for the DEQ (Pillai's Trace: $F = 7.803$; $p = .001$; $\eta^2 = .297$). The results do not differ significantly if we do not control for all covariates.

Table 6.5: Results ANOVAs: Within differences for questionnaires in the experimental group (*N*=63).

Measure	Score t 1	Score t 2	Score t 3	*Cohen's D (t2 – t1)*	*Cohen' D (t3 – t2)*	*Cohen' D (t3 – t1)*	*F* value	*p* value	η^2
HC	7.37	7.57	7.62	.143	.039	.180	1.476	.240	.064
LC	7.38	7.51	7.68	.088	.113	.209	1.660	.202	.072
DEQ	6.98	7.24	7.31	.226	.064	.295	4.681*	.014	.179
FS	5.63	5.68	5.67	.074	-.016	.053	0.332	.719	.015
SWLS	5.11	5.22	5.29	.120	.063	.181	2.639	.083	.109
SPANE	-0.04	-0.34	0.01	-.408	.459	.050	6.834**	.003	.241

Note: HC = Happiness Core; LC = Life Satisfaction Core; DEQ = Domain Evaluation Questionnaire; FS = Flourishing Scale; SWLS = Satisfaction with Life Scale; SPANE = Scale of Positive and Negative Effect. Significance Levels of $p < .05$ and $p < .01$ are displayed by * and **.

Table 6.6: Results ANOVAs: Within difference for questionnaires in the control group (*N*=57).

Measure	Score t 1	Score t 2	Score t 3	*Cohen's D (t2 – t1)*	*Cohen's D (t3 – t2)*	*Cohen's D (t3 – t1)*	F value	p value	η^2
HC	7.16	7.42	7.37	.166	-.030	.129	1.149	.328	.058
LC	7.39	7.58	7.60	.124	.012	.120	0.976	.386	.050
DEQ	7.02	7.19	7.50	.128	.229	.374	7.803**	.001	.297
FS	5.62	5.60	5.65	-.032	.064	.038	0.591	.559	.031
SWLS	5.15	5.30	5.30	.139	.004	.136	2.400	.105	.115
SPANE	0.09	-0.10	0.01	-.217	.123	-.091	0.605	.551	.032

Note: HC = Happiness Core; LC = Life Satisfaction Core; DEQ = Domain Evaluation Questionnaire; FS = Flourishing Scale; SWLS = Satisfaction with Life Scale; SPANE = Scale of Positive and Negative Effect. Significance Levels of $p < .05$ and $p < .01$ are displayed by * and **.

Awareness Effects

As stated in the method part, we added the MAAS for the replication study to investigate whether people merely pay more attention to their SWB or whether their also awareness increases. To investigate this effect, we ran a repeated measures MANCOVA within the two groups and a repeated measures MANCOVA while controlling for all covariates displayed in table 2 between the two groups. We find that awareness significantly increases

in the experimental group (Pillai's Trace: $F = 5.916$; $p = .005$; $\eta^2 = .216$) but not in the control group (Pillai's Trace: $F = 0.318$; $p = .730$; $\eta^2 = .017$). The differences between the two groups are not significant (Pillai's Trace: $F = 2.835$; $p = .064$; $\eta^2 = .054$).

DISCUSSION

Summary

This study aimed to answer the question *"How does more attention to SWB affect SWB?"*. Our hypothesis was that paying more attention to SWB would have a positive effect on SWB in a non-clinical sample. We investigated our research question in two longitudinal studies mixing two different approaches (ESM and DRM). Two groups completed three SWB questionnaires within four weeks, and one of these groups (the experimental group) additionally completed four ESM modules per day and one DRM module per day for two weeks to increase their attention for their SWB.

Both studies show that more attention to one's SWB has no negative effect on one's SWB and that within the experimental group, there are significant positive effects in both studies. These differences do not significantly differ from the control group.

Methodological discussion

The representability of this study for the general population is of course limited, as the sample merely comprises psychology students in Düsseldorf, Germany. Psychology students may be challenging subjects for this kind of study because they probably already reflect their life quite a lot and thus it is harder to increase their attention to SWB and their awareness.

Additionally, the statistical power is limited owing to the sample size. A bigger sample was unfortunately not available for the two studies mainly because of a lack of funding which would have been mandatory to increase the representative sample size for this intense study.

After the initial study, we were not confident whether some of the effects and differences between the two groups might have arisen because we did not have a random sampling between the two groups. As the results seem in general quite similar in both studies, it seems that the ownership of a (Android) smartphone in 2013 was not an important moderator.

In the initial study in the experimental group, only the ABS showed little variance over time. To increase the sensitivity, as the ABS merely gives the option to answer with "yes" and "no" for experiencing certain emotions, we decided to use the SPANE for the replica-

tion study, as it uses a 7-point Likert scale. Unfortunately, the ABS and SPANE scores are not comparable for the two studies.

To be able to explain the effects better, we added the MAAS. Unfortunately, we did not think of this in 2013; thus, we do not have comparable results.

Another limitation is that the follow-up was too short to capture a possible sleeper effect. Deferred effects are likely to exist if heightened awareness of how well one feels affects the level of subjective wellbeing though major life choices, such as the awareness that one feels not well in the presence of one's spouse leading to the decision to divorce. Such decisions take time, and they may be taken years after one has realized how one feels.

The graphical information that participants received in the app might play a role in the increased awareness of the experimental group. Thus, higher attention and awareness could be not just due to the DRM and ESM; rather, it could be due to the graphical feedback. Further, experimental studies controlling for whether participants get graphical feedback should be conducted in the future.

The greatest problem in interpreting the results is that the control group itself also shows increases over time in both studies. Thus, none of the differences between the experimental and control group are statistically significant, likely because of the effect of just filling in three SWB questionnaires. To rule out context effects, we chose a time where the psychology students do not write any exams in the period they participated in the study, and we controlled for bigger changes in the LC score and controlled whether it has an effect that Christmas is closer by starting the replication study a week later. In a follow-up study, there should be three additional randomly assigned groups that merely complete one questionnaire at one of the different questionnaire dates to completely rule out possible context effects. Unfortunately, our sample sizes were too small to split the sample in more groups.

Results discussion

Even when we consider all the methodological issues stated above, it can be concluded that there is not a negative effect owing to more attention to SWB in both studies. In both studies, none of the dependent variables show a significant negative effect between t1 and t3, and the DEQ and SWLS scores show significant increases in the experimental group for both studies additionally to descriptive increases in all other SWB indicators except the ABS in the initial study.

Given that only the participants in the experimental group report significantly increased awareness in the MAAS, it seems that using the DRM and ESM over two weeks can have a specific effect. This increased awareness seems to result in more emotional variance given the significant negative changes in the SPANE, indicating lower ratings for positive emotions between t1 and t2, and significant positive changes between t2

and t3, indicating higher ratings for positive emotions. Thus, it seems that higher general awareness makes participants more aware of their emotions and induces them to rate their SWB in scales that ask them about their overall happiness (HC), overall life satisfaction (LC, SWLS) or their general satisfaction with certain domains (DEQ) but not on a scale that indicates changes in the perception of a meaning in life (FS). From our point of view, participants understand better how satisfied they are with their life when they pay more attention to their SWB because of their increased general awareness.

That similar changes occur in the control group, indicating the same descriptive patterns as in the experimental group (except of the HC scores in the initial study), is not surprising. In the second study, the control group's well-being increased even more in the DEQ, although this should not be overinterpreted as DEQ is not a very common measure, its test-retest-reliability is unknown and it might be more context dependent than the other measures. Participating in a SWB study for four weeks to fill in three longer SWB questionnaires should also increase participants' attention to their SWB. Further, without using DRM and ESM for two weeks, participants should not have higher general awareness but should have a better perception of how satisfied they are with their life.

In comparison to the effect of the study by Bakker et al. (2016), the app group's effect seems comparably high. They find a long-term effect on participant's happiness doing 10 DRM modules of 0.14 on a 0-10 point happiness scale. In this study, we find an effect of about 0.25 points, which is higher than the effect found by Bakker et al. (2016). This result might be due to the following differences: i) participants paid higher attention to SWB because of the mix of ESM and DRM and because the results were always on their phone; ii) the effect might be higher in the first weeks and decreases afterwards; and iii) there might be a context factor that multiplied the effect. In general, it needs to be stated that both studies find a positive effect owing to more attention for SWB.

General Discussion

We see that paying more attention to SWB seems to increase awareness, which results in higher SWB ratings in a non-clinical sample.

Unfortunately, we do not have enough participants that have a very low SWB to compare the effects on participants with a very low SWB to participants with a medium SWB and to participants with a very high SWB in order to test the speedometer metaphor explained in the introduction. Further research should investigate this in more detail.

Thus, our results are limited in the way that we only have findings that support the hypothesis that paying attention to SWB is not having a negative effect on a non-clinical sample but cannot securely state that it has a positive effect on a non-clinical sample nor on a clinical sample.

Final conclusions

In conclusion, the findings show that paying more attention to one's SWB does not have a negative effect on one's SWB in general and might even have a positive effect. Even if it is not yet clear for whom and to what extent more attention to SWB has a positive effect, the results show that it is not generally negative to pay more attention to one's SWB and to give a positive signal for the high media attention for the pursuit of happiness and to tracking happiness in more detail to understand better which activities in which situations can increase SWB.

7 Using the Day Reconstruction Method: Same results when used at the end of the day or on the next day?

ABSTRACT

The day reconstruction method (DRM; Kahneman et al. 2004) constitutes a frequently used method aiming to capture everyday life and everyday feelings. In its original version and in most subsequent studies, respondents are asked to complete the DRM with respect to their previous day on the next day. Yet when asked, respondents prefer to work on the DRM in the evening of the same day, particularly in longitudinal studies where the motivation to complete the DRM proactively on one's own is crucial. Consequently, it is important to consider respondents' preferences about their favoured point in time to fill in the DRM. Thus, the question whether a flexible DRM usage that offers the freedom to work on the DRM in the evening of the same day or on the next day should arise. Reluctance in doing so is reasonable since research on differences in answering patterns between these two points in time is pending. The current study sheds light on this research question by comparing respondents' happiness during the reconstructed episodes in both settings (same day vs. next day) within one person. A DRM smartphone application was used with a group collected from the Innovation Sample of the German Socio-Economic Panel (GSOEP-IS). The results reveal that the point in time during which people fill in the DRM has no significant effect on respondent's happiness ratings. In sum, although an experiment is needed to replicate our findings, our research suggests that researchers might consider (especially for longitudinal studies) to give participants free choice to do the DRM on the evening of the same day or on the next day if they want to reduce response burden in order to increase participation rates.

Keywords: Day Reconstruction Method, Happiness, Time-Use, Timing, Rating Bias

INTRODUCTION

Several decades ago, researchers introduced a method meant to investigate the topic of subjective well-being or happiness as closely aligned to people's everyday lives by measuring their "daily moods over an extended period" (Sandvik, Diener, & Seidlitz, 1993, p. 321). When applying the so-called experience sampling method (ESM: Csikszentmihalyi & Larsen, 2014; Csikszentmihalyi, Larson, & Prescott, 1977; Hektner, Schmidt, & Csikszentmihalyi, 2007), also known as ecological momentary assessment (EMA: Stone & Shiffman, 1994; Stone, Shiffman, & DeVries, 1999), respondents are contacted via a beeper (or nowadays via their mobile devices: MacKerron, 2012; Hendriks, Ludwigs & Veenhoven, 2016) several times a day over the course of a longer time period by asking them to complete questions on what their current activity is, where they are doing it, with whom they are doing it and how happy they are doing it. In the original study, answers were recorded via a paper-pencil format; thus, respondents needed to carry this equipment with them all the time (Csikszentmihalyi et al., 1977). Consequently, researchers were required to make ESM sheets and pencils available, and ultimately, they needed to manually transfer the entire data sets to a PC. Overall, it is not surprising that the ESM/EMA is repeatedly described as a rather time-consuming and costly method to capture happiness during everyday life (Dolan & White, 2007; Kahneman et al., 2004; Krueger & Schkade, 2008). The ESM/EMA has also been criticized to only show excerpts of everyday life instead of depicting the whole day and thus not allowing precise time use information (Kahneman et al., 2004).

To address these shortcomings, Kahneman and colleagues (2004) developed the DRM, where respondents are regularly asked to reconstruct their prior day in single episodes, such as filling in a diary (e.g., 8-9 o'clock breakfast, 9-12 o'clock work). Thus, they indicate which activities they undertook in which time period, where as well as with whom and how happy they were doing so. In this way, the DRM has become a frequently used (e.g., Bakker, Demerouti, Oerlemans, & Sonnentag, 2013; Berkhout et al., 2015; Bylsma, Taylor-Clift, & Rottenberg, 2011; Dockray et al., 2010; Doyle, Delaney, O'Farrelly, Fitzpatrick, & Daly, 2015; Hendriks et al., 2016; John & Lang, 2015; Kahneman et al., 2004; Knabe, Schöb, & Weimann, 2015; Stone et al., 2006) and gradually well-validated method (e.g., Anusic, Lucas, & Donnellan, 2016; Diener & Tay, 2014; Dockray et al., 2010; Kahneman et al., 2004; Schneider & Stone, 2016) that works more efficiently than the ESM/EMA, delivering additional information about people's time use during the course of entire days. According to Kahneman and colleagues' (2004) traditional version, respondents are asked to complete the DRM with respect to yesterday, meaning that one night is in between the day of interest and the current point in time. Overall, the vast majority of investigations are in line with this approach (Anusic et al., 2016; Berkhout et al., 2015;

Bylsma et al., 2011; Doyle et al., 2015; Hendriks et al., 2016; John & Lang, 2015; Knabe et al., 2015; Schneider & Stone, 2016; Stone et al., 2006).

In contrast, only a single study has required respondents to fill in the DRM after work to reconstruct their workday, in the evening directly before retiring (Bakker et al., 2013). Interestingly, if respondents are asked about their preferred DRM notification time, almost 70% would like to complete their diary in the evening before going to bed – whereas roughly 20% prefer the next morning (Ludwigs & Erdtmann, 2017). The other 10% did not have a preference. Consequently, current investigations neglect respondents' preferences regarding what point in time they are most motivated to complete the DRM. In fact, this could be a crucial point with respect to longitudinal studies where respondents are required to fill in their DRM modules every single day. For example, in Bakker et al.'s study (2013), about three-quarter of the employees asked to complete a DRM module during the evening of each of nine subsequent working days during which they ultimately finished all of them. Hence, about one-quarter of the respondents did not give their whole commitment for the study. Similarly, in another investigation (Hendriks et al., 2016), roughly three-quarter of respondents completed the minimum of DRM modules (always in the morning) required to be included in analyses (which was set at 10 of 15 DRM modules). Consequently, about one-quarter dropped out. In sum, both studies kept the DRM notification time constant and ultimately excluded about one-quarter of respondents from their analyses. Thus, making respondents motivated to fill in more DRM modules is essential for this research field. One promising step can be to give respondents the chance to make their day reconstructions at the point in time that they like to do it – either already on the same or on the next day. But how can such a flexible DRM application be organized in a way that is as efficient and as feasible as possible? With consideration to current data collection designs, it becomes obvious that this purpose is difficult to achieve, as some researchers use the self-administered, classical paper-pencil format (Bylsma et al., 2011; Kahneman et al., 2004; Stone et al., 2006), others use interviewers (Anusic et al., 2016; Berkhout et al., 2015; Doyle et al., 2015; Knabe et al., 2015), and several prefer a DRM web version (Bakker et al., 2013; Dockray et al., 2010; John & Lang, 2015; Veenhoven, Bakker, & Oerlemans[9]). All of these methods exhibit some weaknesses when trying to implement flexible DRM usage in longitudinal studies. In terms of the paper-pencil format, it becomes more difficult to control for respondents' commitment. Even if they are reminded by email or notified in another way, this message is not directly linked to the DRM booklet that they are required to work on. Hence, as soon as respondents receive a reminder via email and currently travel by train without carrying the booklet with them, they will probably forget to complete their DRM at a later point in time at home. With respect to the DRM method in which

9 Free website access: http://www.happinessindicator.com (lastly retrieved on: 2017-07-31)

an interviewer is involved, he/she needs to be able to react enormously flexible on the respondents' DRM answering time preferences. Additionally, this approach might be burdensome for respondents because, on the one hand, they may like to work on the DRM in the evening but, on the other hand, they may not want to be called at 11 pm in the evening by an interviewer. Finally, the web version is challenging. Respondents do not carry their PC everywhere, and those with a smartphone available do not have access to the internet at every moment. Consequently, if respondents do not have access to a PC or a web-enabled smartphone, they may simply forget to work on the diary. To solve this methodological challenge, recent digital technologies bring new possibilities for applying the DRM in a flexible way while being efficient and close to participants' behavioural preferences. For example, Hendriks and colleagues (2016) utilized a DRM smartphone application that is capable of working in an offline mode. Since respondents are used to carrying their mobile devices with them, the DRM survey is closely linked to their everyday lives. In addition, notifications sent to remind respondents on their DRM completion can immediately be replied to. In this way, more flexible DRM usage that is superior to currently common DRM implementations via paper-pencil, interviewers or websites can be realized.

Yet, before implementing such a flexible DRM application in research investigations, an important question needs to be answered: Do happiness ratings differ between these two points in time (same day vs. next day)? If the answer is "yes", the results of respondents completing their diary in the morning cannot be compared with results of others answering their diary in the evening, as different things seem to be measured. However, if the answer results in a "no," there is no reason why flexible DRM implementation should be avoided in future investigations. But what arguments play a role in this issue?

When making DRM judgments, respondents refer to the recent past and thus rate their happiness with consideration to retrospective episodes. Two kinds of memory that they can consciously made use of exist: semantic memory vs. episodic memory: The former "registers and stores knowledge about the world in the broadest sense and makes it available for retrieval. [...] [It] enables individuals to represent and mentally operate on situations, objects, and relations in the world that are not present to the senses: The owner of a semantic memory system can think about things that are not here now" (Tulving, 1993, p. 67). In contrast, the latter "enables a person to remember personally experienced events as such. That is, it makes it possible for a person to be consciously aware of an earlier experience in a certain situation at a certain time" (Tulving, 1993, p. 67). As the DRM refers to information as stored and retrieved in episodic memory, respondents should generally utilize this kind of memory when reconstructing their previous day. Another finding supports this point of view: Robinson & Clore (2002a) suggested that the transition from episodic to semantic memory can/could take several weeks. However, the DRM exclusively refers to the last day of memory, making it improbable that episodic memories

have changed to semantic ones. To conclude, respondents seem to apply their episodic memories when recalling single episodes of their previous day.

A challenge that needs to be addressed regarding this kind of memory is that episodic memory declines quickly (Rubin & Wenzel, 1996; Conway, 2001) or that past experiences are more difficult to retrieve as time elapses since encoding (Krosnick & Presser, 2010). Consequently, respondents cannot re-experience a certain emotional experience exactly as they were (Galin, 1994). Accordingly, when asked to reconstruct past experiences after a longer time delay, respondents are prone to unintentional recall biases (Krosnick & Presser, 2010) that can distort their answers and thus their perceived happiness levels. One example for such a bias consists in the "Peak-End effect" (Kahneman, 1999, p. 19): When trying to reconstruct an episode of the previous day, respondents orientate towards the strongest feelings experienced during this period of time ("Peak") as well as towards the final emotions just before episode ending ("End"). As memory loss should have progressed more on the next morning than during the evening, respondents are more likely to be more prone to this bias on the next morning. This finding indicates that the reconstruction of emotional experiences differs, probably leading to diverging episodic happiness judgements. Another crucial point is that respondents are used to referring to contextual information during retrieval process. Unfortunately, the ability to recall contextual details also diminishes quickly (Robinson & Clore, 2002b). Consequently, memory distortions with respect to experiential emotional information should be more probable on next morning than during the evening. Furthermore, another opportunity for a potential bias arises since respondents use their current emotions as a basis for retrieval (Levine, 2002). Research investigations have robustly shown that happiness levels tended to rise over the course of the day (Ayuso-Mateos et al., 2013; Csikszentmihalyi & Hunter, 2003; Daly, Delaney, Doran, & MacLachlan, 2011; Egloff, Tausch, Kohlmann, & Krohne, 1995; Kroenke, Seeman, Matthews, Adler, & Epel, 2012; Mihalcea & Liu, 2006; Sakawa, Ohtake, & Tsutsui, 2015), indicating that happiness levels are generally the lowest in the morning and the highest in the evening. Consequently, negatively connoted moods in the morning are able to distort happiness ratings for past experiences in a negative way. Overall, these findings indicate that because memory is likely worse for recent experiences during the next morning, respondents are subjected to various kinds of biases that can distort their memory for emotional experiences. Overall, this at first sight suggests differing answers of respondents completing the DRM in the evening of the same day vs. on the next day.

In contrast to this perspective, Conway (2001) claimed that episodic memory might be available for 24 hours. Therefore, it can be assumed that the representations of emotional experiences should also be quite accurate during the next day. This idea is supported by findings indicating the important role that sleep plays in the consolidation of memory (Diekelmann & Born, 2010). Memory consolidation "refers to a process

that transforms new and initially labile memories encoded in the awake state into more stable representations that become integrated into the network of pre-existing long-term memories" (Diekelmann & Born, 2010, p. 114) and has already been demonstrated for both emotional (Hu, Stylos-Allan, & Walker, 2006; Nishida, Pearsall, Buckner, & Walker, 2009) as well as episodic memory (van der Helm, Gujar, Nishida et al., 2011; Oyanedel, Binder, Kelemen, Petersen, Born, & Inostroza, 2014; Weber, Wang, Born, & Inostroze, 2014). Consequently, memory consolidation should prevent memory loss, and therefore differences between answers on the same day vs. the next day. Yet, researchers have shown that because of the limitation of memory capacity, sleep not only benefits the consolidation of certain memories but also is responsible for the loss of others (Wilhelm, Diekelmann, Molzow, Ayoub, & Born, 2011). In this regard, Wilhelm and colleagues (2011) tested whether specific memories are particularly consolidated during sleep. Their results revealed that sleep especially benefitted the consolidation of memory if respondents expected that their memory would be used in the future. The design of the current study assures that respondents are aware of their exercise to regularly reconstruct their previous day on the next day. Consequently, sleep should facilitate memory consolidation with respect to the information respondents need for recalling the day. In sum, the following implications can be drawn from this second line of argumentation: Episodic memory does not seem to suffer under the present study conditions, under the prediction that there is the same memory for emotional experiences on the same day vs. the next day. Thus, episodic happiness ratings should not distinguish between the evening of the same day vs. the ratings on the next day.

Overall, two contradicting but both reasonable lines of argumentation have been described, one predicting differences and one predicting no differences between retrospective judgments when done either on the same day vs. on the next day. To conclude, we sum up our research question for the current study:

Do respondents' DRM happiness ratings differ depending on whether they fill in the DRM on the evening of the same day vs. the next day?

The research question is investigated in a sample collected from the Innovation Sample of the German Socio-Economic Panel (GSOEP-IS; Richter & Schupp, 2015). The exact research design is described in detail in the following method section.

METHOD

Sample

The sample was collected after household interviews with 1,869 participants of the Innovation Sample of the German Socio-Economic Panel (GSOEP-IS; Richter & Schupp, 2015). The interviewers showed the participants a video about a study to measure Ger-

man citizens' daily happiness and a screencast of the survey app. The (German) video can be seen using the following link: https://vimeo.com/136258340. Then, the participants were asked whether they would participate in the study for a 50 Euro Amazon voucher in the case that they answer 6 of 7 DRM modules within one week.

Overall, 374 participants participated in the study and filled in all demographic questions and answered at least 1 DRM. They all received a notification to do the DRM at 9pm every evening for 7 days and had 24 hours to answer it. For our study only participants matter who did at least 1 DRM on the evening of the same day and at least 1 DRM on the next day (defined like this when the participant filled in the DRM after he or she slept which we were able to see in the reported DRM episodes). 79 participants match these criteria (289 participants did all DRMs on the evening of the same day and 4 did all DRMs on the next day). In total, we received 5,073 described DRM episodes and happiness ratings, 3,707 were filled in on the same day and 1366 were filled in on the next day. More detailed information on the demographics can be seen in table 7.1.

Table 7.1: Descriptive characteristics of the sample (*N* = 79)

	Mean / %
Age	37.1
Gender	59.5
(% female)	
Income	92.4
(1 = lower than average; 2 =	
(% lower than average)	
Education	39.2
(% Bachelor Degree or higher)	
Household Situation	34.2
(% alone)	
Job	77.2
(% yes)	
Migration Background	10.1
(% yes)	
Religious	70.9
(% yes)	

Note: The table displays the descriptive characteristics of the sample.

Materials

For the study the smartphone app "Happiness Analyzer" (Ludwigs & Erdtmann, 2017; see figure 1) was adapted to survey questions and the DRM. As mentioned before, the current study made use of Kahneman and colleagues' DRM (2004). However, several features differed from the original DRM version and were adapted for several reasons:

(i) The DRM was integrated into a smartphone application (original DRM version: paper-pencil format). By doing so, the potential of this kind of technology for future studies with flexible DRM timings can be best tested. In addition, filling in a DRM via a smartphone application already causes more time efficiency compared with its traditional completion.

(ii) In its original version, respondents were asked to rate every episode with respect to 12 different emotions. Yet, this poses a substantial burden on participants (Hendriks et al., 2016). Additionally, single-item measures on happiness were found to strongly correlate with multi-item measures of the same construct (Knabe, Rätzel, Schöb, & Weimann, 2010), implying that the former could be used as an alternative to the latter. Furthermore, it can be questioned how far individuals are generally capable of judging such a large variety of different emotions for every single episode, indicating that asking for many different emotions may result in answers of relatively low validity. Another crucial point is that there is no consensus about how to address various weightings when aggregating all these multiple emotions (White & Dolan 2009). In contrast, single-item measures on happiness require from the respondent to conduct this challenging aggregation process individually. In this way, the weighting problem can be avoided. In addition, positive and negative emotions could be found to be located on distinctive dimensions (Cacioppo & Bernston 1994), hampering the aggregating process additionally. All these disadvantages of multiple-item measures of happiness led us to the decision to use only a single item on perceived happiness instead of 12 affective items. Thus, in total, participants were asked to describe their full day from one midnight to another in episodes (e.g., lunch from 1-2 pm), indicating where every episodic activity took place, which persons were present and how respondents perceived their happiness doing so. Single-item happiness ratings were obtained by means of a smiley scale developed by Veenhoven in 2004 and episodic activities, locations and associated persons needed to be specified with the help of before quantified and already tested answering categories (Hendriks et al., 2016).

(iii) Contradictory to Kahneman et al.'s (2004) original DRM version, the current study made use of a longitudinal design. Hence, the DRM needed to be completed not only at one time but also at several times successively. The benefit of doing so consists in obtaining results that are in general more representative for the respondents' actual emotions during the reported episodes.

Additionally, demographic data were collected of both DRM groups (same day vs. next day) in the beginning of the app. The questionnaire comprised questions about the individual respondent's age, gender, income, education, household situation, job status, migration background and religion.

Procedure

After participants downloaded the app, they were shown a tutorial on the DRM. Then, they had to answer the personal questions. On the next day they were notified to do a DRM module at 9pm and had 24 hours to fill it in. Over the next six days, the same procedure was followed, and if they answered at least 6 of the 7 DRM modules, they were rewarded with a 50 Euro Amazon voucher code in the app. Through the app, they could examine their average happiness level development over the course of the one-week study, and their happiness level graphs were calculated for all different activities, locations, and social environments (see figure 7.1 for an illustration of the mentioned app).

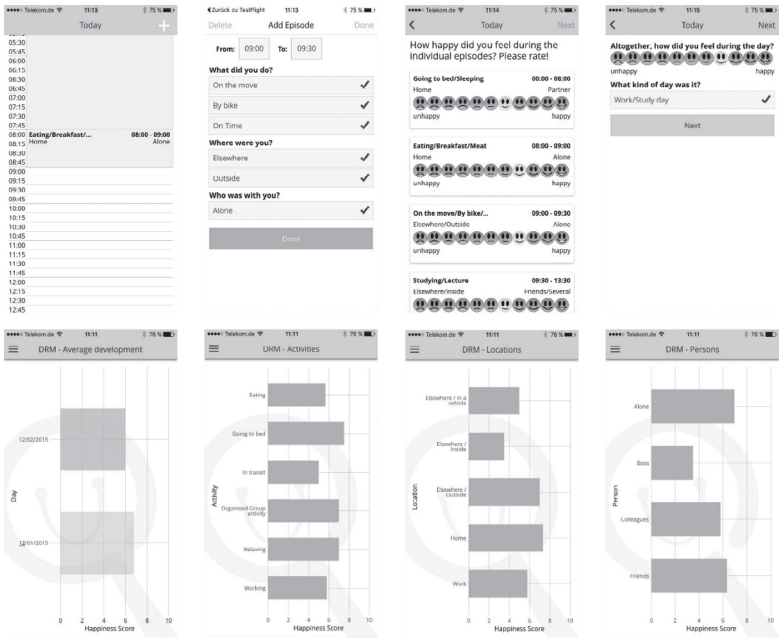

Figure 7.1: Exemplary app extract showing the integrated DRM, including items (row 1), individualized visual results feedback (row 2) and the design framework.

Design

A within-subjects design with two different settings (DRM filled in on the same day vs. DRM filled in on the next day) as independent variables was used. The dependent variable was the happiness rating for every episode.

RESULTS

In order to analyze if there is a difference between the happiness ratings in a DRM when filled in on the same day compared to when filled in on the next day, we calculated

repeated measures ANOVAs for the overall average happiness ratings in both settings and all average single ratings for all locations, social environments and activities in both settings. We find no significant effect between the two settings, neither in the overall average happiness ratings ($F = 0.04$; $p = .846$), nor in the average single ratings. The detailed results for every main location, social environment and activity rating that had at least 30 happiness ratings per setting is displayed in table 7.2 below. The various codes in this table are exactly the ones that the participants saw and defined when filling in the DRM.

Table 7.2: Comparison between the DRM happiness ratings when filled in on the same day or on the next day ($N = 79$).

	Same Day		Next Day			
	#Happiness Ratings	*M (SE)*	#Happiness Ratings	*M (SE)*	*F*-value	*p*
General Rating	3707	6.99 (0.12)	1366	7.00 (0.14)	0.04	.846
Location:						
Home	2403	7.02 (0.12)	873	6.97 (0.15)	0.26	.614
At Work	332	6.77 (0.23)	109	6.85 (0.22)	0.24	.630
Elsewhere	955	7.09 (0.14)	373	7.08 (0.16)	0.01	.932
Social Environment:						
Alone	1538	6.78 (0.13)	520	6.64 (0.16)	1.14	.289
Friend(s)	158	8.00 (0.31)	82	7.80 (0.33)	0.59	.451
Own Child(ren)	432	6.88 (0.21)	171	6.80 (0.30)	0.12	.730
Child(ren)	215	7.25 (0.31)	53	7.16 (0.41)	0.08	.784
Colleague(s)	226	6.70 (0.28)	73	6.62 (0.27)	0.17	.680
Partner	768	7.42 (0.19)	323	7.41 (0.22)	0.01	.953
Relatives	115	7.18 (0.42)	67	7.21 (0.52)	0.01	.964
Activities:						
Sleeping	531	7.14 (0.17)	255	7.09 (0.18)	0.14	.713
Getting ready	370	6.27 (0.18)	148	6.45 (0.18)	1.92	.171
On the move	544	6.77 (0.19)	196	6.85 (0.19)	0.19	.664
Working	351	6.73 (0.21)	116	6.70 (0.20)	0.04	.836
Eating	571	7.24 (0.14)	189	7.28 (0.16)	0.12	.726
Housework	332	6.58 (0.26)	127	6.39 (0.26)	0.85	.361
Relaxing	214	7.62 (0.21)	80	7.46 (0.25)	0.53	.471
Free time	193	7.84 (0.17)	71	7.65 (0.27)	0.94	.340
Leisure Media	309	7.09 (0.18)	98	7.12 (0.22)	0.05	.820

Note: The table displays the comparison between the happiness ratings when the DRM is filled in on the same day compared to when it is filled in on the next day within one person. As can be seen there are no significant differences between the two settings.

DISCUSSION

Most studies are aligned with the design of Kahneman and colleagues' (2004) original study by requiring respondents to reconstruct their previous day on the next day (Anusic et al., 2016; Berkhout et al., 2015; Bylsma et al., 2011; Doyle et al., 2015; Hendriks et al., 2016; John & Lang, 2015; Knabe et al., 2015; Schneider & Stone, 2016; Stone et al., 2006). Yet, when directly asked about their preferred DRM completion day time, most respondents favour the evening of the same day (about 70%), whereas 20% prefer the morning (Ludwigs & Erdtmann, 2017), indicating that current studies usually disregard respondents' preferences. However, especially for longitudinal studies, respondent motivation is of tremendous relevance for gaining large DRM datasets. Thus, why are respondents not allowed to use the DRM in a flexible way, either in the evening of the same day or on the next day, as they prefer? Reluctance to do so was generally reasonable since no study to date has investigated potentially differing answering patterns between respondents completing the DRM in the evening of the same day vs. on the next day. Therefore, we ran this study.

Our results do not show a significant effect between the two settings. This is in line with Conway's argumentation (2001) that the episodic memory does not suffer for about 24 hours and thus episodic happiness ratings should not distinguish between the evening of the same day vs. the ratings on the next day (see introduction).

But it is important to underline that our study is only quasi-experimental and that the sample size is limited. In order to investigate if there is really no significant effect between the two settings it is mandatory to run an experiment with a representative sample where participants are randomly assigned to either do the DRM on the evening of the same day or on the morning of the next day giving participants the same amount of time in both settings (e.g. 9pm until 11pm and 7am until 11am).

Taking this limitation into account, we suggest that researchers might consider (especially for longitudinal studies) to give participants free choice to do the DRM on the evening of the same day or on the next day if they want to reduce response burden in order to increase participation rates.

8 Can Happiness Apps Generate Nationally Representative Datasets? – A Case Study Collecting Data on People's Happiness Using the German Socio-Economic Panel

ABSTRACT

In the last few years, apps have become an important tool to collect data. Especially in the case of data on people's happiness, two projects have received substantial attention from both the media and the scientific world: "Track your happiness," from Kilingsworth & Gilbert (2010), and "Mappiness," from MacKerron (2012). Both happiness apps used the experience sampling method to ask people a few times per day how they feel, what they do, with whom to do it, and where. The collected data are then displayed for the participants in simple graphs to help them understand what makes them happy and what does not. Both studies have collected considerable data without giving participants any rewards. But quantity is not as important as quality with respect to data collection, and thus, understanding whether nationally representative datasets can be collected using such happiness apps is crucial. To address this question, we built a new happiness app and ran a case-study with over 4,000 participants of the innovation sample of the German Socio-Economic Panel (Richter & Schupp, 2015). Participants were informed that the app collects data on everyday happiness after a household interview and asked whether they would like to use the app. In the first year (2015), participants did not receive any reward, and in the second year (2016), participants received a 50 Euro Amazon voucher for their participation. The results showed that our happiness app cannot generate nationally representative datasets if it is not controlled that all demographic sub-groups have access to a smartphone, are highly motivated with a sufficient reward and data is collected with quota sampling.

Keywords: App Surveys, Representativity, Happiness, Experience Sampling Method, Day Reconstruction Method

INTRODUCTION

Online surveys are frequently used to collect data because they save both time for participants and resources for researchers. A problem with online surveys is that it can be difficult to secure data collection for intense studies. Thus, participants are unlikely to answer different online surveys on one day if they are just notified with an e-mail mainly because not everyone has an internet connection everywhere and not everyone reads e-mails multiple times a day. Smartphone applications could address this issue. Participants can be easily notified without continuous internet access to answer surveys at multiple points in time on their smartphone and it seems more likely that a notification will be seen by the participant compared to an e-mail. But a disadvantage of smartphone applications compared to online surveys via e-mail is that much more people have access to a computer compared to having a smartphone which raises questions about the possibility to collect nationally representative datasets using smartphone applications.

In the case of collecting data on people's happiness, it seems crucial to have better data than is typically collected. For instance, many researchers have argued that it is not enough to survey people about their happiness yearly in order to understand the mechanisms or factors that make people happier in certain situations (OECD, 2013). One alternative to standard survey methods that can be used to obtain better data is the experience sampling method (ESM; Csikzentmihalyi & Hunter, 2003), where participants are notified a few times per day to report how they feel, what they do, who they are with, and where. Unfortunately, it takes considerable resources for both participants and researchers to use this method using traditional paper and pencil methods or even online surveys. Thus, in 2010, Kilingsworth & Gilbert developed an app called "Track your happiness" to collect data on people's happiness using the ESM (Kilingsworth & Gilbert, 2010). Just a few months after publishing this app in different media channels, they were able to collect over 250,000 happiness ratings from over 5,000 participants in 83 different countries in the age range of 18 to 88 years (Kilingsworth & Gilbert, 2010). MacKerron (2012) developed a similar app called "Mappiness" to build a dataset in the United Kingdom. After half a year, this app collected 1.5 million happiness ratings from 32,000 participants and had 7,000 stable users (MacKerron, 2012). In conclusion, there is strong evidence that a high quantity of happiness ratings can be collected using happiness apps.

Although ESM provides detailed information about a person's experiences throughout the day, the method typically surveys participants just a few times (usually 4-6 times) a day about their current happiness and current activities. Thus, considerable gaps remain regarding people's activities, their social environment and their exact location. In order to give researchers the data basis that they call for to understand happiness and the underlying mechanisms, it is thus desirable to supplement the ESM with alternative

methods. The day reconstruction method (DRM; Kahneman et al., 2004) solves these issues by asking people to reconstruct their day in episodes for the previous day (e.g., breakfast with partner at home from 8-9am) and to then rate how happy they felt during these episodes. Ludwigs and Erdtmann developed an app in 2013 called "Happiness Analyzer" that combines the ESM and DRM (Ludwigs & Erdtmann, 2017). This tool has also been used in multiple studies to collect a large quantity of happiness ratings over a short time (e.g., Hendriks, Ludwigs & Veenhoven, 2016).

Although smartphone apps allow researchers to collect intense data on people's happiness, the quantity of data does not provide information about the quality of such data. One especially important concern is the national representativity of the sample. The greatest example for this is the "Literary-Digest-Disaster" in 1936, where the newspaper "The Literary Digest" asked all their 10 million readers to participate in a survey about the upcoming American presidential election and got 2.5 million responses but forecasted the election result incorrectly. George Gallup also did a survey using a quota sampling with just 50,000 participants and forecasted the election results correctly. To be able to make nationally representative claims out of a dataset (e.g., Germans became happier from 2015 to 2016), two main requirements need to be met: i) the sample needs to have a similar demographic structure to the population and ii) the response rates need to be in line with this demographic structure. In our example, we need to investigate a sample that has a similar demographic structure to the population and to determine whether different demographic groups are willing to download a survey app to collect data on their happiness in the same way and are using it on a frequent basis to realize good data quality.

To answer these questions, the "Happiness Analyzer" was used in two waves of the innovation sample of the German Socio-Economic Panel (GSOEP-IS; Richter & Schupp, 2015). Directly after a 45-minute interview, participants were asked whether they own a smartphone. If they did own a smartphone, a video of the app was presented to them. After they watched the video, the interviewer asked respondents whether they wanted to use the app. In the first wave in 2015, participants did not get any reward for using the app. In the second wave in 2016, participants were offered a 50 Euro Amazon voucher for using the app.

The next section (2) will explain the sample and the method in more detail. In the third section, we will report data on smartphone ownership, the participation rates and the participation quality. In the fourth section, we will discuss the results to answer the question whether happiness apps can generate nationally representative samples.

METHOD

Sample

Data was collected in cooperation with the GSOEP-IS, which is a nationally representative household panel where participants are interviewed to test new modules that can be integrated in the world-renowned German Socio-Economic Panel (Richter & Schupp, 2015). The first data collection included 2,135 participants of the GSOEP-IS that were interviewed between September 2015 and February 2016. The second data collection included 1,869 different participants of the GSOEP-IS that were interviewed between October 2015 and February 2016. For our study, we focus on three main demographic characteristics reported by respondents in annual CAPI interviews to compare their representativity to the whole German population: i) age; ii) gender; iii) employment status. Employment status is defined by people working full-time or part-time either in a corporate, public or non-profit job. The sample was controlled for national representativity by comparing it to other national statistics. For age, gender and employment status, statistics from the German Census 2011 (German Census, 2011) were used. As a national statistic for smartphone usage we refer to the Kantar World Panel (Kantar World Panel, 2015). To test the sample for significant differences we ran a Chi-Square test. The detailed demographic information of the samples can be seen in table 8.1.

Materials

For the study, four materials were used: i) A video to inform participants about the study; ii) the Happiness Analyzer; iii) graphical feedback; and iv) 50 Euro Amazon vouchers. All materials will be explained in detail in the following sections.

Video

Participants who owned an Android or iOS smartphone saw a video that consisted of two parts: i) A motivating part explaining that the study aims to collect data on people's happiness in everyday life to build a better data basis to inform researchers and politicians how to improve people's happiness; ii) a screencast explaining what the app exactly does and what they would have to do. The (German) video can be seen using the following link: https://vimeo.com/136258340

Table 8.1: Descriptive characteristics of the sample

Variable	First Sample N = 2135	Second Sample N = 1869	χ2(df, N); *p*	National Statistics
	%	%		%
Age			χ2 (1, 4004) = 5.04; *p* = .753	
17-19	*3.3*	*4.1*		Not available
20-29	*10.9*	*10.8*		12.1
30-39	*12.9*	*11.6*		11.8
40-49	*15.2*	*15.0*		14.2
50-59	*18.9*	*19.6*		*14.5*
60-69	*18.8*	*17.9*		*11.1*
70-79	*14.7*	*14.9*		*10.1*
80-89	*5.0*	*5.4*		Not available
90-99	*0.4*	*0.6*		Not available
Gender			χ2 (1, 4004) = 1.72; *p* = .189	
Male	*48.6*	*46.5*		*48.8*
Female	*51.4*	*53.5*		*51.2*
Employment Status			χ2 (1, 4004) = 1.89; *p* = .169	
Having A Job	*49.1*	*51.3*		*51.5*
No Job	*50.9*	*48.7*		*48.5*
Smartphone System			**χ2 (1, 2600) = 51.00; *p* < .001**	
Android	*40.7*	*41.7*		*37.0*
iOS (iPhone)	12.2	*14.4*		*11.0*
WindowsPhone	3.0	*2.7*		*4.0*
Blackberry	0.3	*0.2*		*0.5*
Do not know	21.4	*8.3*		Not available

Note: The table displays the descriptive characteristics of the two samples in percent. Significant differences (p < .05) are displayed bold. Significant differences were tested with a chi-square test. The only significant difference is the distribution of smartphone systems, which seems due to a different awareness of participants if their cellphone is a smartphone or not but not the real distribution of the different smartphone systems. The samples matched in most points to the general German population statistics although both samples are older.

The Happiness Analyzer

For this study, the Happiness Analyzer was used (Ludwigs & Erdtmann, 2017). After the app is downloaded, the Happiness Analyzer first asks participants to answer some subjective well-being questionnaires based on the OECD guidelines regarding how to measure subjective well-being (OECD, 2013) and then asks for some personal information (gender, age, etc.). On the next day, participants are notified to do four ESM modules between 8am and 8pm and one DRM module at 9pm. For the ESM modules, they have two hours to answer, and for the DRM modules, they have 24 hours. For the Happiness Analyzer, it takes about 20-30 seconds to answer an ESM module and about 7-10 minutes to answer a DRM module. Participants had to use the app for one week, equating to 28 ESM modules and 7 DRM modules over seven days. The total time required to answer the surveys, the ESM modules and the DRM modules is about 1.5 hours. Figure 8.1 displays the ESM and DRM modules:

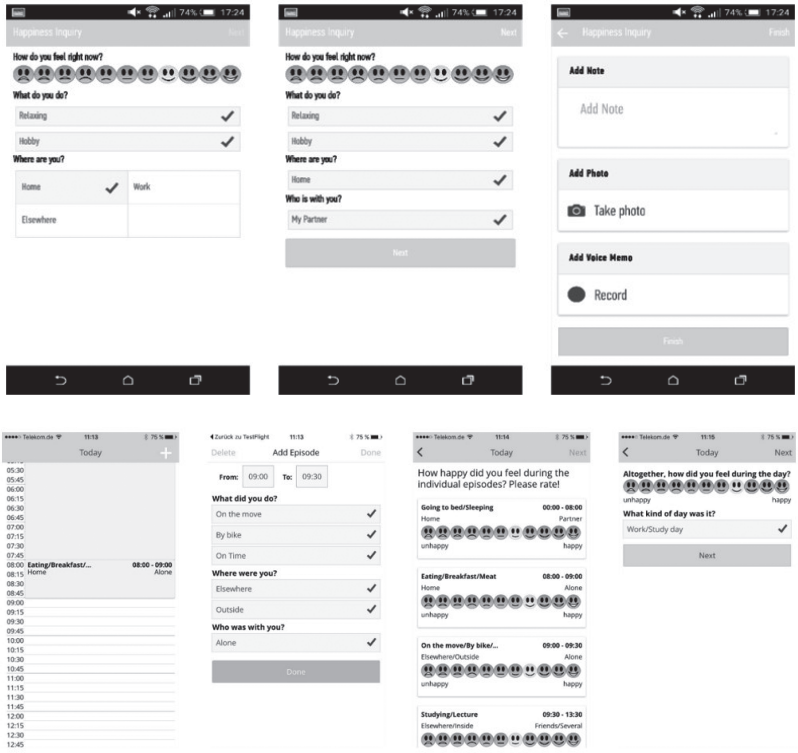

Figure 8.1: Shown is the ESM module (first row) and the DRM module (second row).

Graphical feedback:

As a reward, participants were able to build their own happiness profile. All happiness ratings were displayed in simple graphs showing how happy participants felt in which activities, social environments and locations. Figure 8.2 displays the graphical feedback:

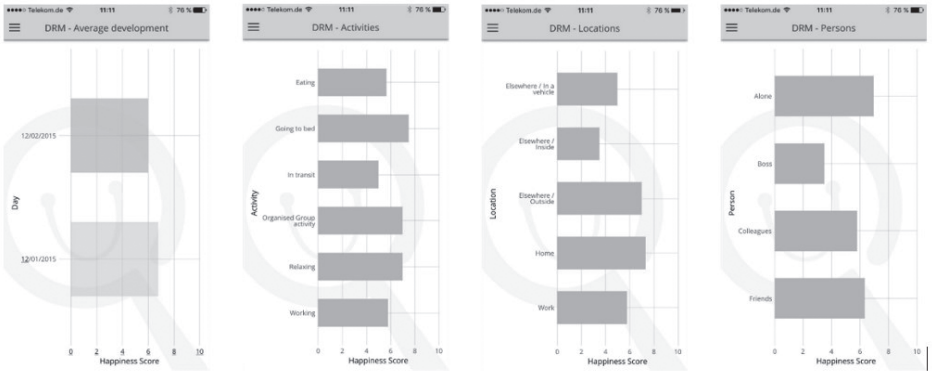

Figure 8.2: Shown is the graphical feedback that participants received.

50 Euro Amazon vouchers

In the second study, participants were able receive each a 50 Euro Amazon voucher if they answered at least 20 of the 28 ESM modules and at least 6 of 7 of the DRM modules. Whether their participation was successful was displayed on a "smiley screen" that showed a green and happy smiley when they answered all surveys, a yellow and neutral face if one DRM module or 5 ESM modules were missed and a red face if more than two DRM modules or more than 8 ESM modules were missed. If they answered at least 6 DRM modules and 20 ESM modules, a 50 Euro Amazon voucher code was displayed in the app after the last participation day. In the first study, participants did not receive any rewards. Figure 8.3 displays the smiley screen and Amazon voucher code:

Procedure

Participants first participated in a 45-minute household interview. Afterwards, they were asked if they own a smartphone and on which system it operates. If they said that they own an Android smartphone or an iPhone, the video was presented to them, and they were asked whether they would want to use the app. If they said that they want to use the app, they signed an informed consent letter and downloaded the app to use it for up to seven days. If participants had any technical issues or general questions, they were able to send an e-mail or call a hotline.

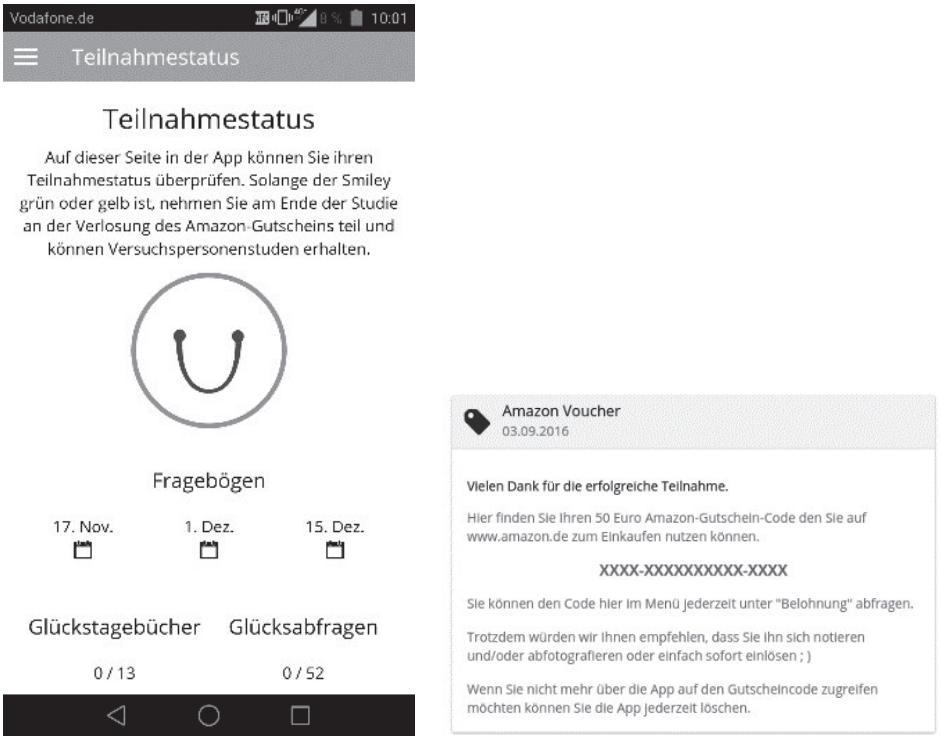

Figure 8.3: Shown is the "smiley screen" (first row) and the Amazon voucher (second row).

Design

A between-subjects design (first sample with no reward vs. second sample with reward) with age, gender and employment status as independent variables was used. The dependent variables were smartphone ownership, participation rates and participation quality in the two different periods of data collection.

RESULTS

To investigate if it is possible to generate a nationally representative dataset using a happiness app (regarding age, gender and employment status) or in our case the Happiness Analyzer, the participation differences (participation rate and participation quality) between the two different samples and between different demographic groups for each sample were calculated with Chi-Square tests. All tests were done once for the whole sample and once for a sample of smartphone owners. Smartphone ownership was defined by owning an Android or iOS Smartphone, which gave participants the option to participate in the study. Participation was defined by a participant downloading the app and answering the initial survey and one diary. A good participation quality was defined by at least 6 of 7 answered DRMs and at least 20 of 28 answered ESMs. All results are displayed in table 8.2, 8.3, 8.4, 8.5 and 8.6.

Table 2 shows that there is no gender bias in smartphone ownership but that nearly double as many people who are working have a smartphone compared to people who do not work and that people older than 50 years are less likely smartphone owners. Table 2 also shows that there was a significant increase in smartphone ownership between 2015 and 2016 ($\chi 2 = (1, 4004) = 4.37$; $p = .037$) but that just about 50% of the German population own a smartphone. Thus, we can conclude that probably every data collection with an app is not nationally representative if not all demographic groups get access to a smartphone for the data collection.

Table 8.2: Smartphone Ownership (Android or iOS)

Variable	First Sample N = 1129 (52.9%) %	Second Sample N = 1050 (56.2%) %	χ2 (1, 4004) = 4.37; *p* = .037
Age	**χ2 (1, 2135) = 646.64; *p* < .001**	**χ2 (1, 1869) = 652.90; *p* < .001**	
17-19	*95.8*	89.6	
20-29	*87.5*	*91.6*	
30-39	*79.6*	*88.0*	
40-49	*69.1*	*77.5*	
50-59	*58.9*	*59.7*	
60-69	*30.9*	*36.4*	
70-79	*15.3*	*15.8*	
80-89	*4.7*	*0.2*	
90-99	*0.0*	*0.0*	
Gender	χ2 (1, 2135) = 0.56; *p* = .454	χ2 (1, 1869) = 0.20; *p* = .654	
Male	*53.7*	*56.7*	
Female	*52.1*	*55.7*	
Employment Status	**χ2 (1, 2135) = 218.09; *p* < .001**	**χ2 (1, 1869) = 249.16; *p* < .001**	
Having A Job	*69.1*	*73.8*	
No Job	*37.2*	*37.6*	

Note: The table displays the descriptive characteristics of the two samples in percent. The general ownership rates are displayed in brackets in the first row in the second and third column. Significant differences (p < .05) are investigated with chi-square tests: i) For the difference between the first and the second sample (see first row, last column); ii) For the demographic differences within a sample (see second and third column). Significant differences are displayed bold. Smartphone ownership differs significantly between the two samples and according to Age and Employment status in both samples.

Table 3 and 4 show that participation rates increased significantly ($\chi2 = (1, 4004) = 254.25$; $p < .001$ & $\chi2 = (1, 2179) = 260.22$; $p < .001$) for all sub-groups (except of participants in the age of 70-79 in the whole sample) for the second sample by rewarding the participants with a 50 Euro Amazon-Voucher. The general participation rate in the first sample is quite low with 3.9% of the whole sample and 7.4% of the smartphone sample. The general participation rate in the second sample is higher with 20% of the whole sample and 35.6% of the smartphone sample. Unfortunately, the general participation is of course biased by smartphone ownership and even in the smartphone sample we see that in the first sample significantly more females participate ($\chi2 = (1, 1129) = 9.30$; $p = .002$) and in the second sample more younger people participate between 17 and 39 years participate ($\chi2 = (1, 1050) = 79.25$; $p < .001$). For employment status there is no significant difference in the participation rates in the smartphone sample.

Table 8.3: Participation Rates Whole Sample

Variable	First Sample N = 84 (3.9%) %	Second Sample N = 374 (20.0%) %	χ^2 (1, 4004) = 254.25; $p < .001$
Age	**χ^2 (1, 2135) = 42.61; $p < .001$**	**χ^2 (1, 1869) = 319.73; $p < .001$ .004**	
17-19	*8.5*	*44.2*	
20-29	*8.6*	*50.5*	
30-39	*6.5*	*36.9*	
40-49	*4.0*	*25.3*	
50-59	*4.7*	*17.2*	
60-69	*1.3*	*6.0*	
70-79	*2.8*	*1.4*	
80-89	*0.0*	*0.0*	
90-99	*0.0*	*0.0*	
Gender	**χ^2 (1, 2135) = 8.13; $p = .004$**	χ^2 (1, 1869) = 0.64; $p = .424$	
Male	*2.7*	*19.2*	
Female	*5.1*	*20.7*	
Employment Status	χ^2 (1, 2135) = 3.78; $p = .052$	**χ^2 (1, 1869) = 34.93; $p < .001$**	
Having A Job	*4.8*	*25.3*	
No Job	*3.1*	*14.4*	

Note: The table displays the descriptive characteristics of the two samples in percent. The general participation rates are displayed in brackets in the first row in the second and third column. Significant differences ($p < .05$) are investigated with chi-square tests: i) For the difference between the first and the second sample (see first row, last column); ii) For the demographic differences within a sample (see second and third column). Significant differences are displayed bold. Participation rates differ significantly between the two samples and according to Age in both samples, Gender in the first sample and Employment status in the second sample.

Table 8.4: Participation Rates Smartphone Sample

Variable	First Sample N = 84 (7.4%) %	Second Sample N = 374 (35.6%) %	χ^2 (1, 2179) = 260.22; $p < .001$
Age	χ^2 (1, 1129) = 5.66; $p = .580$	**χ^2 (1, 1050) = 79.25; $p < .001$**	
17-19	*8.8*	*49.3*	
20-29	*9.9*	*55.1*	
30-39	*8.2*	*41.9*	
40-49	*5.8*	*32.6*	
50-59	*8.0*	*28.8*	
60-69	*4.0*	*16.4*	
70-79	*6.3*	*9.1*	
80-89	*0.0*	*0.0*	
90-99	*0.0*	*0.0*	
Gender	**χ^2 (1, 1129) = 9.30; $p = .002$**	χ^2 (1, 1050) = 1.23; $p = .267$	
Male	*5.0*	*33.8*	
Female	*9.8*	*37.2*	
Employment Status	χ^2 (1,2135) = 0.87; $p = .351$	χ^2 (1, 1050) = 1.60; $p = .207$	
Having A Job	*6.9*	*34.3*	
No Job	*8.4*	*38.3*	

Note: The table displays the descriptive characteristics of the two samples in percent. The general participation rates are displayed in brackets in the first row in the second and third column. Significant differences ($p < .05$) are investigated with chi-square tests: i) For the difference between the first and the second sample (see first row, last column); ii) For the demographic differences within a sample (see second and third column). Significant differences are displayed bold. Participation rates differ significantly between the two samples and according to Age in the second sample and Gender in the first sample.

Table 5 and 6 display that the participation quality increased significantly ($\chi2 = (1, 4004) = 332.67$; $p < .001$ & $\chi2 = (1, 2179) = 337.67$; $p < .001$) for all sub-groups for the second sample by rewarding the participants with a 50 Euro Amazon-Voucher. The participation quality is quite low in the first sample with 1.1% of the whole sample and 2.1%, of the smartphone sample participating in a good quality, which means that 28.6% of the once starting, participate in a good quality in the first sample. The participation quality in the second sample is much higher with a response rate of 17.4% of the whole sample and 31.0% of the smartphone sample, which means 87.2% of the once starting, participate in a good quality in the second sample indicating that the reward amount was chosen right. In line with the participation rates we see again a bias for smartphone ownership and for age with significantly more people participating in the smartphone sample between 17 and 59 years compared to people being older ($\chi2 = (1, 1050) = 64.95$; $p < .001$).

Table 8.5: Participation Quality Whole Sample

Variable	**First Sample N = 24 (1.1%)** %	**Second Sample N = 326 (17.4%)** %	**$\chi2$ (1, 4004) = 332.67; $p < .001$**
Age	**$\chi2$ (1, 2135) = 16.12; $p = .041$**	**$\chi2$ (1, 1869) = 269.91; $p < .001$**	
17-19	*0.0*	*40.3*	
20-29	*3.0*	*44.1*	
30-39	*1.8*	*30.0*	
40-49	*0.6*	*23.5*	
50-59	*1.7*	*15.0*	
60-69	*0.5*	*4.8*	
70-79	*0.3*	*1.4*	
80-89	*0.0*	*0.0*	
90-99	*0.0*	*0.0*	
Gender	$\chi2$ (1, 2135) = 2.26; p = .133	$\chi2$ (1, 1869) = 2.00; $p = .157$	
Male	*0.8*	*16.1*	
Female	*1.5*	*18.6*	
Employment Status	$\chi2$ (1, 2135) = 2.99; p = .084	**$\chi2$ (1, 1869) = 27.15; $p < .001$**	
Having A Job	*1.5*	*21.9*	
No Job	*0.7*	*12.7*	

Note: The table displays the descriptive characteristics of the two samples in percent. The general participation quality rates are displayed in brackets in the first row in the second and third column. Significant differences ($p < .05$) are investigated with chi-square tests: i) For the difference between the first and the second sample (see first row, last column); ii) For the demographic differences within a sample (see second and third column). Significant differences are displayed bold. Participation quality rates differ significantly between the two samples and according to Age in both samples and Employment Status in the second sample.

Table 8.6: Participation Quality Smartphone Sample

Variable	First Sample N = 24 (2.1%) %	Second Sample N = 326 (31.0%) %	**χ2 (1, 2179) = 337.53; *p* < .001**
Age	χ2 (1,1129) = 5.87; p = .555	**χ2 (1, 1050) = 64.95; *p* < .001**	
17-19	*0.0*	*44.9*	
20-29	*3.5*	*48.1*	
30-39	*2.3*	*34.0*	
40-49	*0.9*	*30.3*	
50-59	*2.9*	*25.1*	
60-69	*1.6*	*13.1*	
70-79	*2.1*	*9.1*	
80-89	*0.0*	*0.0*	
90-99	*0.0*	*0.0*	
Gender	χ2 (1, 1129) = 2.51; *p* = .113	χ2 (1, 1050) = 3.05; *p* = .081	
Male	*1.4*	*28.4*	
Female	*2.8*	*33.4*	
Employment Status	χ2 (1, 1129) = 0.64; *p* = .800	χ2 (1, 1050) = 1.95; *p* = .162	
Having A Job	*2.2*	*29.7*	
No Job	*2.0*	*33.9*	

Note: The table displays the descriptive characteristics of the two samples in percent. The general participation quality rates are displayed in brackets in the first row in the second and third column. Significant differences ($p < .05$) are investigated with chi-square tests: i) For the difference between the first and the second sample (see first row, last column); ii) For the demographic differences within a sample (see second and third column). Significant differences are displayed bold. Participation quality rates differ significantly between the two samples and according to Age in the second sample.

Thus, we can conclude that our results show that datasets generated by a happiness app without giving every participant access to a smartphone are biased by smartphone ownership indicating more smartphone owners that work and that are younger. Participation rates and participation quality are highly limited in case participants are not rewarded and are in all cases biased by age indicating younger participants participating more likely and in a higher quality than older participants.

DISCUSSION

Summary

App-based surveys could be the future instrument to collect data for intense studies. Kilingsworth & Gilbert (2010) and MacKerron (2013) developed happiness apps to collect detailed data in intense studies on people's happiness and Ludwigs & Erdtmann developed an app to do so in even more detail (Ludwigs & Erdtmann, 2017). All the apps show that it is possible to collect a high quantity of data without rewarding participants. To investigate the quality of such data collection, a case study was conducted using the

GSOEP-IS. Participants saw a video about an app to track their happiness in everyday life and were then asked whether they want to use the app for one week. In the first data collection period, participants did not get a reward, whereas in the second data collection period, they got a 50 Euro Amazon voucher for using the app.

The results indicate that the participation rate and participation quality are significantly higher if people are sufficiently rewarded, with a general response rate of 17.4% of the whole sample and 31.0% of smartphone owners participating at high quality. Unfortunately, smartphone ownership is biased by employment status and age, indicating that it is not possible to generate a nationally representative dataset using a happiness app without giving all participants smartphone access. But even if participants do own a smartphone and are highly motivated by a sufficient reward, response rates are biased by age. Thus, we can conclude that it is not possible to generate nationally representative datasets using a happiness app without giving participants access to a smartphone, highly motivating them with a sufficient reward and running a quota sampling.

Methodological discussion

Different sampling techniques can have different effects on participation rates. In this study, participants were sampled in person by interviewers. In general, we think that this sampling technique should result in the highest possible participation rates, but it is of course biased by the interviewers' opinion about the study and modern technologies. By just letting the interviewers show the standardized video, we tried to minimize this bias. The participation rates for different sampling techniques should be investigated in future studies.

Collecting data on participant's happiness is a quite personal topic. Thus, simpler topics, such as people's time use, might result in higher participation rates. This study was very intense, as it included a survey, 28 ESM notifications and 7 DRM notifications. Thus, in simpler studies, participation rates might be higher.

This study is not an experiment where we offered rewards in a randomized way in one sample to compare the participation rates. Although it seems that the result would probably have been the same given the strong participation difference between the two samples a real experiment would have been the better design.

Other demographic variables would be important to look at to investigate national representativity in regard to other demographic factors. This will be possible once the data is available from the GSOEP-IS which will probably be the case in late 2018. Thus, readers should feel free to contact the first author for further analysis.

As the results are limited to a German sample, the findings should also be investigated in other nations.

Because more people will own a smartphone in the future the results might change over time so that it is important to replicate our study in the future.

In the first data collection period, interviewers were not briefed to download the app together with the participants. We think that this procedure reduced the final participation rates in the first data collection period because some participants may not have been able to download the app themselves.

A 50 Euro Amazon voucher seemed to be a very high payment, as it equated to an hourly payment of about 30 Euros. Future studies should investigate how participation rates differ when participants are paid more or less.

Conclusion

App-based surveys could be the future instrument to collect data for intense studies and this seems especially the case for the topic happiness. But so far, happiness apps are not able to generate nationally representative samples without using quota sampling. Thus, happiness researchers should ensure that all demographic sub-groups have access to a smartphone and are highly motivated and that they use quota sampling when generating a dataset with a happiness app.

Acknowledgements:

We would like to thank Alexander Kühn and Radek Kubicki for generating the video, Stephan Erdtmann for building the app and hosting the data collection, Lena Henning and the TNS Infratest employees for supporting the data collection, the National Aging Foundation for providing the funding for the 50 Euro Amazon vouchers and the German Socio-Economic Panel for letting us survey the participants of their innovation sample.

9 Concluding Remarks

INTRODUCTION

It is the goal of this last chapter to summarize the dissertation and outline the main conclusions, discuss limitations and implications for further research, outline implications for decision makers and end with an epilogue and some personal comments about the work on this dissertation.

MAIN CONCLUSIONS

First, we reviewed the academic literature and analyzed data in the World Database of Happiness to investigate which subjective well-being assessment methods exist, what the arguments for and against them are and how frequently they are used. We concluded that it is highly beneficial to get better data on people's subjective well-being by using general questionnaires, the Day Reconstruction Method and the Experience Sampling Method with the help of modern technologies.

Second, we outlined our approach to collect this data with the help of modern technology, explained the functionalities and the most recent evaluation feedback from a user feedback evaluation study.

Third, we showed that the tool can provide better data to understand subjective well-being in more detail and explain more variance of the differences between demographic groups.

Fourth, we showed that the tool seems to be beneficial and not detrimental for the user or in other words that participants who used the Happiness Analyzer report higher subjective well-being than before at least in the short-term.

Fifth, we showed that the timing of the data collection does not have a significant effect on the happiness ratings in the Day Reconstruction Method, indicating nearly same results if it is filled in on the evening of the same day or on the next day.

Sixth, we showed that it is possible to collect representative data with the Happiness Analyzer if participants are motivated by a sufficient reward and data is collected with quota sampling.

In sum, this dissertation provides a tool to collect better data on people's subjective well-being, to have more information what defines subjective well-being, maybe know better how subjective well-being is defined and maybe understand why subjective well-being is higher or lower in different groups. This can help researchers and decision makers to better understand in which situations a certain group has differences in their reported subjective well-being compared to another group. Representative databases with data collected with the Happiness Analyzer might help to better inform decision makers to help them to increase their chances to make efficient decisions to improve subjective well-being and to learn continuously from a more detailed evaluation of these decisions.

LIMITATIONS, CHALLENGES & SUGGESTION FOR FUTURE RESEARCH

However, as is the case for nearly every dissertation, all our findings for all the different chapters raise some topics for discussion and challenges for further research.

In our review chapter, we only reviewed how frequently different subjective well-being assessments are used based on the World Database of Happiness. It would be beneficial to do another review on this topic with a broader definition of subjective well-being than the one used for the World Database of Happiness. In general, we do not feel as though we fully solved the puzzle to find the one subjective well-being definition that the whole field agrees on, and we think that it is a very crucial research task for the next few years to have a basis to define, develop and evaluate accurate subjective well-being measures. In general, we realized that there is a lack of detailed psychometric research in subjective well-being research, and we would like to contribute more to this field in the future.

For the method itself, we still do not fully meet the requirement of mixing objective and subjective data. We were successful in collecting location data and network data and adding other objective data sources, but it seems crucial to collect valid biological data, for example, on people's heart rate variability, directly with a smartphone and the

Happiness Analyzer app. Collecting biological data directly with the smartphone would give us the option to collect representative data with manageable costs while collecting biological data in lab studies, for example on people's cortisol levels in their hair or in neuroimaging studies, would take too many resources from both researchers and participants. Biological information would be another source in a triangulation multimethod-approach to have a better database to understand what defines subjective well-being, how it is defined and maybe even why it is different between certain groups to come closer to understand effects related to subjective well-being.

According to our first evaluation study comparing migrants' and locals' subjective well-being and their daily activities, we should underline that this is just one explorative example, where we showed that it helps to explain more variance and understand a subjective well-being puzzle better by collecting data with the Happiness Analyzer. In the future, it seems crucial to show additional similar examples with more representative datasets.

According to our second evaluation study investigating the effect of using the Happiness Analyzer, we need to consider that neither study had a control group that only completed one questionnaire at every point in time. Hence, we cannot say that the effects that we found are not due to other reasons. In addition, both samples consisted of (young) psychology students and were not representative. A representative (long-term) study, with a control group and many different sub-groups, would help provide stable findings on the effect of the Happiness Analyzer.

According to our third evaluation study showing that there are no differences in the happiness ratings depending on if the DRM is filled in on the evening of the same day or on the next day, we need to consider that the study was not an experimental study with a representative sample. We plan to run such an experimental study in the future to determine whether we can replicate our findings.

According to our fourth evaluation study investigating participation rates of the Happiness Analyzer in a nationally representative panel study, we found that it is not possible to create nationally representative datasets with the Happiness Analyzer without having more funding. Additionally, we do not have information thus far how high the participation rates would be in the long term and how high the rewards would need to be to build a nationally representative long-term panel.

A more general limitation is that we cannot prove that every subjective well-being research puzzle or every subjective well-being decision needs a database that matches the OECD gold standard. For example, it seems sufficient to only collect data with general retrospective questions such as "All things considered, how satisfied are you with your life as a whole nowadays?" (ESS; 2013) if the only objective is to compare which of two groups (e.g. two countries) is the happier one or to investigate if there was a general change over time. This data would give information on what changed and knowledge

in which direction it changed but it would be difficult without other data sources to understand why it changed. These other data sources could be further modules of the OECD guidelines such as other questionnaires, the ESM and especially the DRM. But other data sources could as well be other indicators for example social indicators from household surveys such as changes in marital status or economic indicators such as GDP. In this dissertation we were able to show that it can be beneficial to collect better data specifically on subjective well-being to better understand subjective well-being, but we need to state that we do not have proof from different applied studies so far that this database would be sufficient to fully understand subjective well-being in order to be able to evaluate every subjective well-being intervention sufficiently in order to enable a continuous learning.

Figure 9.1 summarizes the limitations, challenges and suggestion for future research.

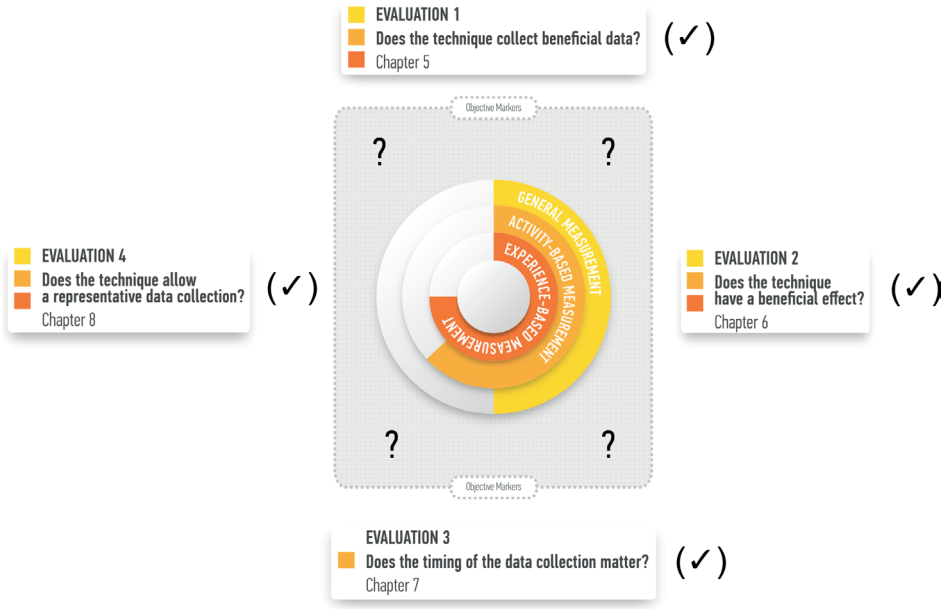

Figure 9.1: Displayed is the onion model and the key questions of the four evaluation studies. The check-marks in brackets display that we were able to investigate the research question but that all studies have limitations and future research is needed. The question marks in the grey zone display that we did not investigate the connection between the subjective and the objective markers in a triangulation approach which is an objective for future research.

In general, we know that we can collect better data on people's subjective well-being with the help of the Happiness Analyzer. But we are not sure at the moment how many researchers would really conduct intense studies using the Happiness Analyzer because of two reasons. First, we are not sure how many researchers have research puzzles that need this detailed subjective well-being data thus far as there are many observations

that can be made with mixing existing datasets on general subjective well-being and other social indicators as well. Second, it is not a simple tool that every researcher could use because it needs to be slightly adapted for nearly every study and funding is needed to reward participants for most samples. Thus, we do not think that the Happiness Analyzer will increase the number of studies that measure subjective well-being according to the OECD gold standard. Ultimately, it is a tool that reduces the effort for collecting data according to this standard, but it does not reduce the effort enough in that many researchers would get the resources to match the gold standard without having a specific reason why they need to match the gold standard.

IMPLICATIONS FOR DECISION MAKERS

The dissertation shows the benefits of collecting better data on people's subjective well-being and gives a tool to collect these data in a more efficient way. Based on this technique long-term data collections that provide a basis to evaluate decisions with the goal to improve subjective well-being better are manageable and would enable continuous learning.

For example, for a mayor of a city it would be beneficial to have this kind of data to be able to make decisions on which interventions should be funded based on the subjective well-being ratings of his or her citizens and not based on top-down decisions by politicians. A mayor could always monitor how different groups feel and compare them to others to figure out which quality of life indicators need to be improved for certain groups to increase subjective well-being. Because data would be collected continuously he or she could learn from the subjective well-being ratings of his or her citizens and make better decisions in the mid- and long-term.

Another example is a manager of a company who realized that it is crucial for his or her business to have happy employees (for example in a service business such as a hotel chain). This manager would be able to decide based on the employees' subjective well-being ratings and could learn from their continuous feedback to increase the likelihood for successful decisions that would improve the employees' subjective well-being in the mid- and long-term.

The third example is the therapist or coach that we already used in the beginning of this dissertation. This therapist would use the Happiness Analyzer to track his or her client's subjective well-being in detail to pick interventions based on the client's feedback, evaluate them based on the client's feedback and learn from the client's feedback when picking the next intervention.

For these applications it is important that decision makers are brave enough to: i) collect feedback from others and take this into account; ii) value the feedback on their

subjective well-being as the basis for their decisions instead of other indicators such as only health indicators or economic indicators which can only be correlates of subjective well-being but not indicators of subjective well-being.

EPILOGUE

In 2013, Stephan Erdtmann and I began to work on the Happiness Analyzer at the end of our Master studies after we read numerous research papers on subjective well-being and the OECD guidelines for measuring subjective well-being. First, it was just an idea to support this meaningful research, but after some time and after discussing the idea with many different scholars, we realized that we should pursue this idea and try to develop and evaluate it on an academic level. Our ambition was to provide a foundation for a way to collect better data on people's subjective well-being, which can result in datasets that can give researchers and practitioners the opportunity to better understand the differences between groups and the effect of interventions. When we set this plan, we had no idea how much work, funding and especially time it might take to realize it, but nearly five years later, we have been able to make it happen. Overall, the tool has been successfully developed, and we feel like we investigated the most important factors that can influence the data collection and the data being collected with the Happiness Analyzer.

Now, our challenge is to motivate researchers, funding institutions, mayors, managers or therapists to use our tool to build long-term panels that would give a basis to better understand subjective well-being.

References

Ackoff, R.L. (1989). From data to wisdom. *Journal of Applied Systems Analysis, 16*, 3–9

Aknin, L. B., Sandstrom, G. M., Dunn, E. W., & Norton, M. I. (2011). Investing in others: Prosocial spending for (pro) social change. In: R. Biswas-Diener (Ed.), *Positive psychology as social change (pp. 219-234*). Springer Netherlands.

Aksel, S¸ Gün, Z., Irmak, T. Y., & Cengelci, B. (2007). Migration and psychological status of adolescents in Turkey. *Adolescence, 42*, 589–602.

Allianz Deutschland AG & RWI (Rheinisch-Westfälisches Institut für Wirtschaftsforschung e.V.). (2011). Depression – Wie die Krankheit unsere Seele belastet. Munich, Germany: Lohse Druckgesellschaft. Retrieved from: http://www.rwiessen.de/media/content/pages/publikationen/sonstige/Allianz-Report-Depression.pdf (lastly retrieved on: 2016-02-17)

Andrews, F.M. & Withey, S.B. (1976). *Social Indicators of Well-being: Americans' Perceptions of Life Quality.* New York, USA: Plenum Press.

Anusic, I., Lucas, R. E., & Donnellan, M. B. (2016). The Validity of the Day Reconstruction Method in the German Socio-economic Panel Study. *Social Indicators Research, 1-20*.

Appleton, S., & Song, L. (2008). Life satisfaction in urban China: Components and determinants. *World Development, 36*, 2325–2340.

Ayuso-Mateos, J. L., Miret, M., Caballero, F. F., Olaya, B., Haro, J. M., Kowal, P., & Chatterji, S. (2013). Multi-country evaluation of affective experience: validation of an abbreviated version of the day reconstruction method in seven countries. *PloS One, 8(4)*, 1-8.

Bacon, N., Brophy, M., Mguni, N., Mulgan, G., & Shandro, A. (2010). *The State of Happiness: Can public policy shape people's wellbeing and resilience?* London, UK: The Young Foundation.

Bakker, A., M. Burger, P. van Haren, W. Oerlemans, and R. Veenhoven. 2016. "Happiness Raised by Raising Awareness: Effect of Happiness Using the Happiness Indicator." *EHERO Working Paper.*

Bakker, A. B., Demerouti, E., Oerlemans, W., & Sonnentag, S. (2013). Workaholism and daily recovery: A day reconstruction study of leisure activities. *Journal of Organizational Behavior, 34(1)*, 87-107.

Bartram, D. (2013). Happiness and 'economic migration': A comparison of Eastern European migrants and stayers. *Migration Studies, 1*, 156–175.

Beaman, A. (2010). *The role of emotional awareness, and positive and negative social support in predicting well-being in recent retirees*. ProQuest Information & Learning, US.

Berk, R. A. (1979). The construction of rating instruments for faculty evaluation: A review of methodological issues. *The Journal of Higher Education, 50(5),* 650-669.

Berkhout, J. J., Helmich, E., Teunissen, P. W., Berg, J. W., Vleuten, C. P., & Jaarsma, A. D. C. (2015). Exploring the factors influencing clinical students' self-regulated learning. *Medical Education, 49(6),* 589-600.

Bjørnskov, C. (2010). How comparable are the Gallup World Poll life satisfaction data? *Journal of Happiness Studies, 11(1)*, 41-60.

Buttel, F.H. & Martinson, O.B. (1977). Ideology and social indicators of the quality of life. *Social Indicators Research, 4(1)*, 353-369.

Bradburn, N. M. (1969). *The structure of psychological well-being.* Chicago, USA: Aldine Publishing.

Bradburn, N. M., & Caplovitz, D. (1965). *Reports on happiness: A pilot study of behavior related to mental health.* Chicago, USA: Aldine Publishing.

Brown, K. W. & Ryan, R. M. (2003). The benefits of being present: mindfulness and its role in psychological well-being. *Journal of personality and social psychology, 84*, 822-848.

Cacioppo, J. T., & Bernston, G. G. (1994). Relationship between attitudes and evaluative space: A critical review, with emphasis on the separability of positive and negative substrates. *Psychological Bulletin, 115*, 401–423.

Bruni, L. (2010). The happiness of sociality. Economics and eudaimonia: A necessary encounter. *Rationality and Society, 22(4)*, 383-406.

Busseri, M. A., & Sadava, S. W. (2011). A review of the tripartite structure of subjective well-being: Implications for conceptualization, operationalization, analysis, and synthesis. *Personality and Social Psychology Review, 5(3),* 290-314.

Bylsma, L. M., Taylor-Clift, A., & Rottenberg, J. (2011). Emotional reactivity to daily events in major and minor depression. *Journal of Abnormal Psychology, 120(1)*, 155-167.

Cameron, D. (2006). General well-being speech. Retrieved from: https://toryspeeches.files.wordpress.com/2013/11/david-cameron-general-well-being-speech.pdf (lastly retrieved on: 2016-02-17)

Chen, J. (2011). Internal migration and health: Re-examining the healthy migrant phenomenon in China. *Social Science and Medicine, 72*, 1294–1301.

Cheng, Z., Wang, H., & Smyth, R. (2014). Happiness and job satisfaction in urban China: A comparative study of two generations of migrants and urban locals. Urban Studies, 51, 2160–2184. Cohen, J. (1992). Quantitative methods in psychology: A power primer. *Psychological Bulletin, 112*, 155–159.

Clark, A. E., E. Diener, Y. Georgellis, and R. E. Lucas. 2008. Lags and Leads in Life Satisfaction: A Test of the Baseline Hypothesis. *The Economic Journal, 118 (529*), 222–243.

Clark, A., Georgellis, Y., & Sanfey, P. (1998). Job satisfaction, wage changes and quits: Evidence from Germany. *Research in Labor Economics, 17*, 95-121.

Clark, E. & Senik, C. (2011). Is happiness different from flourishing? Cross-country evidence from the ESS. *Revue d'économie politique, 121(1),* 17-34.

Commission of the European Communities. (2009). GDP and beyond: Measuring progress in a changing world. Communication from the Commission to the Council and the European Parliament. Retrieved from: http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52009DC0433&from=EN (lastly retrieved on: 2016-02-17)

Conner-Christensen, T. (2003). *Experience-sampling procedures: Are they probes to autonoetic awareness.* Dissertation Abstracts International: Section B: The Sciences and Engineering.

Constantinople, A. (1967). Perceived Instrumentality of the College as a Measure of Attitudes toward College. *Journal of Personality and Social Psychology, 5(2),* 196-201.

Conway, M. A. (2001). Sensory-perceptual episodic memory and its context: autobiographical memory. *Philosophical Transactions: Biological Sciences, 356(1413),* 1375-1384.

Cropanzano, R., & Wright, T. A. (1999). A 5-year study of change in the relationship between well-being and job performance. *Consulting Psychology Journal: Practice and Research, 51(4)*, 252-265.

Cunningham, M. R., Steinberg, J., & Grev, R. (1980). Wanting to and having to help: Separate motivations for positive mood and guilt-induced helping. *Journal of Personality and Social Psychology, 38(2)*, 181-192.

Csikszentmihalyi, M., and J. Hunter (2003). "Happiness in Everyday Life: The Uses of Experience Sampling." *Journal of Happiness Studies, 4 (2),* 185–199.

Csikszentmihalyi, M., & Larson, R. (2014). *Validity and reliability of the experience-sampling method.* In: M. Csikszentmihalyi (Ed.), Flow and the Foundations of Positive Psychology (Chapter 3, pp. 35-54). Dordrecht, the Netherlands: Springer.

Csikszentmihalyi, M., Larson, R., & Prescott, S. (1977). The Ecology of Adolescent Activity and Experience. *Journal of Youth and Adolescence, 6(3),* 281-294.

Daly, M., Delaney, L., Doran, P. P., & MacLachlan, M. (2011). The role of awakening cortisol and psychological distress in diurnal variations in affect: A day reconstruction study. *Emotion, 11(3)*, 524-532.

Danner, D. D., Snowdon, D. A. & Friesen, W. V. (2001). Positive emotions in early life and longevity: Findings from the nun study. *Journal of Personality and Social Psychology, 80 (5),* 804-813.

Dardis, R., Soberon-Ferrer, H., & Patro, D. (1994). Analysis of leisure expenditures in the United States. *Journal of Leisure Research, 26,* 309–321.

Denzin, N. K. (1978). *The research act: A theoretical orientation to sociological methods.* McGraw-Hill, New York.

DGPPN (Deutsche Gesellschaft für Psychiatrie und Psychotherapie, Psychosomatik und Nervenheilkunde). (2013). Gemeinsam für psychische Gesundheit: Erwartungen der DGPPN an die Gesundheitspolitik (Brochure). Retrieved from: https://www.dgppn.de/fileadmin/user_upload/_medien/download/pdf/Broschüren/DGP_Broschuere_Politik_2013_gesamt-Einzelseiten.pdf (lastly retrieved on: 2016-02-17)

Diekelmann, S. & Born, J. (2010). The memory function of sleep. Nature Reviews *Neuroscience, 11(2),* 114-126.

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. (2012). New findings and future directions for subjective well-being research. *American Psychologist, 67(8),* 590-597.

Diener, E., & Chan, M. Y. (2011). Happy people live longer: Subjective well-being contributes to health and longevity. *Applied Psychology: Health and Well-Being, 3(1),* 1-43.

Diener, E. D., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment, 49*, 71–75.

Diener, E., Inglehart, R., & Tay, L. (2013). Theory and validity of life satisfaction scales. *Social Indicators Research, 112*, 497–527.

Diener, E., & Larsen, R.J. (1984). Temporal stability and cross situational consistency of affective, behavioral, and cognitive responses. *Journal of Personality and Social Psychology, 47(4)*, 580–592.

Diener, E., Lucas, R. E., Oishi, S., & Suh, E. M. (2002). Looking up and looking down: Weighting good and bad information in life satisfaction judgments. *Personality and Social Psychology Bulletin, 28(4),* 437-445.

Diener, E., Lucas, R. E., Schimmack, U., & Helliwell, J. F. (2009). *Well-being for public policy.* Oxford, England: University Press.

Diener, E., Lucas, R. E., & Scollon, C. N. (2006). Beyond the hedonic treadmill: revising the adaptation theory of well-being. *American Psychologist, 61*(4), 305-314.

Diener, E., Nickerson, C., Lucas, R. E., & Sandvik, E. (2002). Dispositional affect and job outcomes. *Social Indicators Research, 59(3),* 229-259.

Diener, E., Sandvik, E., Pavot, W., & Gallagher, D. (1991). Response artifacts in the measurement of subjective well-being. *Social Indicators Research, 24(1),* 35–56.

Diener, E., Scollon, C. N., & Lucas, R. E. (2003). The evolving concept of subjective well-being: The multifaceted nature of happiness. *Advances in Cell Aging and Gerontology, 15*, 187-219.

Diener, E. & Seligman, M. E. (2004). Beyond money toward an economy of well-being. *Psychological Science in the Public Interest, 5(1),* 1-31.

Diener, E., Suh, E., Lucas, R. E., & Smith, H. L. (1999). Subjective well-being: Three decades of progress. *Psychological Bulletin, 125(2),* 276-302.

Diener, E., Suh, E., & Oishi, S. (1997). Recent findings on subjective well-being. *Indian Journal of Clinical Psychology, 24*, 25-41.

Diener, E. & Tay, L. (2014). Review of the Day Reconstruction Method (DRM). *Social Indicators Research, 116,* 255-267.

Diener, E., Wirtz, D., Tov, W., Kim-Prieto, C., Choi, D. W., Oishi, S., & Biswas-Diener, R. (2010). New well-being measures: Short scales to assess flourishing and positive and negative feelings. *Social Indicators Research, 97(2),* 143-156.

Di Tella, R. & MacCulloch, R. (2008). Gross national happiness as an answer to the Easterlin Paradox? *Journal of Development Economics, 86(1),* 22-42.

DIW (German Institute for Economic Research) (2016). Socio-Economic Panel (Website). Retrieved from: http://www.diw.de/en/soep (lastly retrieved on: 2016-03-22)

Dockray, S., Grant, N., Stone, A. A., Kahneman, D., Wardle, J., & Steptoe, A. (2010). A comparison of affect ratings obtained with ecological momentary assessment and the day reconstruction method. *Social Indicators Research, 99(2),* 269-283.

Dolan, P., Peasgood, T., & White, M. (2006). *Review of research on the influences on personal well-being and application to policy making*. London, England: DEFRA.

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(1),* 94-122.

Dolan, P., & White, M. P. (2007). How can measures of subjective well-being be used to inform public policy? *Perspectives on Psychological Science, 2(1),* 71-85.

Doyle, O., Delaney, L., O'Farrelly, C., Fitzpatrick, N., & Daly, M. (2015). *Can Early Intervention Improve Maternal Well-being? Evidence from a Randomized Controlled Trial* (Working Paper, No. 2015-015). Retrieved from: https://ideas.repec.org/p/hka/wpaper/2015-015.html (lastly retrieved on: 2016-03-30)

Durand, M. (2015). The OECD Better Life Initiative: How's Life? and the Measurement of Well-Being. *Review of Income and Wealth, 61(1),* 4-17.

Easterlin, R. A. (2003). Explaining happiness. *Proceedings of the National Academy of Sciences, 100(19)*, 11176-11183.

EC (European Commission). (2016). *Eurobarometer surveys* (Website). Retrieved from: http://ec.europa.eu/public_opinion/index_en.htm (lastly retrieved on: 2016-03-22)

Edgeworth, F. Y. (1881). *Mathematical psychics: An essay on the application of mathematics to the moral sciences.* London, U.K.: Kegan Paul; reprinted, New York: M. Kelly, 1967.

Egloff, B., Tausch, A., Kohlmann, C. W., & Krohne, H. W. (1995). Relationships between time of day, day of the week, and positive mood: Exploring the role of the mood measure. *Motivation and Emotion, 19(2),* 99-110.

Ehrhardt, J. J., Saris, W. E., & Veenhoven, R. (2000). Stability of life-satisfaction over time. *Journal of Happiness Studies, 1(2),* 177-205.

Eid, M., & Diener, E. (1999). Intraindividual variability in affect: Reliability, validity, and personality correlates. *Journal of Personality and Social Psychology, 76(4),* 662-676.

Eid, M. & Diener, E. (2004). Global judgments of subjective well-being: Situational variability and long-term stability. *Social Indicators Research, 65*, 245-277.

Ek, E., Koiranen, M., Raatikka, V. P., Järvelin, M. R., & Taanila, A. (2008). Psychosocial factors as mediators between migration and subjective well-being among young Finnish adults. *Social Science and Medicine, 66,* 1545–1556.

Else-Quest, N. M., Hyde, J. S., & DeLamater, J. D. (2005). Context counts: Long-term sequelae of premarital intercourse or abstinence. *Journal of Sex Research, 42(2),* 102-112.

ESS Round 6: European Social Survey (2013): ESS-6 2012 Documentation Report. Edition 1.3. Bergen, European Social Survey Data Archive, Norwegian Social Science Data Services.

Fernández-Dols, J. M., & Ruiz-Belda, M. A. (1995). Are smiles a sign of happiness? Gold medal winners at the Olympic Games. *Journal of Personality and Social Psychology, 69(6),* 1113-1119.

Ferrer-i-Carbonell, A., & Frijters, P. (2004). How important is methodology for the estimates of the determinants of happiness? *The Economic Journal, 114*, 641–659.

Finstad, K. (2010). Response interpolation and scale sensitivity: Evidence against 5-point scales. *Journal of Usability Studies, 5(3),* 104-110.

Flügel, J. C. (1925). A quantitative study of feeling and emotion in everyday life. British Journal of Psychology. *General Section, 15(4),* 318-355.

Fordyce, M. W. (1977). Development of a program to increase personal happiness. *Journal of Counseling Psychology, 24(6),* 511-521.

Fozdar, F., & Torezani, S. (2008). Discrimination and well-being: Perceptions of refugees in Western Australia. *International Migration Review, 42*, 30–63.

Fredrickson, B. L. (2001). The role of positive emotions in positive psychology: The broaden-and-build theory of positive emotions. *American Psychologist, 56,* 218–226.

Frey, B. S., & Stutzer, A. (2002). What can economists learn from happiness research? *Journal of Economic Literature, 40(2),* 402-435.

Friedman, W. J., & Kemp, S. (1998). The effects of elapsed time and retrieval on young children's judgments of the temporal distances of past events. *Cognitive Development, 13(3),* 335-367.

Frijters, P. (2000). Do individuals try to maximize general satisfaction? *Journal of Economic Psychology, 21(3),* 281-304.

Frijters, P., Haisken-DeNew, J. P., & Shields, M. A. (2004). Money does matter! Evidence from increasing real income and life satisfaction in East Germany following reunification. *American Economic Review, 94,* 730–740.

Fuchs-Schundeln, N., & Schundeln, M. (2009). Who stays, who goes, who returns? East-West migration within Germany since reunification. *Economics of Transition, 17*, 703–738.

Fujita, F., & Diener, E. (2005). Life satisfaction set point: Stability and change. Journal of *Personality and Social Psychology, 88(1),* 158–164.

Galin, D. (1994). The structure of awareness: Contemporary applications of William James' forgotten concept of 'the fringe'.. *Journal Of Mind And Behavior, 15(4),* 375-401.

George, J. M. (1991). State or trait: Effects of positive mood on prosocial behaviors at work. *Journal of Applied Psychology, 76(2),* 299-307.

German Census (2011). Retrieved from: https://www.destatis.de/DE/Methoden/Zensus_/Zensus.html. (lastly retrieved on: 2017-10-20)

German Statistical Office (2015). Retrieved from: https://www.destatis.de/DE/PresseService/Presse/Pressemitteilungen/2015/05/PD15_172_631.html (lastly retrieved on: 2017-10-20)

Gesis. (2012). Zum Gallup World Poll. Retrieved from: http://www.gesis.org/fileadmin/upload/dienstleistung/veranstaltungen_fortbildungen/workshops/Infotext_GALLUP_WS_040912.pdf (lastly retrieved on: 2016-03-22)

Goldings, H. J. (1954). On the avowal and projection of happiness. *Journal of Personality, 23(1),* 30-47.

Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the big five personality domains. *Journal of Research in Personality, 37,* 504–528.

Gruber, J., Mauss, I. B. & Tamir, M. (2011). A dark side of happiness? How, when, and why happiness is not always good. *Perspectives on Psychological Science, 6,* 222-233.

Haas, O. & Janssen, B. (2012). Happiness steckt an. *Personalwirtschaft, 42(8),* 56-57.

Harker, L., & Keltner, D. (2001). Expressions of positive emotions in women's college yearbook pictures and their relationship to personality and life outcomes across adulthood. *Journal of Personality and Social Psychology, 80(1),* 112–124.

Harter, J. K., Schmidt, F. L., Asplund, J., & Kilham, E. A. (2010). Causal impact of employee work perceptions on the bottom line of organizations. *Perspectives on Psychological Science, 5,* 378 –389.

Hayne, H., & Imuta, K. (2011). Episodic memory in 3-and 4-year-old children. Developmental *Psychobiology, 53(3),* 317-322.

Hektner, J. M., Graham, C. (2005). The economics of happiness – Insights on globalization from a novel approach. *World Economics, 6(3),* 41-55.

Hektner, J. M., Schmidt, J. A., & Csikszentmihalyi, M. (2007). *Psychometrics of ESM Data*. In J. M. Hektner, J. A. Schmidt, & M. Csikszentmihalyi (Eds.), Experience Sampling Method: Measuring the Quality of Everyday Life (Chapter 6, pp. 103-122). Thousand Oaks, USA: Sage.

Helliwell, J. F., & Wang, S. (2014). Weekends and subjective well-being. *Social indicators research, 116,* 389-407.

Hendriks, M., Ludwigs, K., & Veenhoven, R. (2016). Why are Locals Happier than Internal Migrants? The Role of Daily Life. *Social Indicators Research, 125(2),* 481-508.

Holder, M. D., Coleman, B., & Wallace, J. M. (2010). Spirituality, religiousness, and happiness in children aged 8–12 years. *Journal of Happiness Studies, 11*, 131-150.

Hu, P., Stylos-Allan, M., & Walker, M. P. (2006). Sleep Facilitates Consolidation of Emotional Declarative Memory. *Psychological Science, 17(10),* 891-898.

Huppert, F. A., Marks, N., Clark, A., Siegrist, J., Stutzer, A., Vittersø, J., & Wahrendorf, M. (2009). Measuring well-being across Europe: Description of the ESS well-being module and preliminary findings. *Social Indicators Research, 91(3),* 301-315.

Isen, A. M. (1970). Success, failure, attention, and reaction to others: The warm glow of success. *Journal of Personality and Social Psychology, 15(4),* 294-301.

Isen, A. M., Daubman, K. A. & Nowicki, G. P. (1987). Positive affect facilitates creative problem solving. *Journal of Personality and Social Psychology, 52(6),* 1122-1131.

Jick, T. D. (1979). Mixing qualitative and quantitative methods: Triangulation in action. *Administrative science quarterly, 24(4),* 602-611.

John, D., & Lang, F. R. (2015). Subjective acceleration of time experience in everyday life across adulthood. *Developmental Psychology, 51(12),* 1824-1839.

Jolly, D. (2009-09-14). *GDP seen as inadequate measure of economic health*. The New York Times, 14. Retrieved from: http://www.nytimes.com/2009/09/15/business/global/15gdp.html?_r=0 (lastly retrieved on: 2016-02-17)

Judge, T. A. & Kammeyer-Mueller, J. D. (2011). Happiness as a societal value. *The Academy of Management Perspectives, 25(1),* 30-41.

Judge, T. A., Thoresen, C. J., Bono, J. E. & Patton, G. K. (2001). The job satisfaction-job performance relationship: A qualitative and quantitative review. *Psychological Bulletin, 127(3),* 376-407.

Judge, T. A. & Watanabe, S. (1993). Another look at the job satisfaction-life satisfaction relationship. *Journal of Applied Psychology, 78(6),* 939-948.

Kahneman, D. (1994). New challenges to the rationality assumption. *Journal of Institutional and Theoretical Economics, 150(1),* 18-36.

Kahneman, D., Diener, E., & Schwarz, N. (1999). *Well-being: The Foundations of Hedonic Psychology.* New York, USA: Russel Sage Foundation.

Kahneman, D., & Krueger, A. B. (2006). Developments in the measurement of subjective well-being. *The Journal of Economic Perspectives, 20(1),* 3-24.

Kahneman, D., Krueger, A. B., Schkade, D., Schwarz, N. & Stone, A. (2004a). Toward National Well-Being Accounts. *American Economic Review, 94(2),* 429-434.

Kahneman, D., Krueger, A. B., Schkade, D. A., Schwarz, N. & Stone, A. A. (2004b). A Survey Method for Characterizing Daily Life Experience: The Day Reconstruction Method. *Science, 306*, 1776-1780.

Kahneman, D., Krueger, A. B., Schkade, D., Schwarz, N., & Stone, A. A. (2006). Would you be happier if you were richer? A focusing illusion. *Science Magazine, 312*, 1908–1910.

Kammann, R., & Flett, R. (1983*). Sourcebook for measuring well-being with Affectometer 2.* Dunedin, New Zealand: Why Not? Foundation.

Kantar World Panel (2014).Retrieved from: http://www.kantarworldpanel.com/Global/News/Android-ends-the-year-on-topbut-Apple-scores-in-key-markets (lastly retrieved on: 2017-10-20)

Kantar World Panel (2015). Retrieved from: https://www.kantarworldpanel.com/global/smartphone-os-market-share/intro (lastly retrieved on: 2017-10-20)

Kantor, H. I., Milton, L. J., & Ernst, M. L. (1978). Comparative Psychologic Effects of Estrogen Administration on Institutional and Noninstitutional Elderly Women. *Journal of the American Geriatrics Society, 26(1)*, 9-16.

Kettlewell, N. (2010). The impact of rural to urban migration on wellbeing in Australia. *The Australasian Journal of Regional Studies, 16,* 187–213.

Killingsworth, M. A. & Gilbert, D. T. (2010). A Wandering Mind Is an Unhappy Mind. *Science, 330(6006),* 932.

Knabe, A., Rätzel, S., Schöb, R., & Weimann, J. (2010). Dissatisfied with life but having a good day: Timeuse and well-being of the unemployed. *The Economic Journal, 120,* 867–889.

Knabe, A., Schöb, R., & Weimann, J. (2015). Partnership, Gender, and the Well-Being Cost of Unemployment. *Social Indicators Research,* 1-21.

Knight, J., & Gunatilaka, R. (2010). Great expectations? The subjective well-being of rural–urban migrants in China. *World Development, 38,* 113–124.

Koivumaa-Honkanen, H., Honkanen, R., Koskenvuo, M., & Kaprio, J. (2003). Self-reported happiness in life and suicide in ensuing 20 years. *Social Psychiatry and Psychiatric Epidemiology, 38(5),* 244–248.

Kroenke, C. H., Seeman, T., Matthews, K., Adler, N., & Epel, E. (2012). Mood patterns based on momentary assessment of positive and negative moods over a day and coronary artery calcification in the CARDIA study. Psychosomatic *Medicine, 74(5),* 526-534.

Krosnick, J. A. (1999a). Maximizing questionnaire quality. In: J. P. Robinson, P. R. Shaver, L. S. Wrightsman, J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of political attitudes* (Chapter 2, pp. 37-57). San Diego, USA: Academic Press.

Krosnick, J. A. (1999b). Survey research. *Annual Review of Psychology, 50*, 537-567.

Krosnick, J. A., & Fabrigar, L. R. (1997). Designing rating scales for effective measurement in surveys. In: L. Lyberg, P. Biemer, M. Collins, E. de Leeuw, C. Dippo, N. Schwarz, & D. Trewin (Eds.), *Survey measurement and process quality* (Chapter, 6, pp. 141-164). New York, USA: Wiley.

Krosnick, J. A., & Presser, S. (2010). Question and questionnaire design. *Handbook of Survey Research, 2,* 263-314. Bingley, United Kingdom: Emerald.

Krueger, A. B. & Schkade, D. A. (2008). The reliability of subjective well-being measures. *Journal of Public Economics, 92(8),* 1833-1845.

Larsen, R. J., Diener, E. D., & Emmons, R. A. (1985). An evaluation of subjective well-being measures. *Social Indicators Research, 17(1),* 1-17.

Layard, R. (2005). *Happiness: Lessons from a new science.* New York, NY: Penguin Press.

Lemon, B. W., Bengtson, V. L., & Peterson, J. A. (1972). An exploration of the activity theory of aging: Activity types and life satisfaction among in-movers to a retirement community. *Journal of Gerontology, 27(4),* 511-523.

Lepper, H. S. (1998). Use of other-reports to validate subjective well-being measures. *Social Indicators Research, 44(3),* 367-379.

Levy, S. & Guttman, L. (1975). Structure and dynamics of worries. *Sociometry, 38(4),* 445-473.

Lim, H.-E. (2008). The use of different happiness rating scales: Bias and comparison problem? *Social Indicators Research, 87(2),* 259–267.

Litt, M.D., Cooney, N.L. & Morse, L. (1998). Ecological momentary assessment (EMA) with treated alcoholics: Methodological problems and potential solutions. *Health Psychology 17,* 48–52.

Lu, L. (2001), Understanding happiness: A look into the Chinese folk psychology. *Journal of Happiness Studies, 2,* 407–432.

Lucas, R. E., Diener, E., & Suh, E. (1996). Discriminant validity of well-being measures. *Journal of Personality and Social Psychology, 71(3),* 616-628.

Lucas, R. E., & Donnellan, M. B. (2012). Estimating the reliability of single-item life satisfaction measures: Results from four national panel studies. *Social Indicators Research, 105(3),* 323-331.

Ludwigs, K. & Erdtmann, S. (2017). *The Happiness Analyzer – A new technique for measuring subjective well-being.* Working Paper

Lyubomirsky, S., King, L. & Diener, E. (2005). The Benefits of Frequent Positive Affect: Does Happiness Lead to Success? Psy*chological Bulletin, 131(6),* 803-855.

MacKerron, G. (2011). *Happiness and Environmental Quality* (Dissertation, The London School of Economics and Political Science, Great Britain). Retrieved from: https://core.ac.uk/download/files/170/6590187.pdf (lastly retrieved on: 2016-03-22)

MacKerron, G. (2011). Implementation, implementation, implementation: Old and new options for putting surveys and experiments online. *Journal of Choice Modelling, 4,* 20–48.

MacKerron, G. (2012). Happiness Economics from 35 000 feet. *Journal of Economic Surveys, 26(4)*, 705-735.

Mathers, C. D. & Loncar, D. (2006). Projections of Global Mortality and Burden of Disease from 2002 to 2030. *PLoS Medicine, 3(11),* 2011-2030.

Mauss, I. B., Tamir, M., Anderson, C. L. & Savino, N. S. (2011). Can seeking happiness make people unhappy? Paradoxical effects of valuing happiness. *Emotion, 11,* 807-815.

Maxwell, S. (2001). When to use MANOVA and significant MANOVAs and insignificant ANOVAs or vice versa. *Journal of Consumer Psychology, 10,* 29–30.

Melzer, S. M. (2011). Does migration make you happy? The influence of migration on subjective wellbeing. *Journal of Social Research & Policy, 2,* 73–92.

Menold, N. & Tausch, A. (2015). Measurement of Latent Variables With Different Rating Scales Testing Reliability and Measurement Equivalence by Varying the Verbalization and Number of Categories. *Sociological Methods & Research,* 1-22.

Michalos, A. C. (1980). Satisfaction and happiness. *Social Indicators Research, 8(4),* 385-422.

Michalos, A. C., & Kahlke, P. M. (2010). Stability and sensitivity in perceived quality of life measures: Some panel results. *Social Indicators Research, 98(3),* 403-434.

Mihalcea, R. & Liu, H. (2006, March). *A Corpus-based Approach to Finding Happiness*. In AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs (pp. 139-144).

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review, 63(2),* 81-97.

Miller, G. E., Chen, E., & Cole, S. W. (2009). Healh psychology: Developing biologically plausible models linking the social world and physical health. *Annual Review of Psychology, 60,* 501–524.

Molloy, R., Smith, C. L., & Wozniak, A. K. (2011). Internal migration in the United States. *Journal of Economic Perspectives, 25,* 173–196.

Nishida, M., Pearsall, J., Buckner, R. L., & Walker, M. P. (2009). REM sleep, prefrontal theta, and the consolidation of human emotional memory. *Cerebral Cortex, 19(5),* 1158-1166.

Nowlis, S. M., Kahn, B. E. & Dhar, R. (2002). Coping with ambivalence: The effect of removing a neutral option on consumer attitude and preference judgments. *Journal of Consumer Research, 29(3),* 319-334.

Nowok, B., Van Ham, M., Findlay, A. M., & Gayle, V. (2013). Does migration make you happy? A longitudinal study of internal migration and subjective well-being. *Environment and Planning A, 45,* 986–1002.

Nussbaum, M. C. (2005). *Mill between Aristotle and Bentham.* In: L. Bruni & P. L. Porta (eds.), Economics and Happiness: Framing the Analysis (1st edition, Chapter 6, pp. 170-183). Oxford, England: University Press.

OECD (Organisation for Economic Co-operation and Development). (2013). *OECD Guidelines on Measuring Subjective Well being.* Paris, France: Author. Retrieved from: http://dx.doi.org/10.1787/9789264191655-en (lastly retrieved on: 2017-06-18)

Oerlemans, W., Bakker A. B., & Veenhoven, R. (2014). Retrieved from: http://www.gelukswijzer.nl/happinessindicator/gw/ (lastly retried on: 2017-05-30)

Office for National Statistics. (2012). *First ONS Annual Experimental Subjective Well-Being Results.* Retrieved from: http://www.ons.gov.uk (lastly retrieved on: 2016-02-26)

Oishi, S. (2012). *The psychological wealth of nations: Do happy people make a happy society?* Malden, MA: Wiley-Blackwell.

Oswald, A. J., Proto, E. & Sgroi, D. (2009). *Happiness and Productivity* (IZA Discussion Paper No. 4645). Retrieved from: http://hdl.handle.net/10419/35451 (lastly retrieved on: 2016-02-17)

Oyanedel, C. N., Binder, S., Kelemen, E., Petersen, K., Born, J., & Inostroza, M. (2014). Role of slow oscillatory activity and slow wave sleep in consolidation of episodic-like memory in rats. *Behavioural Brain Research, 27*, 5126-5130.

Pasek, J., & Krosnick, J. A. (2010). *Optimizing survey questionnaire design in political science: Insights from psychology*. In: T. C. Davenport, A. S. Gerber, & D. P. Green (Eds.), the Oxford handbook of American elections and political behavior (Chapter 3, pp. 27-50). Oxford, England: University Press.

Pavot, W., & Diener, E. (1993). Review of the satisfaction with life scale. *Psychological Assessment, 5(2),* 164-172.

Pfaff, T. (2011). *Das Bruttonationalglück aus ordnungspolitischer Sicht: Eine Analyse des Wirtschafts- und Gesellschaftssystems von Bhutan* (Working Paper Series des Rates für Sozial- und Wirtschaftsdaten, No. 182). Retrieved from: http://hdl.handle.net/10419/75326 (lastly retrieved on: 2016-02-17)

Portes, A. (2000). *Social capital: Its origins and applications in modern sociology.* Boston: Butterworth Heinemann.

Pressman, S. D., & Cohen, S. (2005), Does positive affect influence health? *Psychological Bulletin, 131(6)*, 925–971.

Priesner, S. (1999), *Gross National Happiness – Bhutan's vision of development and its challenges* (Discussion Paper). Retrieved from: http://crossasia-repository.ub.uni-heidelberg.de/320/1/GNH_Ch3_Priesner.pdf (lastly retrieved on: 2016-02-17)

Priller, E., & Schupp, J. (2011). Social and economic characteristics of financial and blood donors in Germany. *DIW Economic Bulletin, 1(6),* 23-30.

Putnam, R. D. (2000). *Bowling alone: The collapse and revival of American community.* New York, NY: Simon and Schuster.

Rässler, S. & Riphahn, R.T. (2006). Survey item nonresponse and its treatment. *Allgemeines Statistisches Archiv, 90,* 217-232.

Redelmeier, D. A., Katz, J., & Kahneman, D. (2003). Memories of colonoscopy: A randomized trial. *Pain, 104,* 187–194.

Richter, David & Schupp, Jürgen (2015): The SOEP Innovation Sample (SOEP-IS). *Schmollers Jahrbuch, 135(3),* 389-399.

Robinson, M. D. & Clore, G. L. (2002a). Episodic and semantic knowledge in emotional self-report: Evidence for two judgment processes. *Journal of Personality and Social Psychology, 83(1),* 198-215.

Robinson, M. D. & Clore, G. L. (2002b). Belief and feeling: Evidence for an accessibility model of emotional self-report. *Psychological Bulletin, 128(6),* 934-960.

Rodriguez, A., Latkova ´, P., & Sun, Y. Y. (2008). The relationship between leisure and life satisfaction: Application of activity and need theory. *Social Indicators Research, 86*, 163–175.

Rohrmann, B. (1978). Empirische Studien zur Entwicklung von Antwortskalen für die sozialwissenschaftliche Forschung. *Zeitschrift für Sozialpsychologie, 9(3),* 222-245.

Rosenberg, M. (1965). *Society and the adolescent self-image.* Princeton, NJ: Princeton University Press.

Rubin, D. C. & Wenzel, A. E. (1996). One hundred years of forgetting: A quantitative description of retention. *Psychological Review, 103(4),* 734-760.

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55(1),* 68-78.

Ryan, R. M., & Deci, E. L. (2001). On happiness and human potentials: A review of research on hedonic and eudaimonic well-being. *Annual Review of Psychology, 52(1),* 141-166.

Sacks, D. W., Stevenson, B., & Wolfers, J. (2010). *Subjective well-being, income, economic development and growth* (Working Paper No. w16441). Retrieved from: http://www.nber.org/papers/w16441 (lastly retrieved on: 2016-02-29)

Sakawa, H., Ohtake, F., & Tsutsui, Y. (2015). *Activity, time, and subjective happiness: An analysis based on an hourly web survey* (Discussion Paper No. 926). Retrieved from: http://www.iser.osaka-u.ac.jp/library/dp/2015/DP0926.pdf (lastly retrieved on: 2016-04-01)

Sandvik, E., Diener, E., & Seidlitz, L. (1993). Subjective well-being: The convergence and stability of self-report and non-self-report measures. *Journal of Personality, 61(3),* 317-342.

Saris, W.E., Revilla, M., Krosnick, J.A., & Shaeffer, E.M. (2010). Comparing questions with Agree/Disagree response options to questions with item-specific response options. *Survey Research Methods, 4,* 61-79.

Schmidt, J. A., & Csikszentmihalyi, M. (2007). *Experience sampling method: Measuring the quality of everyday life.* Thousand Oaks, CA: Sage Publications Inc.

Schneider, L., & Schimmack, U. (2009). Self-informant agreement in well-being ratings: A meta-analysis. *Social Indicators Research, 94(3),* 363-376.

Schneider, S. & Stone, A. A. (2016). Ambulatory and diary methods can facilitate the measurement of patient-reported outcomes. *Quality of Life Research, 25*, 497–506.

Schooler, J. W., Ariely, D. & Loewenstein, G. (2003). The pursuit and assessment of happiness can be self-defeating. *The psychology of economic decisions, 1*, 41-70.

Schooler, J. W., & Mauss, I. B. (2010). *To be happy and to know it: The experience and meta-awareness of pleasure.* In: Berridge, K. & Kringlebach, M. (eds.). Pleasures of the brain. Oxford University Press.

Schwarz, N. (1988). Stimmung als Information: Zum Einfluß von Stimmungen und Emotionen auf evaluative Urteile. *Psychologische Rundschau, 39,* 148-159.

Schwarz, N., & Clore, G. L. (1983). Mood, misattribution, and judgments of well-being: informative and directive functions of affective states. *Journal of Personality and Social Psychology, 45(3),* 513-523.

Schwarz, N., Knäuper, B., Hippler, H., Noelle-Neumann, E., & Clark, L. (1991). Rating scales: Numeric values may change the meaning of scale labels. *Public Opinion Quarterly, 55*, 570-582.

Schwarz, N. & Strack, N. (1991). *Evaluating one's life: a judgement model of subjective well-being.* In: F. Strack, M. Argyle, & N. Schwarz (eds.), Subjective well-being: an interdisciplinary perspective (1st edition, Chapter 3, pp. 27-48). Oxford, England: Pergamon Press.

Schwarz, N., & Strack, F. (1999). *Reports of subjective well-being: Judgmental processes and their methodological implications.* In: D. Kahneman, E. Diener, & N. Schwarz (Eds.), Well-being: The foundations of hedonic psychology (Chapter 4, pp. 61-84). New York, USA: Russell Sage Foundation.

Schwarz, N., Strack, F., Kommer, D., & Wagner, D. (1987). Soccer, rooms, and the quality of your life: Mood effects on judgments of satisfaction with life in general and with specific domains. *European Journal of Social Psychology, 17(1),* 69-79.

Scollon, C. N., Prieto, C. K., & Diener, E. (2009). *Experience sampling: promises and pitfalls, strength and weaknesses*. In: Assessing well-being. Springer Netherlands.

Seligman, M. E. P. & Csikszentmihalyi, M. (2000). Positive psychology: An introduction. *American Psychologist, 55(1),* 5-14.

Simon, R. W. & Nath, L. E. (2004). Gender and Emotion in the United States: Do Men and Women Differ in Self-Reports of Feelings and Expressive Behavior? *American Journal of Sociology, 109(5),* 1137-1176.

Smith, T. W. (1979). Happiness: Time trends, seasonal variations, intersurvey differences, and other mysteries. *Social Psychology Quarterly, 42(1),* 18-30.

Steptoe, A., Wardle, J., & Marmot, M. (2005). Positive affect and health-related neuroendocrine, cardiovascular, and inflammatory processes. *Proceedings of the National academy of Sciences of the United States of America, 102(18)*, 6508-6512.

Stiglitz, J., Sen, A., & Fitoussi, J. P. (2009). *The measurement of economic performance and social progress revisited. Reflections and overview* (Working paper, No. 2009-33). Retrieved from: https://hal-sciencespo.archives-ouvertes.fr/file/index/docid/1069384/filename/wp2009-33.pdf (lastly retrieved on: 2016-02-17)

Stimson, R. J., & Minnery, J. (1998). Why people move to the 'sun-belt': A case study of long-distance migration to the Gold Coast, Australia. *Urban Studies, 35,* 193–214.

Stone, A. A., Schwartz, J. E., Schkade, D., Schwarz, N., Krueger, A., & Kahneman, D. (2006). A population approach to the study of emotion: diurnal rhythms of a working day examined with the Day Reconstruction Method. *Emotion, 6(1),* 139-149.

Stone, A. A., & Shiffman, S. (1994). Ecological momentary assessment (EMA) in behavorial medicine. *Annals of Behavioral Medicine, 16(3),* 199–202.

Stone, A.A., Shiffman, S.S., & DeVries, M.W. (1999). *Ecological momentary assessment*. In: D. Kahneman, E. Diener, & N. Schwarz (Eds.), Well-being: The Foundations of Hedonic Psychology (pp. 26–39). New York, USA: Russell Sage Foundation.

Stratton, A. (2010-11-14). Happiness index to gauge Britain's national mood. The Guardian, 14. Retrieved from: http://www.theguardian.com/lifeandstyle/2010/nov/14/happiness-index-britain-national-mood (lastly retrieved on: 2016-02-17)

Suh, E., Diener, E., Oishi, S., & Triandis, H. C. (1998). The shifting basis of life satisfaction judgments across cultures: Emotions versus norms. *Journal of Personality and Social Psychology, 74(2),* 482–493.

Sumner, L. W. (1996). *Welfare, happiness and ethics.* New York, USA: Oxford University Press.

Tatarkiewicz, W. (1966). Happiness and time. *Philosophy and Phenomenological Research, 27(1),* 1-10.

The American Declaration of Independence. (1776). Retrieved from: http://www.ushistory.org/declaration/document/index.html (lastly retrieved on: 2016-02-18)

Thoits, P. A., & Hewitt, L. N. (2001). Volunteer work and well-being. *Journal of Health and Social Behavior, 42(2),* 115-131.

Thompson, R. J., Mata, J., Jaeggi, S. M., Buschkuehl, M., Jonides, J. & Gotlib, I. H. (2011). Concurrent and prospective relations between attention to emotion and affect intensity: an experience sampling study. *Emotion, 11,* 1489-1494.

Tulving, E. (1993). What is episodic memory? *Current Directions in Psychological Science, 2(3),* 67-70.

Urry, H. L., Nitschke, J. B., Dolski, I., Jackson, D. C., Dalton, K. M., Mueller, C. J., & Davidson, R. J. (2004). Making a life worth living neural correlates of well-being. *Psychological Science, 15(6),* 367-372.

Van der Helm, E., Gujar, N., Nishida, M., & Walker, M. P. (2011). Sleep-dependent facilitation of episodic memory details. *Plos ONE, 6(11),* e27421.

Veenhoven, R. (1984). *Conditions of Happiness* (Doctoral dissertation). Retrieved from: http://www2.eur.nl/fsw/research/veenhoven/Pub1980s/84a-full.pdf (lastly retrieved on: 2016-02-22)

Veenhoven, R. (1991). *Questions on happiness: classical topics, modern answers, blind spots*. In: F. Strack, M. Argyle, & N. Schwarz (eds.), Subjective well-being: an interdisciplinary perspective (1st edition, Chapter 2, pp. 7-26). Oxford, England: Pergamon Press.

Veenhoven, R. (1994). Is happiness a trait? *Social Indicators Research, 32(2),* 101-160.

Veenhoven, R. (1997). Advances in understanding happiness (English translation). *Revue Québécoise de Psychologie, 18,* 29-74.

Veenhoven, R. (2000). The four qualities of life – Ordering concepts and measures of the good life. *Journal of Happiness Studies, 1*, 1-39.

Veenhoven, R. (2004). Subjective measures of well-being (WIDER Discussion Papers No. 2004/07). Retrieved from: http://www.econstor.eu/handle/10419/52848 (lastly retrieved on: 2016-03-29)

Veenhoven, R. (2008). Comparability of happiness across nations. School of Sociology and *Social Work Journal, 104,* 211-234.

Veenhoven, R. (2010), Greater Happiness for a Greater Number. Is that Possible and Desirable? *Journal of Happiness Studies, 11,* 605–629.

Veenhoven, R. (2015). *Classification of Happiness Measures* (World Database of Happiness, Measures of happiness, Introductory text). Retrieved from: http://worlddatabaseofhappiness.eur.nl/hap_quer/introtext_measures5.pdf (lastly retrieved on: 2016-02-26)

Veenhoven, R. (2015). *Happiness: History of the concept*
in: James Wright (Ed.) International Encyclopedia of Social and Behavioral Sciences, 2nd edition Vol. 10, pp 521-525, Oxford Elsevier

Veenhoven, R. (2016a). *World Database of Happiness*, Erasmus University Rotterdam, the Netherlands. Retrieved from: http://worlddatabaseofhappiness.eur.nl (lastly retrieved on: 2016-02-25)

Veenhoven, R (2016b). Measures of Happiness, World Database of Happiness, Erasmus University Rotterdam, the Netherlands. Retrieved from: http://worlddatabaseofhappiness.eur.nl/hap_quer/hqi_fp.htm (lastly retrieved on: 2016-02-25)

Ventegodt, S. (1995). *Quality of life in Denmark. Results from a population survey.* Copenhagen, Denmark: Forskningcentrets Forlag.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of Personality and Social Psychology, 54(6),* 1063-1070.

Weber, F. D., Wang, J., Born, J., & Inostroza, M. (2014). Sleep benefits in parallel implicit and explicit measures of episodic memory. *Learning & Memory, 21(4),* 190-198.

Weijters, B., Cabooter, E. & Schillewaert, N. (2010). The effect of rating scale format on response styles: The number of response categories and response category labels. International *Journal of Research in Marketing, 27(3),* 236-247.

Weng, L. J. (2004). Impact of the number of response categories and anchor labels on coefficient alpha and test-retest reliability. *Educational and Psychological Measurement, 64(6),* 956-972.

Wessman, A. E. & Ricks, D. F. (1966). *Mood and personality.* Oxford, England: Holt, Rinehart, & Winston.

Wilber, K. (2008). *Integral psychology*. Consciousness, spirit, psychology, therapy. Boston: Shambhala Publications.

Wismeijer, A. A., Van Assen, M. A., Sijtsma, K. & Vingerhoets, A. J. (2009). Is the negative association between self-concealment and subjective well-being mediated by mood awareness?. *Journal of Social and Clinical Psychology, 28,* 728-748.

White, A. (2007). A Global Projection of Subjective Well-being: A Challenge To Positive Psychology? *Psychtalk, 56*, 17-20.

White, M. P., & Dolan, P. (2009). Accounting for the richness of daily activities. *Psychological Science, 20,* 1000–1008.

Wilhelm, I., Diekelmann, S., Molzow, I., Ayoub, A., Mölle, M., & Born, J. (2011). Sleep selectively enhances memory expected to be of future relevance. *The Journal of Neuroscience, 31(5),* 1563-1569.

Winkelmann, L., & Winkelmann, R. (1998). Why are the unemployed so unhappy? Evidence from panel data. *Economica, 65(257),* 1-15.

Wright, T. A. & Cropanzano, R. (2004). The Role of Psychological Well-Being in Job Performance: A Fresh Look at an Age-Old Quest. *Organizational Dynamics, 33(4),* 338-351.

Wright, T. A. & Staw, B. M. (1999). Affect and favorable work outcomes: two longitudinal tests of the happy – productive worker thesis. *Journal of Organizational Behavior, 20*, 1-23.

WVS Association (World Values Survey Association). (2016). *World Values Survey* (Website). Retrieved from: http://www.worldvaluessurvey.org (lastly retrieved: 06-10-2016)

Zeleny, M. (1987). Management support systems: towards integrated knowledge management. *Human Systems Management, 7(1),* 59–70.

Acknowledgements

First, I want to thank my promotors and my copromotor. I want to thank Lidia Arends who was a great support and helped me to see the bigger picture as well as to work on every detail. I want to thank Harry Commandeur who helped me a lot, especially to combine the different chapters and make my dissertation to a readable book instead of only single papers. I want to thank Martijn Burger who was a great help from the beginning on by showing me which scientific level each paper has to have and helping me to improve the quality of every single paper. Second, I want to thank Ruut Veenhoven, who listened to me at length at our first meeting in Rotterdam in 2013 and helped in the following years to structure my thoughts and connect me to the right people to develop the idea. Third, I want to thank some scholars that gave me a lot of feedback and support from the beginning on: I want to thank Martijn Hendriks for all our discussions, your belief in the method and our great cooperation and friendship. I want to thank Ronnie Schöb and Clemens Hetschko for inviting me to Berlin in 2014 to discuss our idea and to always supporting it from there on. I want to thank Richard Lucas for inviting me to Michigan to discuss our method in detail and for his support for our experimental studies as well as for giving us part of his funding from the National Aging Foundation to pay the rewards for the second study with the German Socio-Economic Panel. I want to thank Jürgen Schupp and David Richter from the German Socio-Economic Panel for supporting the idea and letting us experiment with their innovation sample. And I want to thank all other scholars that gave me feedback to the idea in different meetings and at different conferences. Next, I want to thank the nine people who supported us financially to be able to start the institute "Happiness Research Organisation" that gave us the possibility to attract enough funding to develop the tool and to evaluate it: Alexander Philipp Gruissem, Cort Kibler-Melby, Max Altergott, Igor Kurganov, Olivia Boeree, Thomas Laumer, Ramin Henke, Nasr el Nasr and Marko Neumann. I would also

like to thank my (former) colleagues from the Happiness Research Organisation, especially Lena Henning for working on the projects and Jasmin Mertens for early support. I am especially thankful to thank Stephan Erdtmann, my best friend, whom without his help I would never have been able to develop the Happiness Analyzer. I am thankful to my mom, Ursula, my father, Wilhelm and my brother, Till Ludwigs, as well as my mother and father in law Brenda and Barry Goodrich for their mental support. Lastly, I want to thank my wife Erin Ashley Ludwigs, who moved from the USA to Germany for me so that I could work on the idea, supported me throughout the whole process and gave birth to our beautiful daughter, Hadley Joy, at the end of this journey.

About the Author

Kai Ludwigs is a studied Psychologist (B.Sc, M.Sc, Heinrich-Heine University Düsseldorf) and studied Economist (B.Sc, FernUniversität in Hagen). He is the Founder and Director of the Happiness Research Organisation (HRO), an independent research institute based in Düsseldorf, Germany that is specialized in app-based research on topics such as subjective well-being, quality of life, happiness and satisfaction. Since 2016 he is a board member of the International Society for Quality of Life Studies and of the International Association for Time Use Research. His research interests are the measurement of subjective well-being, quality of life, satisfaction and time-use and the opportunities of modern survey technologies.

Portfolio

Refereed Publications

- Hendriks, M., Ludwigs, K., & Veenhoven, R. (2016). Why are locals happier than internal migrants? The role of daily life. *Social Indicators Research, 125*(2), 481-508.
- Ludwigs, K., Lucas, R., Veenhoven, R., Burger, M. & Arends, L. (in press). How Does More Attention to Subjective Well-Being Affect Subjective Well-Being? *Applied Research in Quality of Life.*

Book Chapters

- Kim, Y., & Ludwigs, K. (2017). Measuring Community Well-Being and Individual Well-Being for Public Policy: The Case of the Community Well-Being Atlas. In *Handbook of Community Well-Being Research* (pp. 423-433). Springer Netherlands.
- Hendriks, M., Ludwigs, K., & Bartram, D. (2017). International migration decisions and happiness: The Migration Happiness Atlas as a community development initiative. In R. Phillips, S. Kenny, B. McGrath (Eds.), *Handbook of Community Development*.
- Ludwigs, K., Henning, L., Arends, L. (in press). Measuring Happiness: A practical review. *Perspectives on Community Well-Being.*

Organizational contributions

- Review Board of the International Society for Quality of Life Studies
- Board Member of the International Society for Quality of Life Studies
- Board Member of the International Association for Time Use Research
- Board Member of the Community Well-Being Institute
- Council for EUROSTAT for HETUS studies

Awards and grants

- Selected study for the Innovation Sample of the German Socio-Economic Panel (2015 and 2016)

- Several grants for the Happiness Research Organisation for applied subjective well-being research studies for cities and companies (over 600,000 Euros between 2015 and 2017)
- Part of several research grants, for example the EU funded project "Design For Wellbeing" (over 1,000,000 Euros total funding) or a project on unemployed people's well-being funded by the German Research Foundation (DFG) (over 500,000 Euros total funding)

Publicity

- Newspaper articles in German newspapers such as Süddeutsche Zeitung, Welt and Rheinische Post
- TEDx Talk about a community well-being project in Frankfurt https://www.youtube.com/watch?v=FI8c14giiWQ

Professional training, seminars, lectures

- Attending the EHERO seminar series at Erasmus University Rotterdam since 2013
- Attending different conference workshops at the conferences of the International Society for Quality of Life Studies in Berlin, Phoenix, Seoul and Innsbruck
- Attending conference workshops at the OECD World Forum 2015 in Guadalajara

Invited professional lectures and presentation

- Professional lectures and presentations at several companies such as Accenture and Coca-Cola
- Professional presentations at different Universities such as Oxford University, University College London, Freie Universität Berlin, Michigan State University, Seoul National University
- Professional presentations at public institutions such as the OECD, the German Statistical Office and the German Socio-Economic Panel.

Conferences, seminars, and workshops

- Conference of the International Society for Quality of Life Studies (ISQOLS), Innsbruck, Austria, September 2017.
- Conference of the Erasmus Happiness Economics Research Organisation (EHERO) and the Erasmus Institute for Philosophy and Economics (EIPE), Rotterdam, the Netherlands, March 2017.
- OECD Conference on Well-Being in the long-term, London, England, December 2016.
- Conference of the International Society for Quality of Life Studies (ISQOLS), Seoul, South Korea, August 2016.

- Conference of the International Association for Time Use Research (IATUR), Seoul, South Korea, July 2016.
- Conference of the International Society for Quality of Life Studies (ISQOLS), Phoenix, USA, October 2015.
- OECD World Forum, Guadalajara, Mexico, October 2015.
- Conference of the International Society for Quality of Life Studies (ISQOLS), Berlin, Germany, September 2014.
- Conference of European Network for Positive Psychology (ENPP), Amsterdam, Netherlands, July 2014.
- Conference of the Erasmus Happiness Economics Research Organisation (EHERO), Rotterdam, the Netherlands, October 2013.

Referee Services

- Applied Research in Quality of Life
- Social Indicators Research

Academic degrees

- B.Sc, Psychology (Heinrich Heine Universität Düsseldorf)
- B.Sc, Economics (FernUniversität in Hagen)
- M.Sc, Psychology (Heinrich Heine Universität Düsseldorf)

Summary

In 2013 the OECD published a guideline that represents the gold standard for measuring subjective well-being in greater detail to collect data in the quality needed as a basis for efficient decisions to improve subjective well-being and the evaluation of those decisions to enable continuous learning. Unfortunately, most existing studies nevertheless do not measure subjective well-being according to this standard, as traditional methods (e.g., paper and pencil or personal interviews) require considerable resources (from both researchers and participants) to capture i) people's subjective well-being at multiple points in time using general questionnaires; ii) people's everyday life and everyday feelings; iii) people's specific feelings in the moment; and iv) a combination of subjective and objective well-being measurements. To resolve this issue, we developed an app as a mobile assessment tool, the "Happiness Analyzer".

In this dissertation we review and discuss existing subjective well-being measurements, outline the functionalities of the Happiness Analyzer and show that it is possible to collect beneficial data with the technique, that the technique seems to have a beneficial effect on the user, that the timing of the data collection is flexible and that it is possible to collect representative datasets with this technique.

In conclusion, we developed and evaluated a new technique for measuring subjective well-being and invite researchers, mayors, managers and therapists to use it to better understand people's subjective well-being.

SUMMARY IN DUTCH

De OECD heeft in 2013 richtlijnen gepubliceerd die worden gezien als de gouden standaard voor het gedetailleerd meten van subjectief welzijn zodat kwalitatief goede data kunnen worden verzameld die kunnen dienen als basis voor efficiënte besluitvorming betreffende het verhogen van subjectief welzijn en deze besluitvorming te evalueren en daarvan te leren. Helaas volgt het meeste onderzoek deze standaard niet omdat traditionele methoden (bijv. schriftelijke vragenlijsten of persoonlijke interviews) veel vergen van zowel onderzoekers als participanten betreffende het krijgen van meer inzicht in i) subjectief welzijn in verschillende tijdsperiodes, ii) het dagelijks leven en geluk van mensen, iii) specifieke gevoelens van mensen op specifieke momenten, en iv) het zowel meten van subjectief en objectief welzijn. Om dit issue te tackelen hebben we een app ontwikkeld die dient als meetinstrument, de "Happiness Analyzer".

In dit proefschrift bediscussiëren we bestaande meetinstrumenten van subjectief welzijn, schetsen de karakteristieken van de Happiness Analyzer, en tonen aan dat het mogelijk is om waardevolle data met deze techniek te verzamelen, dat het gebruik van deze techniek ook positieve effecten kan hebben op participanten, dat de timing van de data collectie flexibele is, en dat het mogelijk is om representatieve datasets te genereren met deze techniek.

De kern van dit proefschrift is dus de ontwikkeling en evaluatie van een nieuwe techniek om subjectief welzijn te meten. We nodigen onderzoekers, beleidsmakers, managers en therapeuten uit deze techniek te gebruiken om kennis op te doen over subjectief welzijn.

SUMMARY IN GERMAN

Im Jahre 2013 hat die OECD Guidelines veröffentlicht, die heute den Gold-Standard zur Messung von subjektivem Wohlbefinden darstellen. Diese detaillierteren Daten werden gebraucht damit effiziente Entscheidungen zur Steigerung von subjektivem Wohlbefinden getroffen und dauerhaft evaluiert werden können. Leider messen die meisten Studien subjektives Wohlbefinden nicht nach diesem Standard, da traditionelle Forschungsmethoden (z.B. Papier und Bleistift oder persönliche Interviews) einen hohen Ressourcenaufwand verlangen, sowohl von Forschern als auch Teilnehmern, um: i) Das subjektive Wohlbefinden der Menschen an verschiedenen Zeitpunkten mit Fragebögen zu messen; ii) Das alltägliche Leben und Erleben der Menschen zu erfassen; iii) Die gegenwärtigen Gefühle der Menschen im Moment zu erfassen; und iv) Eine Kombination von subjektiven und objektiven Wohlbefindensmaßen zu schaffen. Um eine ressourcenschonende Messung von subjektivem Wohlbefinden nach dem OECD Gold-Standard zu ermöglichen, haben wir die App „Happiness Analyzer" entwickelt.

In dieser Arbeit erfassen wir zunächst existierende Methoden zur Messung von subjektivem Wohlbefinden und diskutieren diese. Hiernach erklären wir die Funktionen des Happiness Analyzer und zeigen, dass es möglich ist mit der App sinnvolle Daten zu erheben, dass die App einen positiven Effekt auf die Teilnehmer hat, dass es flexibel ist wann genau gemessen wird und dass es generell möglich ist repräsentative Daten mit der App zu erheben.

Zusammenfassend können wir festhalten, dass wir eine neue Technik zur Messung von subjektivem Wohlbefinden entwickelt und evaluiert haben und nun Forscher, Stadtverantwortliche, Manager und Therapeuten einladen diese Technik zu nutzen, um das subjektive Wohlbefinden der Menschen besser zu verstehen.