CAN HAPPINESS APPS GENERATE NATIONALLY REPRESENTATIVE DATASETS?

A Case Study Collecting Data on People's Happiness Using the German Socio-Economic Panel

Kai Ludwigs, Ruut Veenhoven, Richard Lucas, David Richter and Lidia Arends

Applied Research in Quality of Life. Online since April 2019 at https://link.springer.com/article/10.1007%2Fs11482-019-09723-2#s

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 financial rewards. But quantity is not everything that matters 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), a different group of 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

1 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 longitudinal 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 additional data to what 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 additional 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, different 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 good participation quality rates. In the fourth section, we will discuss the results to answer the question whether happiness apps can generate nationally representative samples.

2 METHOD

2.1 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). Due to SOEP regulations data was collected in two different waves with different participants and not in one wave. 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 2016 and February 2017. 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 1.

2.2 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

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 1 displays the ESM and DRM modules:

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 2 displays the graphical feedback:

50 Euro Amazon vouchers

In the second study, participants were able to 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 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.

3 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 good participation quality rate) 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 2, 3, 4, 5 and 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 only 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 3 and 4 show that participation rates increased significantly ($\chi 2$ = (1, 4004) = 254.25; p < .001 & $\chi 2$ = (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 ($\chi 2$ = (1, 1129) = 9.30; p = .002) and in the second sample more younger people participate between 17 and 39 years participate ($\chi 2$ = (1, 1050) = 79.25; p < .001). For employment status there is no significant difference in the participation rates in the smartphone sample.

Table 5 and 6 display that the good participation quality rates increased significantly ($\chi 2$ = (1, 4004) = 332.67; p < .001 & $\chi 2$ = (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 good participation quality rate 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 good participation quality rate 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 ($\chi 2$ = (1, 1050) = 64.95; p < .001).

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 good participation quality rates 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.

4 DISCUSSION

4.1 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 longitudinal 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, different participants got a 50 Euro Amazon voucher for using the app.

The results indicate that the participation rate and good participation quality rate 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.

4.2 Methodological discussion

Unfortunately, we do not have any other information about participants' attitude towards mobile phones apart from if they own a smartphone or not. Vicente & Lopes (2016) interviewed 1,501 mobile phone users and differentiated three different groups: i) "Apathetic" users who only have a mobile phone to be able to communicate with people and be available; ii) "Social and hedonic" users who mainly see having a mobile phone as a status object but also think that mobile phones can be addictive and can affect privacy; iii) "Busy and active" users that use their mobile phone frequently for example as a working tool. Given their analysis "Apathetic" and "Social and hedonic" users might have a smartphone in our study but would not really be willing to participate in a study where they are notified multiple times a day to give feedback. Therefore, only the "Busy and active" users would participate but we do not have any data on the distribution in our sample but suggest collecting data on participants' attitude towards their mobile phone for further investigations.

There is an increasing rate of people owning a smartphone already between 2015 and 2016 in our data and according to Statista (2019) the rate will increase from 61% monthly active smartphone users in Germany in 2016 to 78% in 2022. This means that the response rates in our study might look very different when we would run the same study in some years again.

Different sampling techniques (in person, per phone, per online advertisement, per print advertisement) can have different effects on participation rates. In this study, participants of a household survey were sampled in person by interviewers. In general, this is the most intense sampling technique and 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 participants' 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 although we do not have any source to support this thought.

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 because we would not have the confounding time factor where for example participants' attitude towards their mobile phones might have changed.

Other demographic variables would be important to look at to investigate national representativity regarding other demographic factors such as education, income levels, etc. The GSOEP collects hundreds of other demographic factors in their representative household surveys since 1984 so that other factors will be available, which will probably be the case in 2020. Thus, readers should feel free to contact the first author for further analysis.

Because we did not offer participants to participate in the study in a different format (e.g. online with e-mail notification or paper-pencil with a beeper) we cannot differentiate if participants did not participate because they did not want to participate with a smartphone, or they did not want to participate because of the topic of the study. For example, Guo et al. (2016) show that response rates differ significantly depending on the survey mode. Surveying participants on their smartphone might also affect their responses compared to other survey modes such as web-surveys or paper-pencil surveys. Although for example in a recent study Sarracino, Riillo & Mikucka (2017) do not find significant differences in the answers to well-being questions if answered in a phone interview or in a web survey.

As the results are limited to a German sample, the findings should also be investigated in other nations. Germany is a wealthy country with a comparably high smartphone penetration rate and thus we cannot generalize our findings for other nations.

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, answering patterns and participation quality differ when participants are paid more or less because other studies showed effects due to reward variations (e.g. Guo et al. 2016).

5 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.

REFERENCES

Csikszentmihalyi, M. & Hunter, J. (2003). Happiness in everyday life: The uses of experience sampling. *Journal of Happiness Studies, 4,* 185-199.

German Census (2011). https://www.destatis.de/DE/Methoden/Zensus_/Zensus.html. Retrieved: October 20th 2017.

German Statistical Office (2015). https://www.destatis.de/DE/PresseService/Presse/Pressemitteilungen/2015/05/PD15_172_631.html Retrieved: October 20th 2017.

Guo, Y., Kopec, J. A., Cibere, J., Li, L. C., & Goldsmith, C. H. (2016). Population survey features and response rates: a randomized experiment. *American journal of public health, 106(8),* 1422-1426.

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.

Kahneman, D., Krueger, A. B., Schkade, D. A., Schwarz, N. & Stone, A. A. (2004). A survey method for characterizing daily life experience: The day reconstruction method. *Science, 306*, 1776-1780.

Kantar World Panel (2015). https://www.kantarworldpanel.com/global/smartphone-os-market-share/intro. Retrieved: October 20th 2017

Killingsworth, M.A. & Gilbert, D.T. (2010). A wandering mind is an unhappy mind. *Science, 330,* 932-932.

Vicente, P., & Lopes, I. (2016). Attitudes of older mobile phone users towards mobile phones. *Communications-European Journal of Communication Research, (1),* 71-86.

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

MacKerron, G. (2012). Happiness and Environmental Quality. PhD thesis, London School of Economics and Political Science. Available from: http://etheses.lse.ac.uk/383/. Retrieved: October 25th 2017

OECD (2013). OECD Guidelines on Measuring Subjective Well-being. OECD Publishing. http://dx.doi.org/10.1787/9789264191655-en. Retrieved: October 26th 2017

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

Sarracino, F., Riillo, C. A. F., & Mikucka, M. (2017). Comparability of web and telephone surveys for the measurement of subjective well-being. *Survey Research Methods, 11*(2), 141–169.

Statista (2019). https://www.statista.com/statistics/568095/predicted-smartphone-user-penetration-rate-in-germany/ Retrieved: March 12th 2019

Table 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.

Table 2: descriptive characteristics of the two samples in percent

	Smartphone Ownership (Android or iOS)		
Variable	**First Sample N = 1129 (52.9%)**	**Second Sample N = 1050 (56.2%)**	**$\chi 2$ (1, 4004) = 4.37; *p* = .037**
	%	%	
Age	**$\chi 2$ (1, 2135) = 646.64; *p* < .001**	**$\chi 2$ (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	$\chi 2$ (1, 2135) = 0.56; *p* = .454	$\chi 2$ (1, 1869) = 0.20; *p* = .654	
Male	*53.7*	*56.7*	
Female	*52.1*	*55.7*	
Employment Status	**$\chi 2$ (1, 2135) = 218.09; *p* < .001**	**$\chi 2$ (1, 1869) = 249.16; *p* < .001**	
Having A Job	*69.1*	*73.8*	
No Job	*37.2*	*37.6*	

Note:. 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: Participation Rates Whole Sample

Variable	**First Sample N = 84 (3.9%)**	**Second Sample N = 374 (20.0%)**	**χ2 (1, 4004) = 254.25; *p* < .001**
	%	%	
Age	**χ2 (1, 2135) = 42.61; *p* < .001**	**χ2 (1, 1869) = 319.73; *p* < .001**	
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	**χ2 (1, 2135) = 8.13; *p* = .004**	χ2 (1, 1869) = 0.64; *p* = .424	
Male	*2.7*	*19.2*	
Female	*5.1*	*20.7*	
Employment Status	χ2 (1, 2135) = 3.78; *p* = .052	**χ2 (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 4: Participation Rates Smartphone Sample

Variable	**First Sample** **N = 84 (7.4%)**	**Second Sample** **N = 374 (35.6%)**	**$\chi2$ (1, 2179) = 260.22; *p* < .001**
	%	%	
Age	$\chi2$ (1, 1129) = 5.66; *p* = .580	**$\chi2$ (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	**$\chi2$ (1, 1129) = 9.30; *p* = .002**	$\chi2$ (1, 1050) = 1.23; *p* = .267	
Male	*5.0*	*33.8*	
Female	*9.8*	*37.2*	
Employment Status	$\chi2$ (1,2135) = 0.87; *p* = .351	$\chi2$ (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: Good Participation Quality Rates 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*	

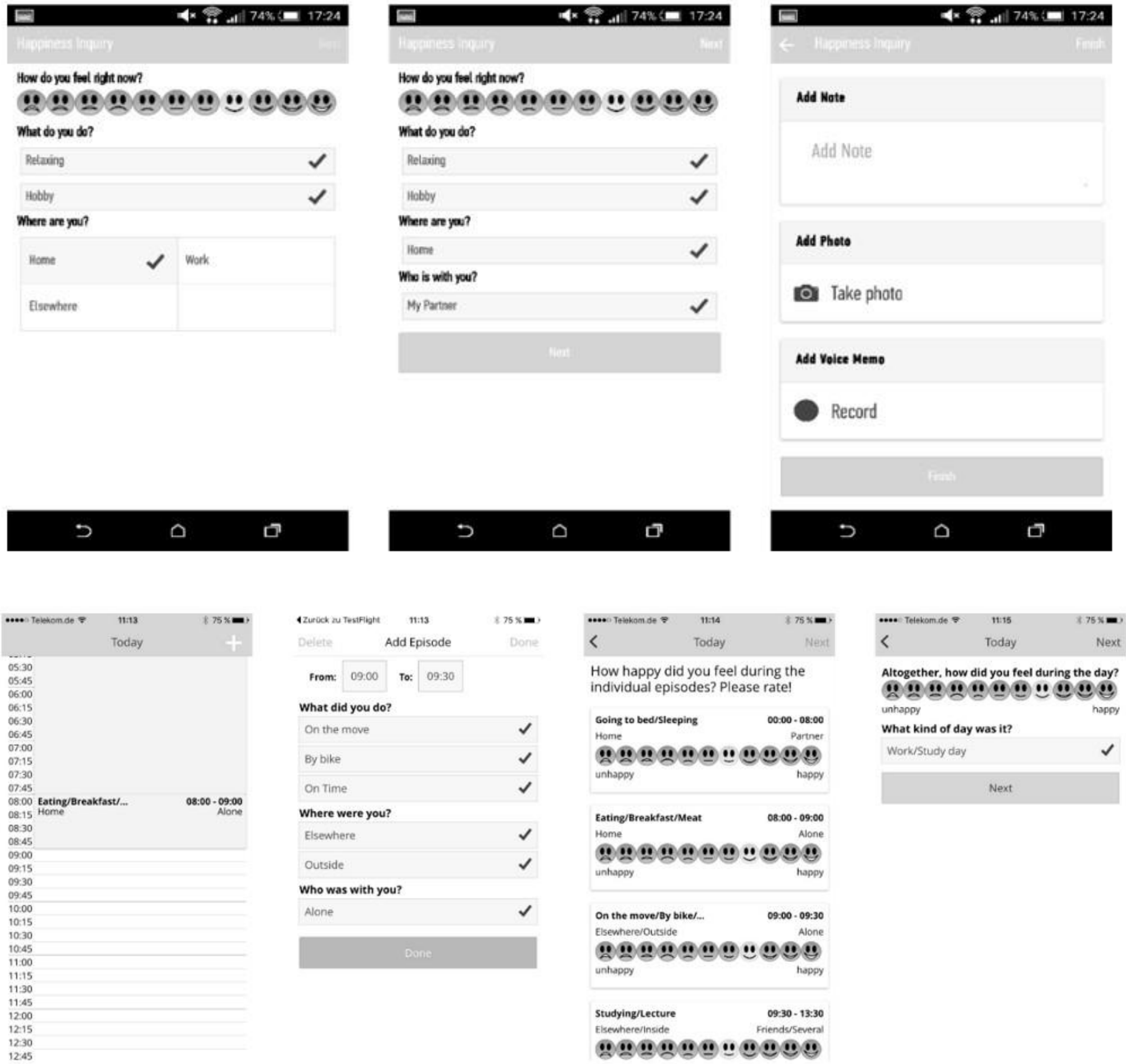
Note: The table displays the descriptive characteristics of the two samples in percent. The general good participation quality rates are displayed in brackets in the first row in the second and third column. Significant differences (p < .05) are investigated with chi-square tests: i) For the difference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

Table 6: Good Participation Quality Rates Smartphone Sample

Variable	**First Sample N = 24 (2.1%)**	**Second Sample N = 326 (31.0%)**	**$\chi2$ (1, 2179) = 337.53; *p* < .001**
	%	%	
Age	$\chi2$ (1,1129) = 5.87; p = .555	**$\chi2$ (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	$\chi2$ (1, 1129) = 2.51; *p* = .113	$\chi2$ (1, 1050) = 3.05; *p* = .081	
Male	*1.4*	*28.4*	
Female	*2.8*	*33.4*	
Employment Status	$\chi2$ (1, 1129) = 0.64; *p* = .800	$\chi2$ (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 good 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. Good Participation quality rates differ significantly between the two samples and according to Age in the second sample.

Figure 1: Sceenshots:ESM module (first row) and the DRM module (second row)



In the ESM module participants had to rate how they feel, what they are doing, where they are and who is with them and had afterwards the option to add a note or picture or a voice recording. In the DRM module participants had to define their day in episodes at the end of a day or on the next day for the previous day (e.g. 9am to 9:30am) and then had to define what they did in these episodes, where they were and who was with them. After they defined the episodes they had to rate how they felt during these episodes and at the end had to rate how they felt during the whole day and what kind of a day it was.

Figure 2: The graphical feedback that participants received.

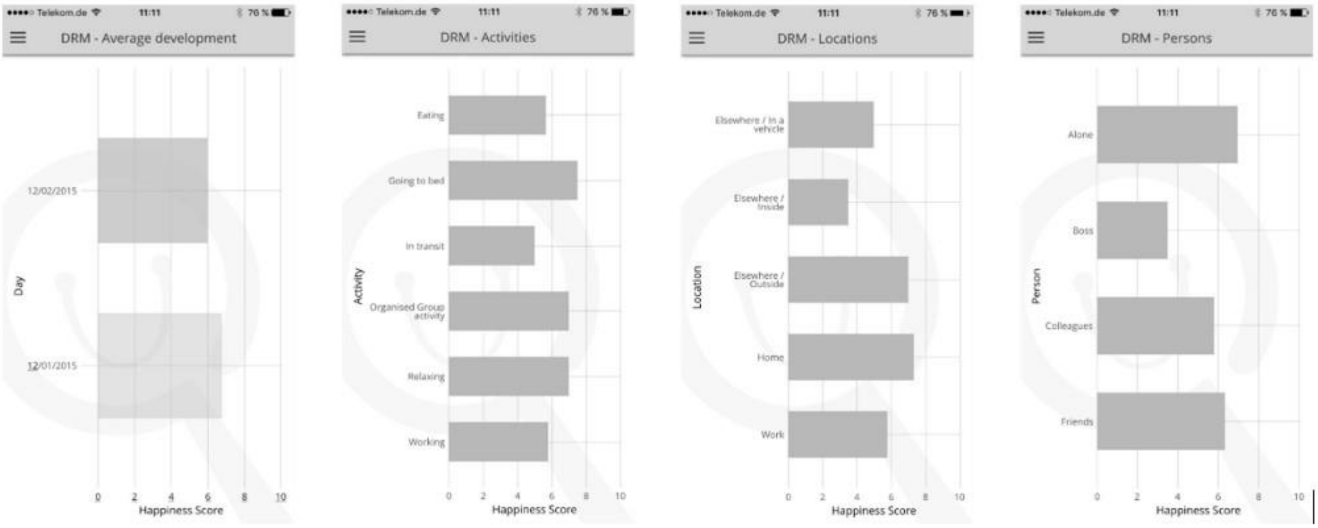
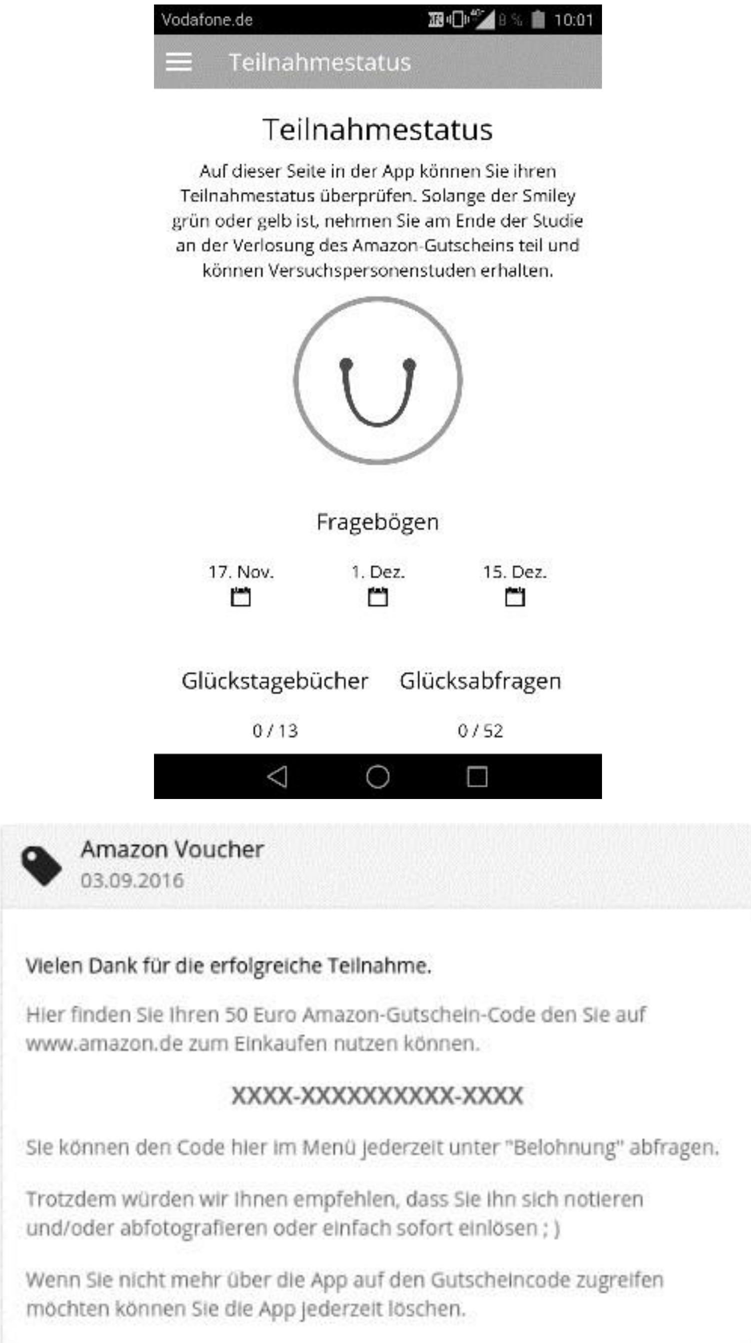


Figure 3: The "smiley screen" (first row) and the Amazon voucher (second row)



Unfortunately, these screens are only available in German but do not display much more than explained in the text above. In case a translation is needed please contact the main author.