STATISTIK AUSTRIA

Die Informationsmanager

Final Report

"Preparatory work to revise the harmonised European Time Use survey for the implementation in the round 2020"

EUROSTAT-GRANT Agreement No. 07141.2016.001-2016.772

Statistics Austria

Vienna, December 2017

For further information please contact:

Project Manager and Time Use Survey Austria:
Sonja Ghassemi-Bönisch
Guglgasse 13
1110 Vienna
Tel.: +43 1 711 28 - 7103
sonja.ghassemi-boenisch@statistik.gv.at

Subjective well-being:
Kathrin Gärtner
kathrin.gaertner@statistik.gv.at

Ivo Ponocny
ivo.ponocny@modul.ac.at

Christian Weismayer
christian.weismayer@modul.ac.at

Software development:
Friedrich Csaicsich
friedrich.csaicsich@statistik.gv.at

Gabriel Kittel
gabriel.kittel@statistik.gv.at

Usability testing:
Matea Paškvan
matea.paskvan@statistik.gv.at

Marc Plate
marc.plate@statistik.gv.at

Sonja Hinsch
sonja.hinsch@univie.ac.at

Constanze Volkmann
constanze.volkmann@univie.ac.at

Design of the app:
Manuela Heidenreich
manuela.heidenreich@statistik.gv.at

Table of Contents

1   Executive Summary

The aim of this project was to develop a Time Use Diary based on browser-technology that works on mobile devices and test this diary for usability. For achieving this target, a literature- and a best –practice–review was done. Also the findings of the 'Modul University Study' and the literature review to subjective well-being were important to develop this app and helped to have a better theoretical base for going into the field. The app was working satisfyingly but there is still potential for improvement and to make it more user-friendly. The report describes results from all phases of this project and gives recommendations on the basis of our findings.

There are already some examples for the measurement of time use via app. Many commercial apps are available on the free market and show that there is an interest for this topic for the users. Apps for time use research are at the beginning. In some of the studies for example in the pilot of the Netherlands institute of social research (2013), not only activities during the day were recorded via app but also well-being. There are also several traditional time-use-surveys which ask for well-being and the subjective evaluation of activities, either at the end of the diary (Finland, Germany and Norway) or for every time slot (France, Italy, Poland and UK).

From a well-being perspective evaluated time use (which could be described as the micro measurement of well-being)- becomes increasingly prominent as a measure of general well-being. In a wider frame measuring subjective well-being within time use surveys can also help to entangle the role of context for (micro-) well-being and – as an input for economic theory and research- to clarify the concept of utility and the relationship between utility and time use.

In the Modul study of living conditions described in chapter 5 Ponocny et al could not only show that the 11 –point scale might be to detailed and skewed to measure momentary well-being but also that this momentary well-being is not only influenced by the pleasantness of the activity conducted but also by other factors and circumstances.

On the basis of an literature review on well-being measurement and of the results from the MODUL study we chose the following design regarding well-being measurement in our app: Two different questions are provided regarding evaluated time use: one about the momentary hedonic status (*happy*) and another one about being pleased by the activity which is carried out at the moment (*I like what I do*). Additionally respondents are asked about being stressed. All three items were measured on a unidimensional 7- point scale from "very" to "not at all" via push-messages four times a day.

In our project we developed a time use diary based on browser-technology that works on mobile devices and tested this diary for usability.

The app- prototype is developed in accordance with the actual HETUS-Guidelines from 2008 and has two important new features: look- up tables (i.e. testing verbatim mode of data collection with predefined list of activities that will appear after typing some initial word letters) and the possibility for copying information across time slots.

For the TUS-app, the decision has been made to implement a hybrid app using the Cordova framework. Cordova provides a container for each target platform that contains the almost platform-independent web app providing the functionality to the user. Plugins allow access to system

functionality like storage and local notifications. The user interface has been developed using the Ionic framework which allows adaptation of the app to the look and feel of native apps thus providing a uniform overall user experience. The app is available for the operating systems Android and iOS.

After the app was developed and internally tested a usability test with a test-design that includes video and telephone interviews was implemented. The test showed that the app was working adequately but there is still potential for enhancement and to make it more user-optimised.

The respondents of the test indicate that they are already used to apps and have higher demands on the standard regarding the usability of an app. For the next time use survey the technical impact has definitely to be considered. The loading time of an app and different devices with different operating versions can result in performance problems - especially for older mobile phones. A challenge the respondents faced were the 144 time slots. The app should adapt to the user's personal lifestyle and not force an artificial reality upon them. Especially the ten minute slots, although they are meant to help the respondents to be as accurate as possible, cannot reach this aim and the respondents tend to get imprecise. The solution for this might be to establish a calendar-like-version of the app with a start-and end button for of the activity.

We advise to use the build-in-dictionary of the mobile device in connection with the autocomplete function of the mobile phone and not a determined and given activity list. In terms of data preparation we suggest not spending too much time in developing the activity list and invest a lot more time in automatic text recognition. The usability test showed there are different types of time diary users. The one who fills out the activities afterwards, others report their activities 'live' during the day and some even in advance. The app has to cover all needs.

The app should give the participants immediate feedback. This is essential and it serves the need of the respondent. The detailed reconstruction of the day is burdensome and the users of an app are used to get an instant gratification. This gratification can be a graphic demonstration of the personal time for the different activities. The app should make fun without being childish or suspect and give the respondents some extra value immediately.

We advise to also get in touch and use ideas from open source projects i.e. "Tabi project" or ESSNET Serv. A new test phase should be definitely considered before the next wave in 2020. Enough testing time is really necessary. Unforeseen bugs and challenges that can occur in the field phase will need extra time for getting new solutions.

2 Introduction

The last time use survey in Austria was conducted in 2008/09. The next European wave of Time Use Surveys will start in 2020 and therefore it will be necessary to enhance the design of the survey to increase data quality and response rates.

With the Grant of Eurostat it was possible to develop a mobile app for time use surveys and introduce a new technology in this survey. Thus respondents have the possibility of filling out the diary in a more convenient and contemporary way. With the newly developed app-prototype, it is possible to go one step further into the application of new design features for time use surveys.

It is not clear if Austria can attend the next wave. It depends on the financing of the next survey. With this app-prototype it is possible to go one step further to the next wave of time use surveys.

3 Measuring Time Use via App

The following chapter gives a description of the awarded Grant with the time table of this project, a best-practice-review of apps and a discussion why surveys via app are going to be the first choice.

3.1 Description of the awarded Grant

The aim was to develop a time use diary based on browser-technology that works on mobile devices and test this diary for usability. For achieving this target a literature- and best–practice–review was done. The results of the Modul University Study and the literature review on subjective well-being were important to develop this app and helped to have a better theoretical base for going into the field. Results from the study of the Nederland's "Using smartphones in survey research: a multifunctional tool" were also taken into account and all this components assisted the development and testing of an time use app.

The TUS-diary includes all essential features for users to fill out the diary and follows the HETUS-Guidelines. These are for example the 10-minutes time slot, main and secondary activity, where was the activity done and with whom. The prototype exhibited two important new features: look- up tables (i.e. testing verbatim mode of data collection with predefined list of activities that will appear after typing some initial word letters) and the possibility for copying information across different time slots. Another feature was created, that made it possible to start the diary more individually. The respondents had the possibility to tell when they got up at this day and for the activity sleeping it was possible to fill out more timeslots at once.

Statistics Austria implemented lookup-tables for the time use activities to make it easier for the respondents to fill in the diary. Lookup -tables are lists of various activities which offer different options every time users start to type a certain activity. For example, if they type "wo": "work", "workout" and "worship" are possible activities which are offered. The use of the diary should be intuitive and the design minimalistic and attractive for the respondents at the same time.

Another feature that we developed is to easily copy the activity for the various timeslots. So if the respondents have the activity "watching TV" for several hours, they can copy it to the next time slots in a simple way.

As suggested by HETUS regulations a well-being item was included into the diaries. Push-notifications were sent by the app. As a first step for the construction of such an item, a literature and best practice review for subjective well-being questions in connection with time use was conducted. As suggested by the Expert Group on Quality of Life, time use surveys may serve as an excellent instrument for the collection of subjective well-being (SWB)-related information. Since the affective component of quality of life takes place in time, observations over time in short intervals may seem, at least from a certain perspective, as the via regia to judge "how life is".

The prototype followed the recommendation of the HETUS-Guidelines but the experiences are also beneficial for other surveys, for example the Household Budget Survey and of course other Time Use Surveys. Therefore our project will be useful for other projects as well and has a sustainable impact.

Project Task – Description	Target	Status
Literature review best-practice review of web diaries based on browser technology (secondary review), Primary Analysis of TUS-data of a survey from Modul University with connex to subjective well-being questions	M + 3	January- March 2017
Conception of the prototype "TUS-Diary for mobile devices" including the development of lookup-tables for the activities and a feature to copy activities in the diaries	M + 6	March – May 2017
Development of the prototype diary day for mobile devices	M + 8	May – August 2017
Usability test Testing of the prototype including cognitive test and probing with regard of subjective well-being question in connection with time use	M + 9	September and October 2017
Analysis Analysis of the usability test	M + 11	November 2017
Report Final technical and methodological report with results of the analysis and potential proposals for the guidelines of HETUS	M + 12	November and December 2017

3.2 Best-practice-review

In principle, TUS methodology is well-defined in guidelines developed by Eurostat (2008; Harmonized European time use surveys) or the UN (Guide to Producing Statistics on Time Use, 2005). But the digital development requires also new approaches in survey techniques. It is already common sense that is necessary to modernise surveys and to make it easier for respondents to participate in different modes and thereby to increase the response rates and the quality of the data.

The new techniques have their advantages and disadvantages. It is clear that a mobile time-use app has the potential to activate new population groups and to get a better response rate in these

groups. At the other hand, apps on smartphones are not convenient and suitable for all respondents. Some people are critical of apps and their way of data collection. And there is also a group of persons who have no or not much experience with smartphone apps (see "ICT usage in households and by individuals"). So for this population apps will not work and therefore there is still a need for paper-diaries. A multi-mode- mix will be a possibility to cope with this situation.

Another issue is that the different data collection modes can have an impact on the results. Especially the comparison of the results will need more attention. The comparison of the results within the survey with a multi-mode-approach will probably indicate a mode-effect. But even if it is just an online-survey, a change of method (e.g. introducing an app) will influence the historic comparison of the results of former time use surveys. But nevertheless we have to cope with this and go with the digital development.

The new development will also cause new costs, i.e. more IT resources for programming the survey infrastructure. But there is the potential to have savings at the coding process. A requirement for it is automatic recognition software that allows coding as many activities as possible and learns within the coding process.

Table 2: Advantages and disadvantages of a Time-Use- App

Advantages	Disadvantages
Answers of new population groups => Higher response rates	Not for all respondents useful => still recruiting with paper necessary
Improvement of the data quality and getting an image of the reality	Change of method
Potential savings at the coding process	New development => new costs

3.2.1 Reason for "mobile first"

As presented by Peter Lugtig[1] and Vera Toepoel at the last ESRA conference in Lisbon 2017, a mobile first design should be used for all surveys were it is possible. Especially young people are more and more reluctant to take part in official surveys. On the other hand, the young generation is the group with the most intensive mobile phone use in the population. If one wants to reach them, mobile phone friendly surveys have to be used. But when is a survey mobile phone friendly? According to Lugtig and Toepoel it has to be short (or consist of short parts), it should be small in data volume (no pictures), it should be easy to handle on different phones (no long dropdowns, no scrolling) and every item should be shown on an extra page (no scrolling). Surveys which are constructed for mobile phones always work quite well on desktop computers, which is not the case the other way around.

---

[1] http://www.peterlugtig.com

Also Florian Keusch[2] emphasised that the increasing popularity of smartphones poses new challenges for researchers but also opens up new opportunities for novel ways of data collection. For example, a rising share of respondents access web surveys that were originally designed to be taken on large screen computers, on their smartphones. Researchers are now also able to collect data from smartphone users – such as geolocation, online behaviour and browser history, app usage – through passive mobile data collection via apps. Keusch outlines also that compared to surveys that rely on self-reported data, passive data collection has the potential to provide richer data (because it is almost constantly collected), to decrease respondent burden (because fewer survey questions need to be asked), and to reduce measurement error (because of less forgetting and social desirability).

Also the input of Albrecht Wirthmann from the Big-Data- Working Group from Eurostat and Ole Mussmann from CBS Nederland at the "Task force on Innovative Tools and Sources for TUS" showed that with sensors new ways of collecting time use data are possible. Both mention to be transparent with privacy issues. However, it will be necessary to get the agreement of the respondents to collect passive data from smartphones.

3.2.2   Commercial Time Use Apps

On the free market or better in the app stores, time use apps are available to enhance people's everyday life. Often these apps are constructed for work-related-activities but also to get a better overview of one's personal life. Some apps track the amount of time a person uses for different apps. Other time use apps emphasize a fitness and health approach. They track the GPS-location and the speed and pace of steps. Most of the apps have a function for starting and ending the activity. Another feature provided by all of the apps is to visualise the amount of time of the different activities. So the app gives the participants feedback regarding their own behaviour. These features have the potential to also increase the motivation to participate in 'official' time use surveys.

Table 3: Screenshot Toggle –App

Source: https://play.google.com/store/apps/details?id=com.toggl.timer

---

[2] https://floriankeusch.weebly.com/research.html
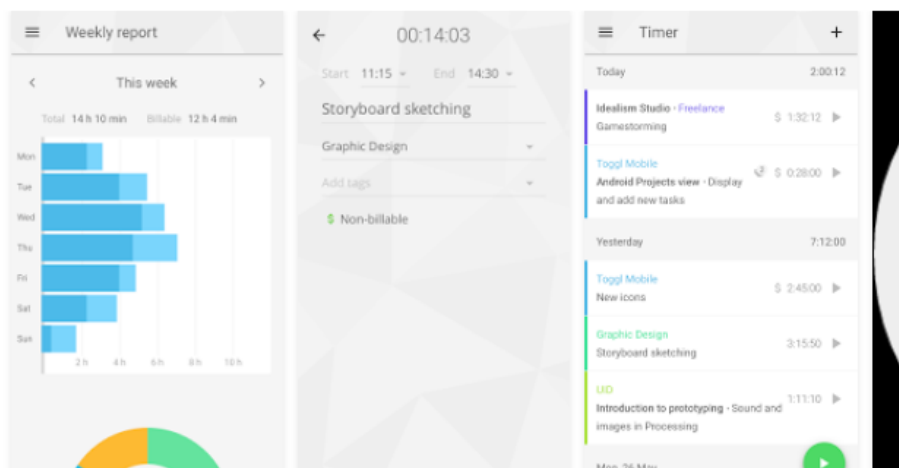
Table 4: Screenshot of Timley App

Source: https://play.google.com/store/apps/details?id=com.timeapp.devlpmp&hl=en
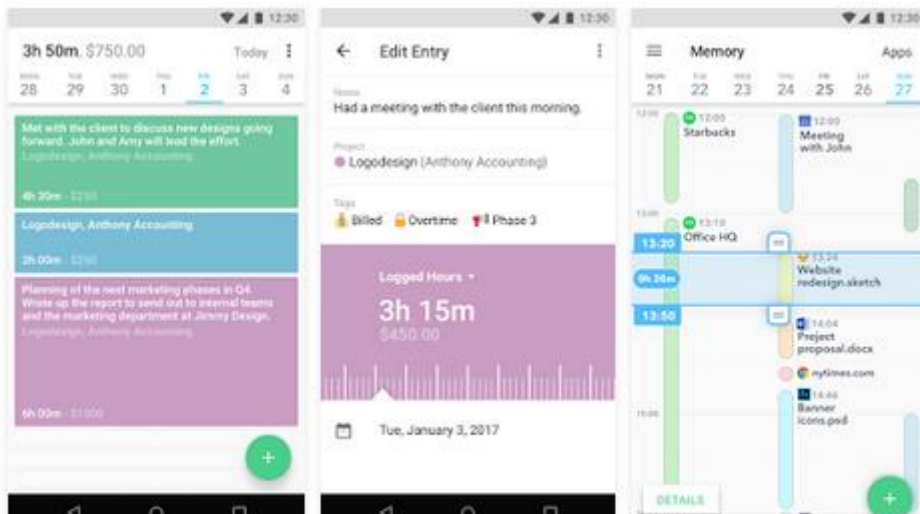
3.2.3 Time-Use-Web-Diaries

In the report "Using smartphones in survey research: a multifunctional tool" (2013[3]) of the Netherlands Institute for Social Research it is described in detail, how the app was constructed and tested. There was also a subjective well-being item implemented with the beeping method. The respondents were able to use the app as a tool to record their time use in a mobile diary and indications were found that the data quality is in line with previous time use studies. Some participants reported problems with finding the relevant activities or considered the time intervals as too short. Another important issue related to smartphone surveys was data privacy and representativeness. In the sense that not everybody has access to these devices or does not want to give all the personal information.

In Belgium the University Research Group TOR[4] has established a modular time use research infrastructure called "MOTUS". This survey-infrastructure gives the possibility to ask a personal adopted questionnaire and to include a web-diary with further plug-in-features i.e. GPS. One of the benefits of MOTUS described by the research group is related to controlling the quality of the time-diary and all elements of the survey itself. The administrative features are also adjustable. To reduce the unspecified time, it is possible to use a starting and ending time of the activity. All respondents are asked to provide the context in which the activity took place, i.e. who else was present. One important issue they mention is that personal contact is important. There is a need for a person who explains the aim of the survey and guarantees anonymity, and immediately answers the most pressing questions. The assumption is that this might offer more confidence than a letter with an encrypted username and password especially in the more burdensome time-use surveys.

---

[3]

https://www.scp.nl/english/Publications/Publications_by_year/Publications_2013/Using_smartphones_in_surv ey_research_a_multifunctional_tool

[4] http://www.vub.ac.be/TOR/news

14

The research about new time use online-diaries based on a web browser or an app was challenging. There are only few detailed documentations. On the task force for innovative tools at Eurostat, Hungary presented that a web-diary is planned for measuring time use. It is for sure informative to learn from their experiences with this tool. For the future developments, an expert-survey with a questionnaire regarding time use innovations would be productive and give more information.

3.2.4   Open source project "Tabi" [5]

In November 2017 the Tabi project went online. The goal of this project is to collaboratively develop an app for data collection regarding time use. As the developers say it is open source under the MIT license. This means that everybody can use the app and it's backend, including also national institutions. They want to target especially the needs of (national) statistical institutes.

The idea to share experiences makes sense and even if the code cannot be shared between the national institutes, know-how can be shared and this would save a lot of time. This project should be observed for the upcoming developments.

---

[5] https://tabi.gitlab.io

4 Overview of Subjective Well-being (SWB)

4.1 Why measuring SBW?

Measuring *subjective* well-being within time-use diaries can be done with two different aims in mind. Well-being items can be included to foster well-being research in general but it can also be used to answer specific time use questions like: What do certain activities mean for people conducting them? What activities are to be considered as leisure? Which have to be counted as reproductive activities? How might well-being measures help to answer these questions? The latter will be discussed in chapter 4.1.2. In the following chapter 4.1.1, however, we are going to examine why well-being items should be included into time use surveys to learn more about subjective well-being.

4.1.1 Why measuring SWB within TUS: from a SWB-research point of view

There are different reasons why measuring SWB within TUS could be beneficial for time use research: maybe the most prominent point is that evaluated time use (which could be described the micro measurement of well-being)- becomes increasingly prominent as a measure of general well-being. Micro-measurement of time use might however also help to overcome or at least counterbalance biases connected with the happiness paradox. In a wider frame measuring SWB within TUS can also help to entangle the role of context for (micro-) well-being and – as a an input for economic Theory and research- to clarify the concept of utility and the relationship between utility and time use.

Evaluated time use as a component of SWB

Subjective well-being is a topic of increasing importance (and therefore research efforts) not only in fundamental research but also from a policy point of view, in particular after the recommendations given in the well-known and influential Stiglitz-Sen-Fitoussi report (2010, see also United Nations Secretariat, 2012). In the discussion about GDP as a single measure for progress and wealth, well-being measures gained substantial importance. A full *replacement* of the GDP by happiness indexes such as in Bhutan is not intended in Western countries, but assessment of hedonic status and including the results into national reporting has become standard in the EU, the OECD and the UN (Bache & Reardon, 2016).

Though Kahneman, Diener and Schwarz admit in their preface (p. xi) to their Well-Being book (1999) that knowledge gains about hedonics might not immediately trigger national strategies, they express their hope that it "will be relevant to policy", in particular when scientific understanding of hedonics will have reached higher levels and may more efficiently support GDP-based evaluations. We believe it is fair to say that this journey has begun, but that it still pays to consider the methodological questions on a very fundamental level.

The two probably most influential theoretical foundations of well-being assessment, subjective well-being and objective happiness, treat the ongoing hedonic state as one of the basic pillars at least, or even as the basic fundament. It is a truism that the qualities of our lives are related to how we feel when we live them, thus evaluated time use could be considered as the most natural way to assess QoL (Krueger et al., 2009, call it the "currency of life"). However, this is not the only approach to evaluating our living conditions: other, more globally oriented popular notions for discussing "how life is" are, among others, life satisfaction, well-being (in particular subjective well-being, but also material, psychological, Ryff, 1989, or social, Keyes, 1998), flourishing (Keyes, 2002), happiness, capabilities and functionings (Sen, 1984). Diener (2000) defines *subjective well-being* as a positive

cognitive evaluation of life and the presence of positive and absence of negative feelings, together with satisfaction in important life domains. Though life and domain satisfactions are known not to be constant over time, and assessed life satisfaction may be triggered by momentary influences (Schwarz & Clore, 1983), they express in principle rather steady components of life, whereas affective experience and emotions happen in time and have a strongly temporal character. In Diener's SWB concept, they are just one component, in addition to cognitive evaluation, and also an input to the latter. Thus, strictly speaking, life satisfaction has in fact a dual function: as global or final judgment and as (just) one of the many streams of hedonic state or as one of several contributors to the hedonic balance.

In contrast, the *objective happiness* approach (cf. Kahneman, 1999) focuses on the quality of timely experience directly and considers the total amount of happiness as the temporal integral of momentary happiness over time. The term *objective* expresses – admitting the subjective character of happiness in principle – that the aggregation of momentary experiences into a single value is not based on an arbitrary, idiosyncratic combination function, but by a clearly defined rule which is the same for all persons. Leaving the way of aggregation to the individual would open the door for memory biases, personal positivity or negativity biases, primacy or recency effects, individual weighting tendencies of positive or negative peak experience, and so on. However, assessing objective happiness in the strict sense would require a continuous recording of momentary felt affect, which is of course not possible but has to be approximated by instantly filled in diaries, day reconstruction or experience sampling (see below).

For assessment purposes, Kahneman and Krueger (2006) propose an elegant solution to simplify the challenging integration task: the U-index which equals the share of time which was spent in an unpleasant rather than pleasant state. It has obviously desirable scale properties (avoiding the discussion of the measurement level of subjective ratings), and poses less difficulties in quantifying than rating scales. The price is a potential loss of information, since only a dichotomous decision about the hedonic balance is provided. In particular, negative episodes may fail to be reported if their effects are not extreme or considered bearable. In addition, it should be noted that the standard problem of judging hedonic states numerically is still present, namely, where exactly respondents see the borderlines between "rather pleasant" and "rather unpleasant". This limits the attractiveness for application in regular time use surveys, in particular in a non-clinical context.

Certainly, the objective happiness approach is more strongly related to time use data than the subjective well-being one, since the global judgment is replaced by the time-dependent data, as a bottom-up approach. Kahneman cites Zajonc (1980) as authority supporting the idea that judgment is not necessarily a cognitive process at all: examples show that merely affective evaluation can bypass or at least be faster than a cognitive one. In fact, most persons would probably not deny that there is a continuous sensory hedonic evaluations but they would disclaim that they are constantly thinking about how they feel.

Recording evaluated time use provides a level of detail which allows for investigating other popular scientific topics, apart from an aggregation to a global judgment, for example the tracking of *flow*, i.e. peak experience in the meaning of Csikszentmihalyi (1975): the pleasant state of being captured by the current activity, a perfect balance between challenge and individual resources, which makes people forget about time. On the other hand, and leaving the field of positive psychology, also pain or other unpleasant feelings happen in time, such as stress which stands in perfect contrast to flow: a

straining imbalance between the perceived challenge and the perceived resources (in the meaning of Lazarus, 1966). So evaluated time use is not only a key to well-being, but also to ill-being, work satisfaction, therefore also to economic productivity (Harter, Schmidt, & Hayes, 2002), and even to clinical phenomena such as pain, anxiety or depression.

Overcoming biases

As Cummings and Nistico (2002), but recently also Ponocny et al. (2016) point out, global life satisfaction ratings are prone to bypass negative circumstances in life, which constitutes a "positivity bias" or a "happiness paradox" (Staudinger, 2000). This is understandable, considering – as mentioned before – that a judgment about oneself is not only consequence of affective experience, but at the same time an input: considering one's own life as bad will also invoke negative emotions and may therefore be avoided.

Thus, the subjects' use of the scale may not only lead to biases, but is in general a complex judgmental process with many unobservable steps, including individually different concepts of the notions, the expectations, and the anchor values (Schwarz & Strack, 1999). However, as Schwarz and Strack mention, some of the difficulties are mitigated in diary studies: The accessibility of emotions is easier if related to a specific time (Schwarz, 2012), with less cognitive processing interfering. As an example, social desirability regarding the hedonic value of children will be very salient for global judgments, but probably much more relaxed when talking about a specific day or hour.

Other well-known effects refer to the strong influence of novel events, the influence of the current state of mood, and heuristic effects such as order effects and peak-end-evaluation (Schwarz & Strack, 1999).

The role of context

In spite of the vast amount of well-being assessment, little is known about the concrete causal paths leading from living conditions to hedonic judgments, and the correlations between observable circumstances and subjective life evaluations keep staying small (Argyle, 1999). In other words: individual subjective well-being is hard to predict from other knowledge about that person. From the response, it can hardly be concluded which aspects of life are addressed to, and with which benchmarks they are compared (for an overview, see again Schwarz & Strack, 1999). Time-use surveys provide valuable background information about the actual life of the respondents, and may help to understand where people's ratings come from, by the "simultaneous assessment of contextual and experiential variables" (Schwarz, 2012, p. 37). However, they cannot replace global judgments such as usual single-item questions about life satisfaction, since it does not need to refer to immediate daily activities. Social status, worries about the future, grief, regret of what has happened, counterfactual thoughts, anticipation of positive events, or similar, will not fully manifest themselves in diary content where the awareness will be centered on main activities rather than on thoughts coming along. But at least they will help isolate the activity-grounded hedonic experience from the other. By this, evaluated time us will facilitate the search for potentially improvable circumstances, and identify main drivers of mischief, such as frustration during work or loneliness.

DeLongis, Coyne, Dakof, Folkman and Lazarus (1982) conclude that health is better predicted by everyday hassles than by major life events; in addition, stressors have been shown to take effect via everyday events (Pillow, Zautra, & Sandler, 1996).

Context information also helps to better judge within-person variability, due or not due to context variables, and therefore the dependence on time or timely correlated circumstances. Knowing about intra-individual variability, on a micro-level, may also help understand phenomena in the population and the variation there. Some startling puzzles in SWB research may be overcome on the detailed level of evaluated time use, knowing how the people actually spend their time. For example, a comparison between time could give insight why adaptation to a change in life situation takes place, should it be related to a shift in time use patterns.

Addressing economic considerations: the utility approach

It is a common view in the social indicators movement that the GDP within a country is only a weak indicator for the utility of living conditions, but so would also be life satisfaction or happiness ratings on a too general level. In fact, the notion "utility" is more complex than that and has components which should be carefully separated. Kimball and Willis (2006) provide an overview of the relation between utility and happiness: Economic theory uses the term utility for understand human choices, so that it rather expresses anticipated hedonic consequences, which is in fact a *predicted utility* or *decision utility.* Happiness, in contrast, expresses actually felt *experienced utility*, whereby Kimball and Willis discriminate between short-term ("elation") and long-term ("baseline mood"). Clearly, evaluated time use will focus on the momentary, current affect, the *flow utility*, which Kahneman (1999) calls *instant utility*. Global judgements of life, such as questions about life satisfaction or global happiness, on the other hand, refer to *lifetime utility*, which respondents will judge on the basis of *remembered utility*.

When we evaluate the hedonic value of time periods, one may be tempted to assume the utility arises by how one spends this time. This is tautologically true, but not necessarily in the meaning that the affect should be attributed to what actually happens *during* this time or what kind of activity is conducted which would be a causal inference. (Certainly, the current events are still involved to the degree that they at least *enable* the enjoyment of previous or past events). It may as well be that people are still happy for what has happened before or will happen after, or for possibly disregarded side conditions such as the weather, and is hardly related at all to what we are doing. We will adopt these points later on, but drop the term *process utility* here for the case where the hedonic level *is* due to what we are doing.

Concrete economic applications of time use surveys refer, among others, to non-paid work, such as household and care, to health economy (physical exercises and sleep), to consuming and demands (eating, leisure activity), or work-life-balance (Gershuny, 2011). If time use is evaluated, this adds non-monetary utility (input and output) information to merely economic estimates of produced value. For example, the value of care investments cannot only be judged by costs on the input side, but also by hedonic consequences as the output. As another example, GDP could increase when at the same time process utility might decrease. Gershuny (2011) gives examples for the actual estimation of marginal utilities, such as non-TV leisure, care and sleep.

At the end, there have even been proposals for national accounts based on time use (Kahneman et al., 2004, Krueger, 2009) based on the U-index as described above (in the part about evaluated time use).

Time use surveys have also been used to calculate extended versions of the GDP, enriched by contributions due to unpaid work, which show a larger contribution by women ("women's GDP")

than the traditional GDP does (see Schaffer & Stahmer, 2006, Franz, 1996, and Keuning & Timmermann, 1995). They are not based on evaluated time use; however, including the hedonic component could help judging whether the difference in economic participation also corresponds to one in hedonic outcomes.

4.1.2 Why using well-being to enhance quality of TUS data?

Quality –enhancements can happen on different levels. A deeper knowledge on the subjective quality of time use enhances our possibilities for analysis and also analysis of data quality in general, but it also helps us to gain specific insights into the field of care-and housework and division of families in general.

Meaning of activities

There is an ongoing debate about what it is we want to measure in time use studies: activities how a non-participating observer would record them or activities how they are perceived and described by those who are active (or not active and at rest). Is it the activity which we are interested in or the meaning of this activity? If participants have the opportunity to state the activity in their own word, these words will (at least in many cases) contain also information about the meaning of an activity. If we only offer restricted answer categories, information about meaning will in many cases get lost. Asking for well-being and/or the subjective evaluation of a situation could help to gain information about the subjective meaning of activities.

Identifying non-paid work

Imagine the following situation: You want to use TUS data to find out if men do more reproductive (non-paid) work than women. Therefore, you select certain activities as cleaning, doing homework with kids, washing the elderly etc. as reproductive work. You would also consider cooking to be reproductive work but then you stumble across a few diaries where some women cooked at the weekends literally for hours. Maybe one of them cooks meals for the whole week to put them in the freezer and reheat them when she has no time for cooking. But maybe cooking is one of her hobbies and the cooks for hours because she invited friends and that is what she enjoys to do at the weekend: preparing one very special dish after the other while her husband cleans the kitchen and the rest of the house and tries to prevent the kids from disturbing her. In the latter case she surely enjoys what she is doing, in the former case probably not that much. And this is where subjective well-being measures come into play. They can help us classify certain activities and give us more information about the personal meaning of these activities for the respondents.

On the other hand, it should be kept in mind that activities can`t be classified by well-being measures alone. Classifying activities as leisure only because many people enjoy doing them would be misleading. Work (paid or unpaid) has to be considered as work regardless if it is enjoyable or not. It also has to be taken into account that enjoyment, satisfaction and happiness don`t have to be caused by the actual activity but might also be due to certain other circumstances, such as people present or more general emotional states.

One of the examples for different evaluations of time spent on certain activities and the corresponding interpretations can be found at Krueger et al. (1999). In their Table 1.10 they demonstrate such differences regarding the perception of seemingly equal situations: Spending time

with the boss is less pleasant for men than for women, but vice versa for spending time with children, since for women it rather means caring and household.

Fairness in general

The expression fairness does not only refer to the amount of time spent for work (in general or in relation to the disposable income) but might also be used in the context of time spent for enjoyable activities. Chores and resources could be viewed as distributed fairly within a household, when everyone has the same amount of burdensome and enjoyable things to do. Or maybe when the "resulting burden" or the general utility (as referred to in the last chapter) are the same. If one wants to analyse the division of labour and also the division of resources to be spent on leisure activities from this perspective one does not only need information on activities during a day but also on the subjective quality of this activities. This information can only come from a time use survey that also asks for the subjective quality of activities and therefore for well-being in a given time period.

Meeting needs of respondents

Our experiences with earlier waves of the time use survey and with other diary-studies show that participants like to express themselves and to give detailed information about aspects of their lives, which are relevant to them. Because of this they normally prefer free answer formats instead of predefined categories for activities but also provided in some cases quite detailed information about activities and circumstances. Asking for subjective well-being and thereby for the subjective meaning of activities might be a good opportunity for respondents to tell us more about their lives and thereby be more satisfied with their opportunities to express themselves.

4.2   How to measure subjective well-being?

A review of different time use surveys shows that there are nearly as much ways to measure well-being and enjoyment as there are time use surveys. Diaries can vary according to the place, where questions about quality of time and activity are asked (at the end of a diary or within the diary itself), in the extent in which the activity itself is addressed (in contrast to the time period) and in the question mode (additional columns in the diaries or ad hoc questions – *push messages* – via apps).

4.2.1   Where to ask for subjective well-being?

In some countries as in Norway, Germany and Finland time use surveys contain only one or a few questions regarding enjoyment of certain activities during the day (see table 5)

Table 5 Enjoyment of activities: general questions at the end of the questionnaire

Finland	When you think about the activities you have entered in your diary, which of them do you enjoy most?
Germany	Of all activities you recorded during this diary day: Which was your greatest pleasure? 2. For which activities do you wish to have had more time? 3. Of all activities, you recorded during this diary day: Which one did you enjoy not?
Norway	Were there any activities during the day that you experienced as particularly positive? What was the most positive and at what time of the day was this activity? 2. Were there any activities during the day that you experienced as particularly negative? What was the most negative and at what time of the day was this activity?

Including questions like this at the end of the dairy is a simple and economic way to gather some information on enjoyment. But as in most cases, only the most enjoyable category is addressed (and in some cases the most unenjoyable). There is no information on enjoyment of different activities and well-being in general which leads to the conclusion that enjoyment or subjective well-being should be assessed for every activity/every time slot in order to answer questions like those formulated in 4.1.

4.2.2   Time slots vs. push messages (beeping method)

As the surveys mentioned in the last chapter only assess enjoyment of certain activities, methods where well-being or enjoyment is assessed within the diary capture the whole day and thereby all activities during the day. This method, chosen by France, Italy, Poland and the UK (see also Table 6) covers all activities/time slots during a day and is therefore very comprehensive but at the same time very demanding for the respondents "Asking respondents to record how happy, sad, stressed or irritable they feel during each activity in the diary would definitely make them feel tired, irritable and rushed."(The Netherlands Institute for Social Research, 2013, p.21). Respondents' burden should be increased significantly. It is also not sure if respondents are able to report feelings and emotions for specific time periods.

From the view of well-being research, the experience sampling method (ESM) which is the technical/ scientific term for the beeping method is the via regia to gather information about momentary well-being. In this method the respondent get text message/push-messages (with an additional sound when the mobile phone is not turned mute) on their phones (or other devices) at random time points during the day and are asked to answer one or a few questions about their current well-being. In the Netherlands this method was tested in a feasibility study at the Netherlands Institute for Social Research (2013). Respondents got 6 "beeps" (push- messages) during the day they could answer within a certain period of time. Otherwise they would disappear again. In this pilot peep messages could as well be ignored. However, respondents should be of course encouraged to do so only, if it´s absolutely not possible or unsafe (driving!) to answer the messages.

Six messages every day could be quite burdensome or even annoying as well, bit results of this first pilot show, that respondents didn`t suffer that much or at least did not complain (Anne Elevelt, personal correspondence). However, in choosing the number of push messages on a certain day one not only has to think about respondents' burden but also about the number one needs to cover different types of activities to a sufficient amount. If one has only a small sample one certainly has to ask more often during the day to cover all relevant forms of activities.

4.2.3   Activities vs. time slots

In both cases, when asking for time slots and when using the beeping-method, one has to think about the scope of interest and choose the questions accordingly: are we interested in the evaluation of certain activities or in mood and well-being during a certain time span. However, in some cases it´s not easy to draw a clear distinction between those two.

Following Kahneman's objective happiness approach, it is clear that momentary judgments should evaluate the current emotional status. But in a usual time use survey with an evaluation column, the awareness of the respondents is at least in part drawn to the activity with which they spend the current time, or researchers might be tempted to interpret the evaluation of the time as an

evaluation of the synchronous activity. Although this will coincide in many cases, it is not necessarily the activity which causes the emotion.

This point seems not very well reflected in the literature. At least implicitly, evaluations of time spent during an activity and of the activity itself are somehow identified with each other (cf. Pääkkönen, 2013), when publishing tables or even rank orders about the most pleasant or unpleasant activities. This is, to a certain extent, even true for Krueger et al. (2009), who contrast immediate rating of *activities* to evaluated time us *during activities*, but still use formulations such as "Americans find child care substantially more unpleasant than do the French" (p 74) when in fact well-being "during" an activity is addressed.

But precedent or upcoming events may outweigh the hedonic value of the current occupation, or – in more seldom cases – contrasting emotions may not extinguish each other but exist at the same time. A near at hand example would be physical pain but joy about the beauty of nature during a hiking tour, or shock and relief after a near accident. But the difference may root even more deeply, mapping the well-known discrimination between hedonic and eudaimonic aspects: learning for an exam might be exhausting or boring, nevertheless one would judge the activity as successful and therefore positive when the desired progress has been achieved. Interpreting evaluated time use as current *hedonic* affect will be flawed, then, since the rating is closer to a cognitive satisfaction then than to pleasure or enjoyment, judged by the criterion whether the respondent is happy when the activity is over. This is another version of the well-known *correlation is not causation* situation: the activity is not the only possible reason for the feelings during the activity (though probably the most important one in the majority of cases), and activities may well be confounded by other influencing factors.

The discrimination between enjoyment of time and satisfaction with time use should be more salient for work and other compulsory or extrinsically motivated activities, but inter-individual variation can be expected for leisure time ratings as well. Krueger et al. (2009) summarize the reasons as follows (p. 11): pleasure can arise from other than leisure activities or be quite modest for certain individuals, changes over time, heterogeneous emotional experiences, and multidimensional emotional responses. The role of cultural and socialized influences must not be underrated: Situations signaling some kind of success (even dining together may function as indicator for success in family life), rituals, or festivities may exert some kind of institutionalized well-being – for people with referring learning experience. In general, activities can produce extrinsic values, or may generate enjoyment directly in the meaning of process benefits (cf. the discussion about the different types of utility in 2.1.1 above). In a similar vein, Kubovy (1999) contrasts bodily pleasures vs. pleasures of the mind, which are considered to arise from a cumulation of emotions.

What can be said: wording of the item should follow the mayor aim of inclusion of a well-being item. If one wants to foster general knowledge about well-being one would probably focus on quality of time whereas if one wants to learn more about what activities might be part of the leisure time one might want to focus on pleasantness of activities. How did other time use surveys handle this question?

Table 6 Overview of SBW-questions

France	Was that moment pleasant or unpleasant?
Italy	Is this a pleasant moment?" rank the feeling on a scale from -3 to +3, with -3 being "not pleasant at all" and +3 being "very pleasant"
Poland	Was your time spent on main activity from col.2 generally pleasant/nor pleasant, nor unpleasant/ unpleasant?
UK	Diaries include an enjoyment column after the final who else was present column

As Table 2 shows, two countries explicitly focus on time ("moment" whereas we doubt that this is an adequate translation). Poland chose a mixed version "time spent on activity" and UK probably focused on the activity (enjoyment) but the exact wording wasn`t stated.

4.2.4   What questions to ask

After one came to a decision if to ask for the evaluation of the activity or well-being in a current time span (or both) one still has to decide how to ask for it and what items to use. Whereas there seems to be a widespread consensus how to ask for well-being on a global level (life-satisfaction is the prominent measure here) there is much more variety in the measurement of momentary well-being.

One way to ask for momentary well-being is to ask for emotions. However, the mere heterogeneity of positive and negative experience requires compromises regarding the grade of detailedness in general, but in particular in the time use survey context where people are expected to rate the hedonic status more than once. It is impossible to assess all different aspects of good or bad affect simultaneously, including emotions such as anger, stress, disappointment, coping, self-esteem, etc. (Krueger et al., 1999, ask for being happy, tired, stressed, sad, interested, and affected by pain). Nevertheless, we usually can feel a net affect which is more or less positive or negative, again referring to Kahneman's (1999) idea of an index describing the time of dominantly negative affect. Although it is clearly known that the good vs. bad-quality of experience is not unidimensional, it is normally not the case that marked good and marked bad emotions are felt at the same time (see also Kahneman, 1999, p.8). However (p. 11), there is some evidence that at least the zero point between "good" and "bad" is not arbitrary but comparable between persons (which does not apply to unipolar rating scales, as mentioned above).

Besides classical emotions like happiness and sadness there are some which are especially interesting in the context of time use. When an assessment of time use-related burden is intended stress seems to be an appropriate variable since it has a high chance to change with a frequency as desired. Additionally, it is relevant for mood (see again DeLongis et al., 1992, about the role of daily hassles) and widespread, also in Austria where a quarter of persons confirmed living a stressful life (Ponocny et al., 2015), and it is by its character closely linked to what happens during the day.

4.2.5   Scales

When applying rating scales, the number of categories and their anchoring has to be chosen. In particular, they may be formulated as unipolar or bipolar scales.

Regarding the number of categories, the seminal paper by Kahneman and Krueger (2006) as well as the American Time Use Survey apply a 7-categories scale, which is a well-arguable choice following

the standard arguments: middle categories are avoided which may be taken as mental effort-saving alternative by not forcing a choice between rather positive or negative, and 7 may be considered as a compromise between over-differentiation and providing too few response options (see also Chapter 5.1). Ponocny et al. (2016) confirms the difficulties to discriminate between the meanings of an 11-point life satisfaction scale. On the other hand, the skewed shape of many well-being responses restricts most of the responses to the right half of the scale, so that fewer than three categories there would reduce the potential information drastically.

Regarding hedonic evaluations, unipolar scales starting with, for instance, "not happy" are not recommendable since it does not make clear whether "not happy" stands for a neutral or a negative state (probably neither to the respondent nor to the researcher), so end points such as "unhappy" vs. "happy" will be better understandable. Labelling all categories (and not just the endpoints) seems to make the responses better understandable at first sight, but does not necessarily help when the verbal expressions do not add really tangible information about the categories (e.g., agree "a bit", "mostly"). Large-scale surveys often skip them when it comes to subjective well-being questions (as an example, the European Quality of Life Survey, Eurofound, 2005).

5 The MODUL study of living conditions

In the last chapter some problems in the measurement of subjective well-being were mentioned as scales, activity classification and the connection of circumstances and emotions/well-being which were also addressed in a project titled "Living conditions, quality of life, and subjective well-being in regions: A methodological pilot study with explorative interviewing and quantitative measurement" (Ponocny et al., 2015) which was funded by the Anniversary Fund of the Austrian National Bank (OeNB). This chapter on the results of the MODUL study of living conditions presents selected results concerning those three topics.

5.1 Problem: Scale/distribution

Typically scales ranging from 1-10 are used for overall life satisfaction and happiness ratings (e.g. "1 – very satisfied" to "10 – very dissatisfied", and "1 – very happy" to "10 – very unhappy"). Independent from its resulting problematic distribution it is still the most typically used range in the literature due to advantages that arise from a constant measurement over time, especially for longitudinal or panel studies.

Overall satisfaction and happiness ratings were asked for twice at 10 different locations in Austria, in a paper-pencil questionnaire and in a short socio-demographic questionnaire handed over to interviewees before face-to-face interviews were conducted.

The questions on the questionnaire were formulated in the following way:

- "How satisfied are you with your life overall?" ('1 – extremely dissatisfied', '10 – extremely satisfied')
- "How happy are you with your life overall?" ('1 – extremely unhappy', '10 – extremely happy')

From the two histograms below (1,272 respondents for the overall satisfaction, and 1,261 respondents for the overall happiness) the negative or left skewness is obvious. It becomes clear that neither the satisfaction nor the happiness ratings of respondents are measured in a statistically feasible way. Mean values drift towards the more positive side of the scale whereby the occurrence of less positive ratings are underrepresented.

How satisfied are you with your life overall?

How happy are you with your life overall?
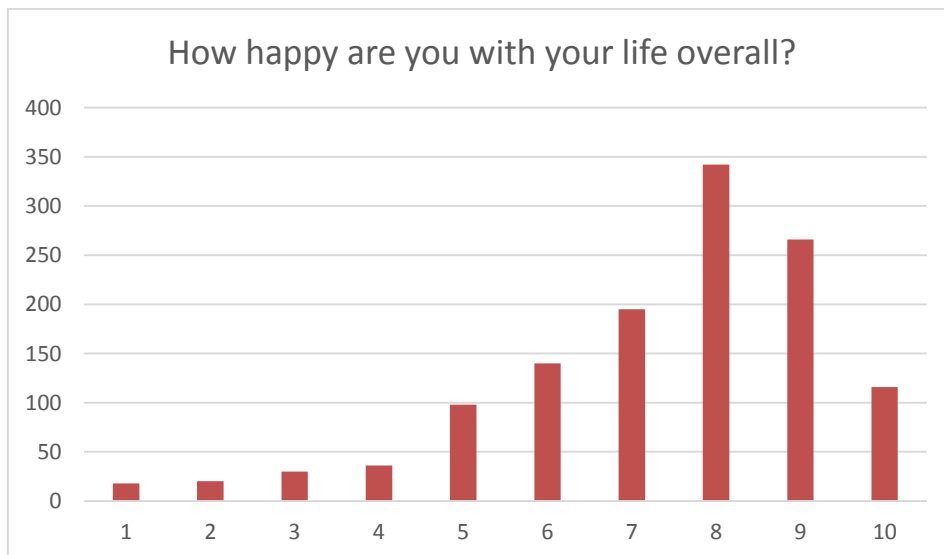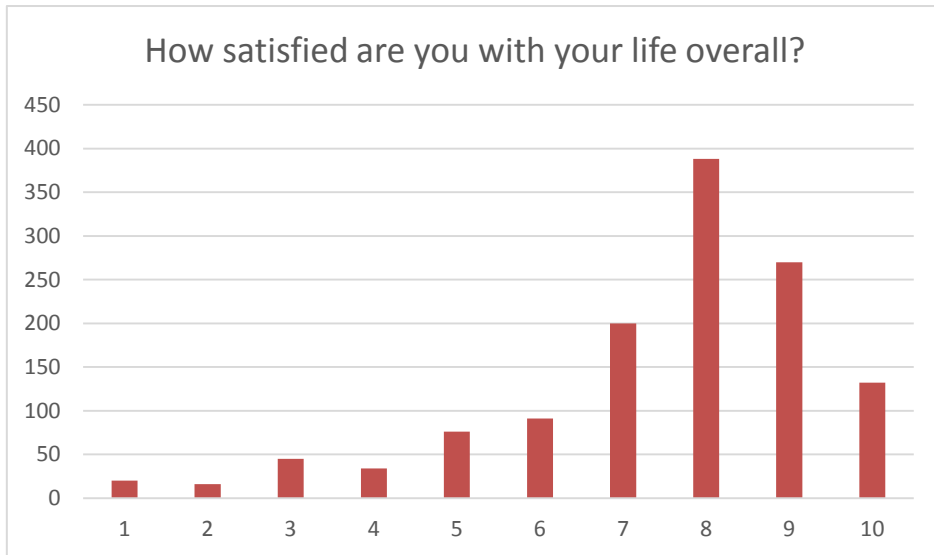
The same is true for the two overall satisfaction and happiness rating questions in the socio-demographic questionnaire completed in course of the semi-structured interviews. Hereby the poles of the two scales were defined in a different way compared with the formulation of the two questions mentioned above:

- "How satisfied are you with your life overall?" ('1 – very satisfied', '10 – very dissatisfied')
- "How happy are you with your life overall?" ('1 – very happy', '10 – very unhappy')

Distributions of 501 (overall satisfaction ratings) and 499 (overall happiness ratings) respondents are even worse compared with the ones above. It becomes clear that formulations like "extremely" instead of "very" have a positive impact on the usefulness of the scale.

How satisfied are you with your life overall?

How happy are you with your life overall?
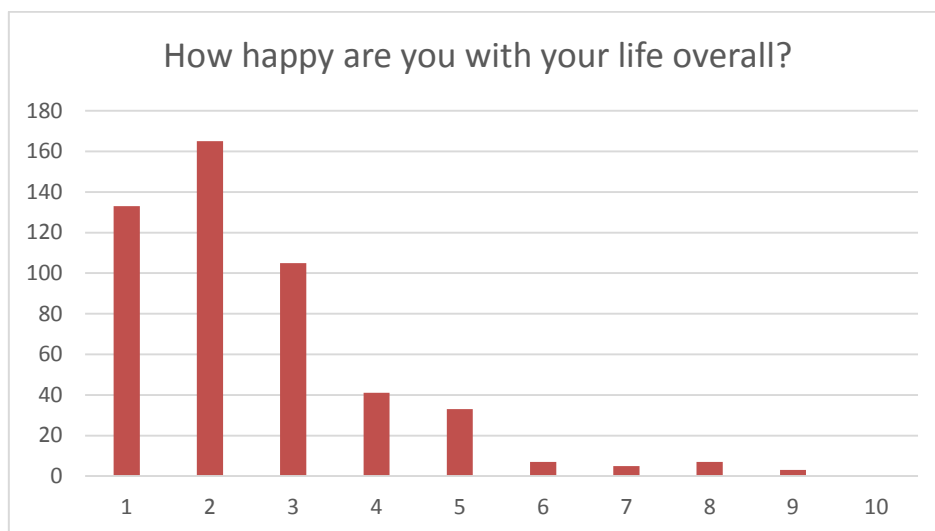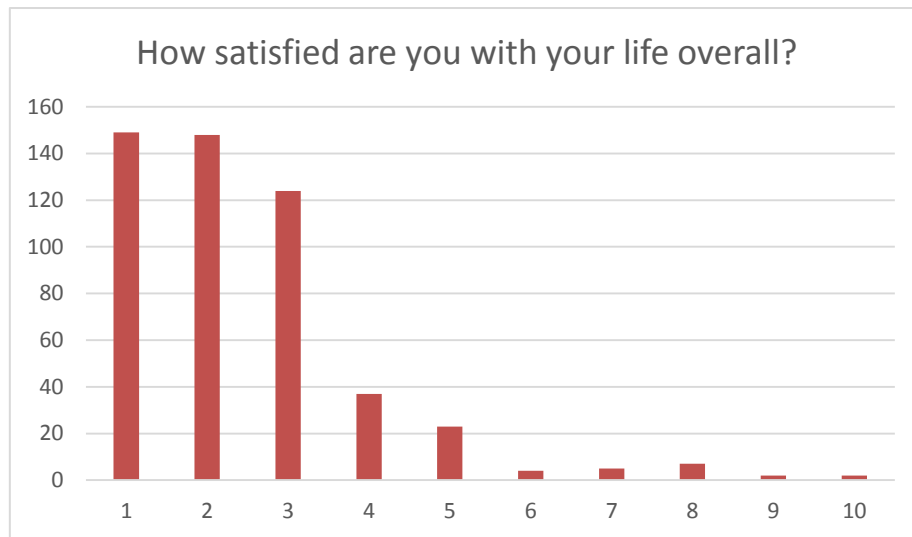
The four histograms above present the overall life satisfaction and happiness respectively. But diary evaluations tackle much narrower time frames. A diary example of a single day as observed in the MODUL study of living conditions is given below. Initial respondents were randomly selected by making use of a list of inhabitants received by the community's mayor and out of telephone books. Further respondents were collected by means of snowball sampling starting with the initial respondents derived from the previous step. Overall, 341 respondents completed the 48 half-hourly evaluations over the period of 7 days. Apart from various different information tracked in the diary (e.g. time, location, accompanying person, activity, and commented feelings), respondents numerically evaluated their feelings on the following question "I feel…" on a scale from '1 – very good' to '10 – very bad' every half hour over the period of one week as well. Aggregated half-hourly ratings over all respondents reveal a clear dominance of values towards more positive evaluations. Therefore, the same problem as discussed above becomes clear for the diary study as well.

Summarized, a narrower scale (ranging from e.g. 1 to 7) with extreme poles is recommended [e.g. The American Time Use Survey (ATUS) makes use of a 6-point scale ('0 – not present', '6 – very strong') to rate the strength of the following six feelings: pain, happy, tired, sad, stressed, and meaningful.]

Table 7

Date			Rating for whole d			Was this day … ?			a normal day (work, school,…)	If unusal, why?		1	holiday (weekend, off duty, no school,…)	
21	7	2012	"1 very satisfied"			(please check)		1	an unusual day			1	sick	
Day	Month	Year	"10 very dissatisfied"		3								on vacation	
													other reasons	

Time	Activity (sleeping, shopping, working, …)	Duration	Where am I? (at home, in cinema,…)	Duration	Who else is present? (nobody, partner, children, colleagues,…)	Duration	I feel: "1 very good" "10 very bad"	Duration	Indicate why you feel like that. (What is the reason for this feeling?)
0:00-0:30	going clubbing	begin	revo		9 friends		4		totally worn out and tired
0:30-1:00									
1:00-1:30									and it rains when I walk home alone
1:30-2:00									
2:00-2:30									
2:30-3:00									
3:00-3:30									
3:30-4:00									
4:00-4:30									
4:30-5:00									
5:00-5:30									
5:30-6:00									
6:00-6:30	sleep								
6:30-7:00									
7:00-7:30									
7:30-8:00									
8:00-8:30									
8:30-9:00									
9:00-9:30									
9:30-10:00		end							
10:00-10:30	read in bed		at home	begin	alone		5		feel sick
10:30-11:00									
11:00-11:30									
11:30-12:00									
12:00-12:30	family crisis talk				Mom, Dad, Partner		7		everybody said or angry, tell about sorrows
12:30-13:00									and what annoys them; miserable mood
13:00-13:30	eating				family		2		but at the end everything is worked out
13:30-14:00	chatting				family		2		good mood at lunch, but feel sick still and taste hardly anything
14:00-14:30									
14:30-15:00	chatting				friends		1		
15:00-15:30				end					all cheer me up to make me feel better
15:30-16:00	playing the piano		on the roof		alone		1		
16:00-16:30									good music, reflecting and relaxing
16:30-17:00	upload fotos on computer		at friends	begin	alone		1		remembering Spain, many funny fotos
17:00-17:30									
17:30-18:00	painting	begin			alone		1		totally relaxed
18:00-18:30									
18:30-19:00									
19:00-19:30		end							
19:30-20:00									
20:00-20:30									
20:30-21:00									
21:00-21:30									
21:30-22:00	watching TV documentary				alone		2		very interesting, but I am tired and worn out
22:00-22:30									
22:30-23:00									
23:00-23:30									
23:30-24:00				end					

Table 8

	0:00-0:30	0:30-1:00	1:00-1:30	1:30-2:00	2:00-2:30	2:30-3:00	3:00-3:30	3:30-4:00	4:00-4:30	4:30-5:00	5:00-5:30	5:30-6:00	6:00-6:30	6:30-7:00	7:00-7:30	7:30-8:00	8:00-8:30	8:30-9:00	9:00-9:30	9:30-10:00	10:00-10:30	10:30-11:00	11:00-11:30	11:30-12:00
Monday	2.37	2.34	2.34	2.30	2.33	2.31	2.36	2.33	2.34	2.36	2.40	2.50	2.68	2.73	2.81	2.90	3.01	3.00	3.09	3.09	3.04	3.06	3.01	3.01
Tuesday	2.30	2.31	2.30	2.31	2.28	2.28	2.24	2.29	2.23	2.19	2.33	2.35	2.53	2.73	2.87	2.87	2.82	2.87	3.00	3.00	2.97	2.93	2.97	2.99
Wednesday	2.06	2.08	2.07	2.00	2.05	2.05	2.05	2.05	2.10	2.16	2.11	2.31	2.54	2.61	2.84	2.82	2.81	2.95	2.91	2.93	2.94	2.94	2.92	2.93
Thursday	1.97	1.93	2.02	2.00	2.00	1.97	2.00	2.02	1.99	1.96	2.07	2.21	2.41	2.50	2.52	2.66	2.79	2.85	2.90	2.84	2.79	2.81	2.80	2.86
Friday	2.04	2.08	2.08	2.05	2.00	2.00	2.08	2.08	2.09	2.16	2.19	2.38	2.46	2.58	2.76	2.76	2.74	2.74	2.80	2.86	2.81	2.82	2.87	2.79
Saturday	2.18	2.18	2.22	2.21	2.19	2.14	2.16	2.16	2.17	2.14	2.15	2.19	2.23	2.27	2.33	2.39	2.47	2.60	2.63	2.68	2.73	2.81	2.85	2.78
Sunday	2.07	2.08	2.10	2.05	2.13	2.13	2.11	2.09	2.09	2.10	2.11	2.10	2.14	2.21	2.18	2.18	2.29	2.34	2.38	2.40	2.45	2.41	2.37	2.44

	12:00-12:30	12:30-13:00	13:00-13:30	13:30-14:00	14:00-14:30	14:30-15:00	15:00-15:30	15:30-16:00	16:00-16:30	16:30-17:00	17:00-17:30	17:30-18:00	18:00-18:30	18:30-19:00	19:00-19:30	19:30-20:00	20:00-20:30	20:30-21:00	21:00-21:30	21:30-22:00	22:00-22:30	22:30-23:00	23:00-23:30	23:30-24:00
Monday	2.89	2.87	2.88	2.90	2.83	2.90	2.82	2.81	2.92	2.84	2.77	2.70	2.69	2.65	2.59	2.60	2.51	2.54	2.54	2.51	2.50	2.51	2.51	2.54
Thuesday	2.80	2.86	2.86	2.93	2.94	2.94	2.89	2.82	2.65	2.66	2.69	2.64	2.61	2.64	2.51	2.55	2.54	2.51	2.56	2.59	2.45	2.33	2.29	2.23
Wednesday	2.77	2.79	2.75	2.86	2.86	2.89	2.87	2.83	2.76	2.75	2.72	2.67	2.63	2.64	2.56	2.56	2.39	2.31	2.24	2.21	2.17	2.10	2.05	2.01
Thursday	2.60	2.60	2.70	2.70	2.73	2.71	2.69	2.65	2.63	2.59	2.65	2.63	2.47	2.44	2.46	2.47	2.43	2.44	2.42	2.36	2.20	2.15	2.12	2.07
Friday	2.70	2.68	2.70	2.69	2.70	2.70	2.68	2.70	2.74	2.67	2.69	2.67	2.54	2.60	2.53	2.46	2.41	2.47	2.45	2.43	2.39	2.23	2.25	2.19
Saturday	2.73	2.71	2.67	2.65	2.59	2.60	2.55	2.55	2.46	2.42	2.46	2.50	2.56	2.51	2.38	2.30	2.28	2.28	2.21	2.15	2.11	2.08	2.17	2.09
Sunday	2.29	2.42	2.45	2.42	2.50	2.52	2.47	2.46	2.48	2.47	2.49	2.48	2.51	2.44	2.47	2.43	2.48	2.49	2.47	2.40	2.36	2.29	2.31	2.34

5.2 Problem: Activity classification

Two options are possible. One that makes use of pre-classified activities (classification schemes) as being frequently used, another one asking respondents to describe their activity with their own words (open response).

In the MODUL study of living conditions diary participants were asked to state their activity using their own words, being in line with the latter option mentioned above. The longest activity comes with 125 characters. Overall 2,937 different terms were used in order to describe the different activities.

The following explanations that reveal more insight into the time frame chosen will be based on the "coffee" example. This means that a specific activity was under study in order to be able to go into more detail within one activity. Further ones like "cooking" or "going for a walk" were used as well and led to the same conclusions with regards to a single activity.

Too broad time spans lead respondents to fill in several activities – typically "main" and "side" activities that were mentioned within the same time frame or cell within the diary (e.g. "getting up, personal hygiene, coffee"). However, this seldom happened and therefore half-hourly evaluations are appropriate to avoid multiple activities within one time frame.

A different problem shows up if time spans are too narrowly defined. They make the evaluation task for the respondents too burdensome. Logically, an automatic extension of an activity is a must in order to avoid too many identical repeated tasks from the study participants. In the MODUL study of living conditions – paper pencil – people made use of manually indicated brackets around several time frames.

However, summarized, the half-hourly classification represent an appropriate time span with regards to one's activities. Here it has to be mentioned that location and accompanying person come with much broader time spans justifying the avoidance of narrower time spans.

After cleaning (tokenization, lemmatization instead of stemming, stopwords removal – numerical information, punctuation…), reducing the number of different terms that appear in the activity column to the most frequent ones, binary weighting, co-occurrence detection, cosine similarity weighting, and applying agglomerative hierarchical clustering (AHC) with Ward's method, the most frequent 57 terms retain the strongest activities contained in the activity description fields.

E.g. for the coffee example, different situations and settings are mentioned in the course of the activity description as well: accompanying persons or locations, coffee as one out of several activities, being on the way to drink a coffee, coffee in the morning vs. at lunch, coffee prior an exam, coffee aside doing housework… Summarized, automatically detected topic groups reveal different written descriptions for the "same" activity. Therefore drinking coffee has different meanings in different situations.

Optional description or fields (e.g. location, accompanying person…) should be used in order to solve part of the problem, if respondents prefer to additionally differentiate between coffee drinking in different settings.

With regards to the topic detection method applied here, a posteriori conducted activity categorization by human beings might be too burdensome and it therefore has its justification. However, a big disadvantage of the conducted classification technique is that activities showing up too seldom will not be detected at all, or at least it is hard or even not possible to separate them

from other activities in a data-driven way. Offering the respondents a clear classification with an a priori categorization has its justification as well and is possibly more efficient for the respondents instead of open responses.

5.3 Problem: Rating statements

It was found that even if the time of the day, the activity, the location, as well as the accompanying person were the same, or summarized all entries are the same for the same person except from the day of the week, numeric emotional ratings were still different. This seems to be plausible between weekdays and weekend. But even on two different weekdays evaluations were different.

So the quality of time spent with a certain activity might be driven by other aspects apart from just the activity itself. It is not only the sole activity, like discussed in the literature by the 'flow'-concept, but manifold other reasons are drivers of related feelings and the resulting numeric evaluation. Just being familiar with the activity carried out does not give insight into ones well-being whilst performing an activity and just knowing about the activity does not allow for generalizations over time for the same person.

This lead to the conclusion that an additional need for a detailed description of ones feelings has to be used to explain such variation. Therefore, open statements next to the numerical rating in the very last column of the diary screenshot above allowed respondents to additionally comment on their numeric evaluation of feelings during a certain activity. Different reasons were located.

The longest emotional description for a certain activity rating was 324 characters long. Continuing the coffee example from above it becomes clear that the emotional evaluation tackles different reasons even for the same activity (e.g. "somehow tired", "coffee in the garden, nice conversation", "positive interruption of work", "feeling to be finished with some work"…). Consequently, descriptions for certain ratings are manifold. This is also the reason for different emotional ratings albeit the same activity of the same person at the same time with the same accompanying person was under consideration.

After cleaning (tokenization, lemmatization instead of stemming, stopwords removal – numerical information, punctuation…), reducing the number of different terms that appear in the activity column to the most frequent ones, binary weighting, co-occurrence detection, cosine similarity weighting, and applying agglomerative hierarchical clustering (AHC) with Ward's method, the most frequent 102 terms retain the strongest activities contained in the activity description fields.

Topic clusters representing argumentations for the same activity point into different directions. A qualitative interpretation together with the raw statements lists different reasons: e.g. conditions related with the activity ("nice weather"), one's own motivation for a certain activity ("I like it") [comment: this is closest to the common understanding of such evaluations], accompanying persons making the activity become something special ("walking around with spouse"), actual well-being of a person from two different perspectives: psychological ("stressed") and physiological ("back pain"). This also makes clear that emotional evaluations of activities have to be explained by the study participants. Answers to the question "I feel…" capture all of the above and even more different fields. The activity itself could play just a minor role.

Summarized conclusions from the MODUL study of living conditions:

- Apart from the activity itself, participants list manifold drivers for their ratings.
- These drivers originate from different perspectives related to each activity. A possible multidimensional construct might be inherent in the evaluation of a single activity.
- These drivers vary over time for the same person for the same activity at the same time with the same accompanying person.
- Motivated ratings reveal additional information on the meaning of sole numerical evaluations.
- One-time evaluations are not able to uncover such variability. Evaluated time use is justified.

6 Considerations and conclusions concerning the design of the well-being questions

In this chapter we will first outline the general decisions we made concerning methodology, than we will give some insights in the process of finding appropriate items and finally we will describe the chosen items and design.

6.1 General decisions

First of all we decided to use the beeping method instead of detailed diary recordings of well-being. This decision was driven by theoretical considerations as well as pragmatic ones. As the theoretical reasons are described in the chapter above, here we will concentrate on the pragmatic ones. First of all mobile surveys should be as small and handy as possible. Huge grids with day structures and additional columns for well-being questions would not fit easily on a screen of a mobile phone. Even if one could find a good und lucid design to ask this question the additional response burden would not be easy to bring together with busy mobile respondents, who fill out their diaries on the fly. Four to six push massages a day seemed much less burdensome to us.

The other reason was more driven by the project logic. As this project is already directed to the use of mobile phones, technical possibilities of the mobile phone-technology should be used and tested, which can`t be used and tested in paper-pencil or desktop only- designs. The possibilities to test new innovative designs in such an experimental project should be used.

6.2 Process of item selection

In a first round we collected different items for momentary well-being in general as well as for the evaluation of the activity and tested them with friends and relatives. In the following team meeting we came to the conclusion that items like "How do you feel doing this activity at the moment?" "How would you describe your momentary well-being?" or "how would you evaluate your average well-being "are too long and/or to abstract and academic.

For the following round of mini-tests, we developed three sets of three to seven items we presented to different colleagues and asked them to answer them directly. The first set contained seven questions, whereby the first six were introduced with the question: "How are you at the moment?" followed by happy, angry, sad, stressed, worried and tired and answering options on a scale between yes and no. The last question was "Do you like, what you are doing at the moment?" again with a yes-no- scale. The second set used basically the same design but was condensed to three items: only happy, stressed and the item concerning the activity were kept.

The design of the last set was influenced by the design, which was used in the pilot study at The Netherlands Institute for Social Research (2013). It only contained scales for the dimensions stressed, happy and tired.

In the following small and informal test-series we came to the conclusion, that seven items were too many to answer them immediately without getting angry or stressed. So we decided to keep with three items, which meant, that we had to decide, if to ask for the tiredness or for the quality of the activity. As there were many theoretical and practical reasons to ask for the quality of the activity separately we decided to skip the tired-item.

But our small test also brought another finding: the yes- no scale was seen as dichotomous rather than continuous. Therefore we first thought about switching to a bidirectional scale with happy-sad stressed-calm etc. However we did not adopt this concept because we finally came to the conclusion that there are no unambiguous "opposites" for happy. Being unhappy doesn`t mean automatically to be sad. Therefore we decided unidimensional scales as described below.

6.3   Conclusions

So for this study, the following structure was chosen, based on key conclusions from the literature as well as from the small pretests described above:

Due to the necessity to separate currently experienced total affect from the affect which is contributed by the currently performed activity, two different questions are provided regarding evaluated time use: one about the momentary hedonic status (*Ich bin gerade/I´m at the moment…glücklich/happy*), and another one about being pleased by the activity which is carried out at the moment (*Was ich gerade tue, gefällt mir…/I like what I do*).

In addition, there was space for one particular emotional aspect, going beyond a global positive-negative-assessment. From the different options being stressed was chosen because of its relevance, its connection to how the time is spent, and its frequency which matches with the rhythm of the assessments (*Ich in gerade/I´m at the moment…getresst/stressed*).

All three items were measured on a unidimensional 7- point scale from "very" to "not at all".

7 The App – technical information

The following chapter describes the technical structure of the App. There will be an overview of the technical features the app, how the installation worked and what software was used.

7.1 Installation

The TUS app is available for two different types of smartphones: Apple's iOS (iPhone) and Google's Android, which is the mobile operating system running on most smartphones from other vendors than Apple. Each of those platforms provides a store, which is a method for distributing new apps: App Store on iOS and Play Store on Android.

In order to install the app, the respondent searches for the German term "Zeitverwendung" and taps the search result entry that matches the icon and description of the TUS app. Then, installation follows the standard process of installing an app on the smartphone, to which users of the device are accustomed. If the app has already been installed before, the store redirects the user to the installed version.

7.1.1 Google Play Store

The following pictures show the overview page and details page of the app descriptor on Google Play Store. If the user taps on the green "Installieren" (install) button, the app will be installed.
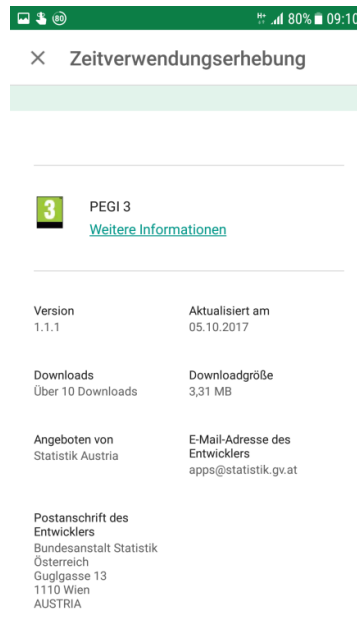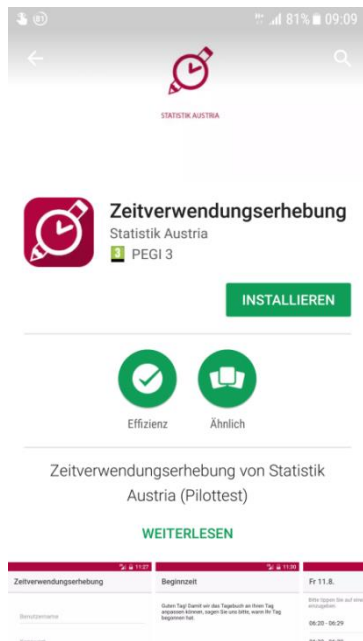
7.1.2 Apple App Store

The following pictures show the overview page and details page of the app descriptor on the App Store. In this example, the app has already been installed before, so if the user taps on the blue "Öffnen" (open) button, the device loads and displays the already installed app.
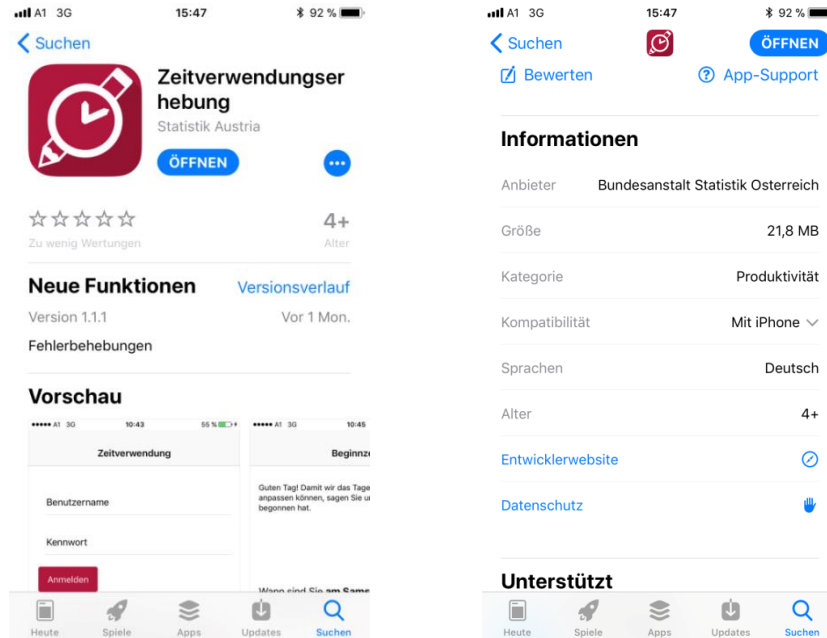
7.2 Usage

This section describes how to use the app on the mobile device, if it has already been installed following the instructions in the previous section. The following sections show the different views (screens) of the app and describe their functionality.

7.2.1 Login

Initially, the app shows the login screen. If the respondent has been provided with a username and a password by the NSI, they shall be entered here. Otherwise, the user will not be able to use the app.

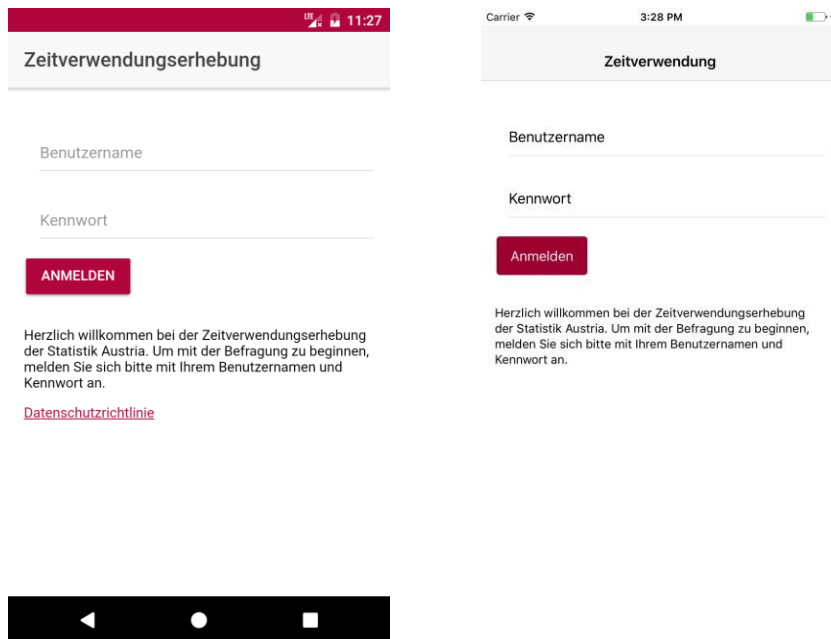Authorization data is stored locally, allowing the respondent to close and reopen the app without having to enter the user id and password again. Only after finalizing the diary, authorization data is cleared from local storage and has to be entered again.

In the current version of the app, each user id is assigned a fixed day for the survey. If the respondent is supposed to participate in the survey for more than one day, they need to be provided with additional user ids. Future versions of the app may allow answering the survey for more than one day with a single user id.

Note that the Android version of the login screen provides a link to the privacy policy, while the iOS version does not. This is due to different requirements that Google and Apple place on apps. While Google requires the link to be placed on the login screen, Apple states that a link to the privacy policy has to be provided with the app's description page in the App Store instead.

Another difference between the Android and iOS versions is the maximum allowed length of the app's title. The German Android title translates to "time use survey" while the iOS title had to be shortened to "time use" in order to fit into the constrained space.

7.2.2 Start time

After login, the respondent selects the start time of their day. This is supposed to mean the time of getting up from bed in the morning for most respondents. The survey period starts with the selected time and ends ten minutes before the selected time on the following day.

The respondent selects the start time by tapping on the grey "hh:mm" text. Then, the operating system specific time controls are displayed, which allow the respondent to enter the time value.
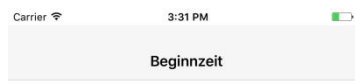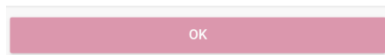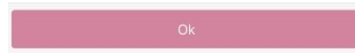
7.2.3   Initial list of time slots

After the respondent has logged in and selected the start time as described in the previous sections, the list of time slots is displayed. This is the central point of the app that allows the respondent to enter activities by tapping on the appropriate time slot. If the respondent swipes up or down, the time slots currently not visible will be shown.

Note that the 23:50 time slot is followed by the 00:00 time slot for the next calendar day – unless the respondent has selected 00:00 as start time, in which case the 23:50 time slot is the last one.

The following pictures show the list of time slots of an empty diary.
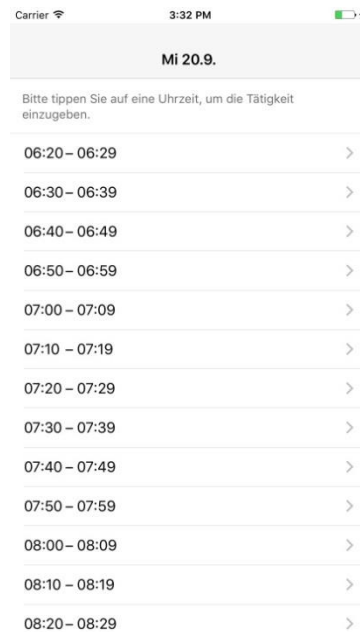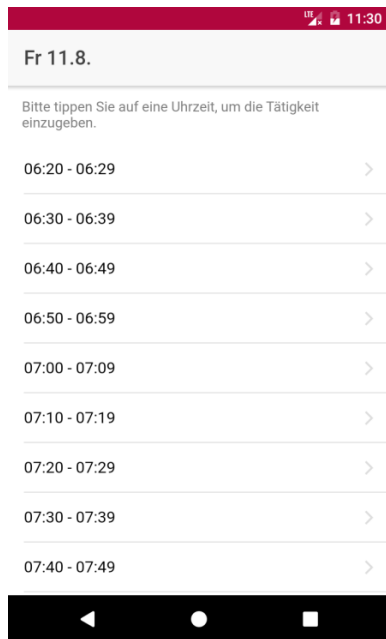
7.2.4 Details view

If the respondent taps one of the time slots, the details view is displayed and allows the user to fill in the activity. The first two items contain the primary and secondary activities which are described further below. Tapping on the third row ("Wer war dabei?") switches to a view that allows specifying the accompanying persons. Tapping the bottom row ("Wo war das?") switches to another view that allows selection of the location.
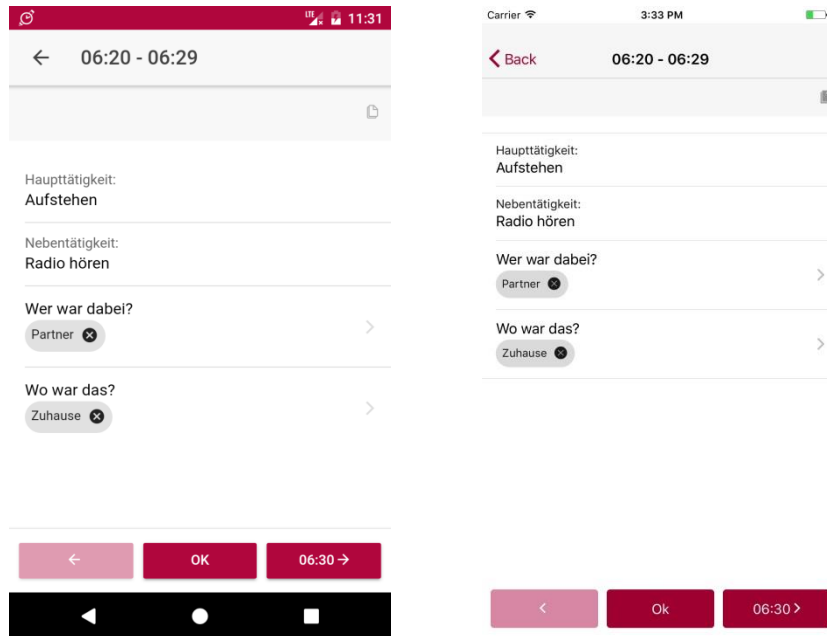
7.2.4.1 Navigating

The three red buttons on the bottom row allow the respondent to directly navigate to the previous time slot (left button), the next time slot (right button) or to the list of time slots (middle button). Swiping left or right is an alternate method for navigation to the previous or next time slot.

7.2.4.2 Copying entries

In order to facilitate entry of activities that last longer than ten minutes, a copy button is provided in the top right corner of the activity view, provided that a previous entry exists and is filled out correctly. In the pictures above, the respondent is supposed to have entered 06:20 as start time. Thus, since the 06:20 time slot is the first one, the copy button is inactive (greyed out). Otherwise, if the respondent taps the copy button, the information contained in the previous time slot is copied into the current one.

7.2.4.3 Activity selection

If the respondent taps on the labels "Haupttätigkeit" (main activity) or "Nebentätigkeit" (secondary activity), the activity view is displayed. Here, the respondent must provide a textual description of the main activity and may provide a textual description of the secondary activity. The textual description of the activity may be typed in free form or selected from a dictionary list described in the next section.

7.2.4.4 Dictionary search

In order to help the respondent to fill out the activity description, after typing the third letter the app displays a list of dictionary entries that match the yet incomplete description. The initial dictionary is stored centrally on the server and contains terms derived from responses from the previous time-use

survey. If the intended activity description is displayed, the respondent may complete the entry by tapping on that description. If the list does not contain an appropriate activity description, the respondent may ignore the list and finish typing. In addition to the dictionary list, the respondent may use the dictionary feature provided by the mobile device, if available.

In order to allow the app to handle typographical errors or alternate spellings, a fuzzy search algorithm is employed to find matching entries. The app incorporates the "fuse.js" programming library[6] for fuzzy search, which implements the Levenshtein distance algorithm[7] in order to find matching entries and sort them by similarity to the respondent's input.

The following picture shows a list of suggestions provided to the respondent after typing three letters. Note that the respondent typed "shh" which results in the app showing a list of terms starting with "sch" due to the fuzzy search algorithm described above:
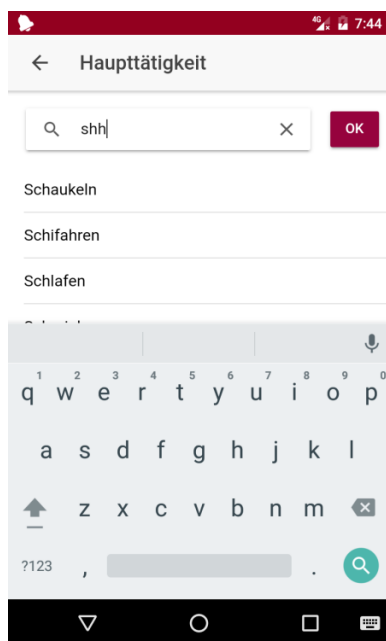
Future versions of the app may learn from the respondent's previous answers by adding them to the locally stored copy of the dictionary, while leaving the central dictionary unchanged.

7.2.4.5 Long-lasting activities

In order to prevent the respondent from having to copy or even repeatedly type long-lasting activities into several consecutive ten-minute-slots, certain dictionary entries may be designated for allowing variable duration. This feature is especially targeted at the activity of sleeping which concludes the day for most respondents. So, if the respondent selects or types sleeping as main activity, the app asks for the activity's end time and automatically fills out all of the corresponding consecutive time slots.

---

[6] http://fusejs.io/
[7] https://xlinux.nist.gov/dads/HTML/Levenshtein.html

7.2.5 Subjective Well-being questions

In addition to the activities, the app collects the respondent's answer to well-being questions. The app tries to ensure that well-being questions are filled out just in time in order to gather accurate results. So, at certain times predefined in the respondent's user profile, the app is interrupted and the well-being page is displayed:
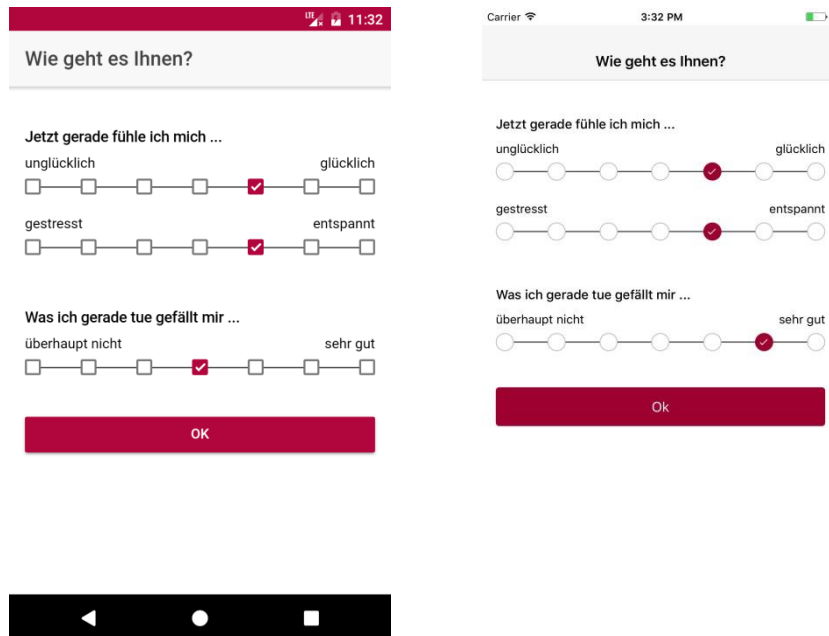
In order to provide the visual appearance of a slider while avoiding suggesting a default value, the input control of each well-being question consists of several checkboxes connected with a line. Initially, all of the checkboxes are empty. If the user taps on a checkbox, it is activated while all other checkboxes in the same row are deactivated. Only if all of the subjective well-being questions are answered, the Ok button is active.

If the app is not in the foreground (i.e. active) at the specified time, the device issues a notification which is displayed at the typical screen location, which may be the top margin, a popup window or the lock screen depending on the mobile device type. If the user taps on the notification within ten minutes, the subjective well-being question is displayed. After ten minutes, the notification is cancelled automatically. Unfortunately, the usability tests have shown that the cancelling mechanism does not work under all circumstances, so the final survey may contain well-being questions answered out of time. Those answers may be easily identified by comparing answer time to question time, though.

7.2.6 Completed list of time slots

If the respondent has scrolled to the bottom of the list (by repeatedly swiping in the down-up-direction), there is a button that allows to finish the diary. The following picture shows the completed survey with the button to finish the diary ("Tagebuch abschließen") at the bottom of the list. Note that the checkmarks in the rightmost column state that the corresponding entry has been filled out completely. On incomplete entries, a red exclamation mark will be displayed instead.

If the main activity of all time slots is filled out, the app will close the diary and send all the data to the server. Note that incomplete entries (i.e., with empty location or unspecified accompanying

persons) do not prevent the diary from being marked as finished, while totally empty entries lead the app to display an error message.
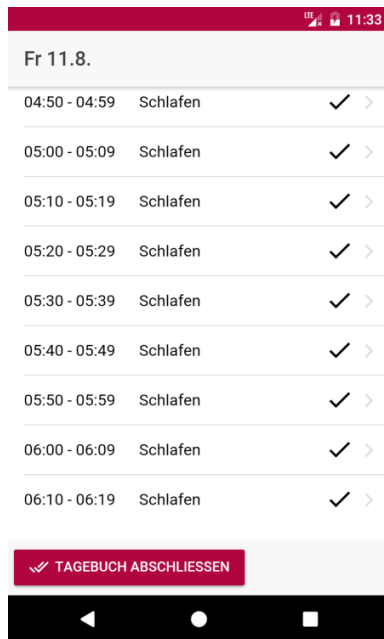
7.2.7   Final page

After the respondent has tapped the "finalize" button described in the previous section, the final page is shown which by default just confirms that the respondent intends to finalize the diary and send data to the server. After finalizing the diary, data may not be changed anymore. In the current version of the app, this is ensured by logging out the user and preventing further login with the same id. Future versions of the app may provide a read-only view of the data after completion.
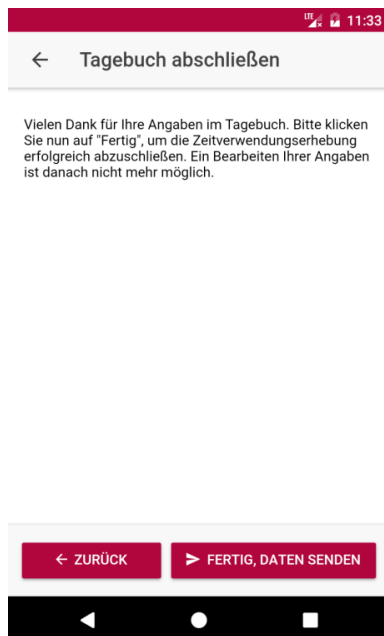
7.2.7.1   Retrospective subjective well-being question

If the respondent has not answered any well-being question before finalizing the diary, the well-being page described further above is merged into the final page in order to allow answering a single

subjective well-being question retrospectively. The user profile designates the timeslot for the retrospective question:

Note that unlike in the case of the well-being questions issued on time, answering the retrospective well-being questions is optional.

This chapter has described the installation and use of the TUS app. The next chapter describes the architecture of the app and provides rationales for the design decisions that have led to the current appearance and functionality of the app.

8 Software Architecture

Software architecture describes the fundamental structure of an application. The architecture of the TUS app has been developed based on several requirements, which are explained in the following sections: the application shall operate without Internet access (offline), shall frequently send data to the server if online, shall issue well-being notifications even if the application is currently closed and shall run on most of the mobile devices (smartphones) currently in use.

8.1 Requirements

Functional requirements describe the features of an application, that is, what the application should do. Non-functional requirements describe environmental or performance characteristics of the application, that is, how the application should accomplish the functional requirements.

8.1.1 Offline operation

A goal of the application is to allow the respondent to fill in the current activity just in time in order to gather accurate data. Since respondents are possibly situated in a location with poor Internet access or no Internet access at all, the application must allow the respondent to update their diary without being online. Consequently, the application must provide a means to store entered data locally in order to prevent loss of information not yet transmitted to the server due to power outage (e.g. empty battery), device failure or the respondent deliberately powering off the mobile device.

8.1.2 Online operation

In contrast to the offline requirement, the NSI is interested in access to the most current data, albeit incomplete or subject to change. Change patterns in current data allow the NSI to detect if the respondent has difficulties in using the app. If the respondent aborts the diary, the NSI can operate on the incomplete data. So, the requirement is that the application shall frequently send data to the server, provided that Internet access is available in sufficient quality.

8.1.3 Local notifications

An important feature of the survey is to provide the respondent with subjective well-being questions. It is desired that these questions are answered on time. It is an essential feature of a smartphone to interrupt its user in order to notify them of a current event, for example a phone call or a text message. The application shall use the device's notification functionality in order to issue the well-being question. The well-being notification shall occur even if the application is currently closed.

8.1.4 Sufficient number of devices

For every application it is desired that it can be used by as many users as possible. However, supporting older devices or multiple operating systems may exceed time and budget constraints. So, a balance must be found in order to enable the majority of smartphone users to participate in the survey while ignoring devices which are too old or have a too low market share. As of 2017, the most popular types of smartphone are Google's Android and Apple's iOS (iPhone).

8.2 Target devices

This section describes the process that has been followed to decide which mobile devices shall be supported by the app. Several websites exist which allow to assess the current market share of

mobile device platforms[8]. As of May 2017, Android has worldwide market share of 65.6 percent, iOS has 32.3 percent, and others (e.g. Windows phone, Blackberry) together have 2.1 percent.

Market share of mobile operating systems (May 2017)
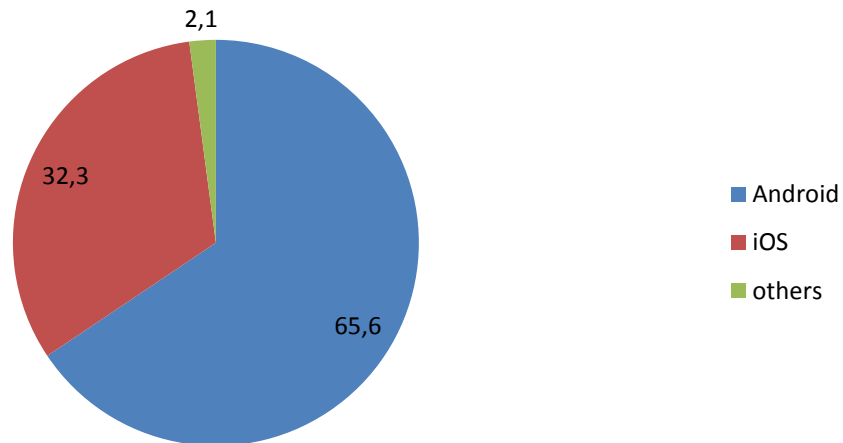
The diagram clearly shows that it is reasonable to focus on supporting Android and iOS, while disregarding the other available platforms.

8.2.1 Operating system versions

Within each operating system platform, a decision has to be made on the oldest version to support. Older operating system versions usually run on devices with less power, while providing fewer features and having more bugs than newer devices, resulting in additional effort for their support. Thus, the considerations written in the previous sections apply also to older versions. Additionally it is to be expected that an already low market share of an older device type will almost always further decrease in the future.

---

[8] https://david-smith.org/iosversionstats/
https://developer.android.com/about/dashboards/index.html
https://www.netmarketshare.com/operating-system-market-share.aspx

The following diagram shows the market share of Android versions.

Android versions (May 2017)
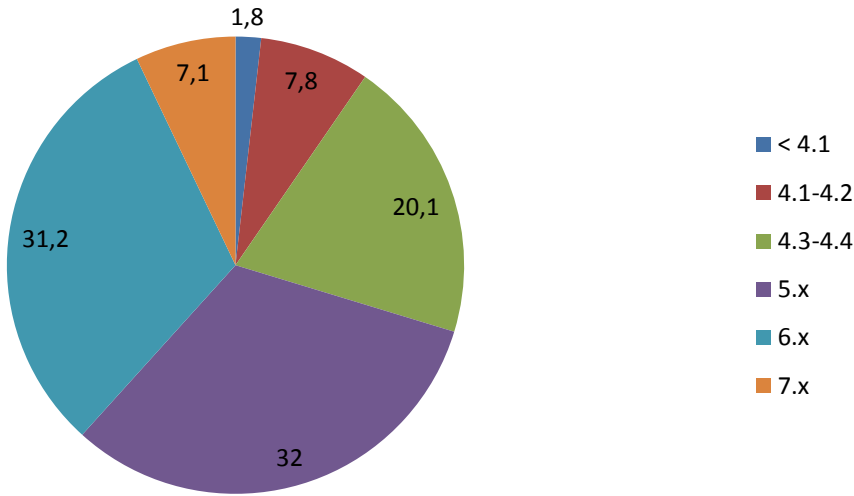
The diagram would suggest that it is reasonable to support versions 4.1 and higher. The final app has been restricted to supporting versions 4.3 and higher, though, which corresponds to devices produced or updated in 2013 or later. The reasons for this restriction will be described in the section on hybrid apps below.

The following diagram shows the market share of iOS versions.

iOS versions (May 2017)
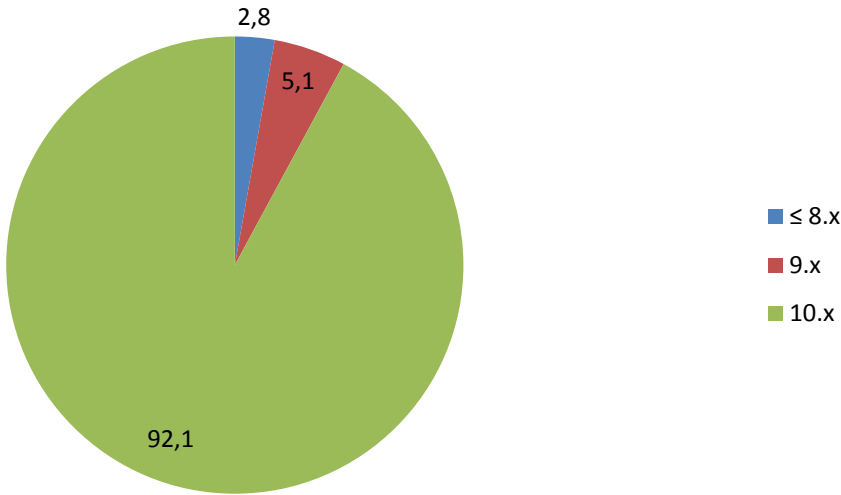
The diagram suggests that it is reasonable to support iOS version 9 and higher.

8.2.2   Market share of supported devices

Based on the market shares described in the previous sections and known technical limitations of specific older operating system versions, the decision has been made to support Android versions 4.3 and higher, and iOS versions 9 and higher. This results in a market share (in May 2017) of

$$(1 - 0.078 - 0.018) * 0.656 + (1 - 0.028) * 0.323 = 0.907$$

Thus, based on that decision, the mobile app supports about 90.7 percent of the smartphones in use worldwide.

8.3   Development platform

Currently, several approaches to develop a mobile app are in use. With every computing device, there is the possibility to develop an app targeted specifically for that device, called a native app. On the other hand, there exist tools and frameworks that allow development of an app suited to run on multiple device types. The following sections describe the main approaches to develop a mobile app.

8.3.1   Native development

Native apps, since being developed specifically for a certain operating system are able to exploit all of the features provided by that device with the best performance and consequently are superior to all of the other options. The main disadvantage is that additional effort has to be made in order to support other operating systems. The technical differences between Android and iOS do not allow any code developed for one system to be reused on the other system. In many organizations, development skills are only present for one of those platforms.

Since the decision has been made for the TUS app to support iOS and Android, native development is not feasible.

8.3.2   Web apps

Recently, there has been the trend in software technology to provide new applications as web apps. Modern web browsers meanwhile support the presentation of feature-rich user interfaces with almost no difference to traditional applications. Adaptive web design allows interfaces that change their layout in order to adapt to different screen sizes. The main drawback is that web apps are still limited in their access to basic system functionality. Plus, web apps usually are not distributed via the usual channels (stores), resulting in additional effort of directing respondents to the web site.

Specifically for the TUS app, web apps do support local notifications, but cease issuing notifications if being closed by the user. So, the requirement stated above to present well-being questions even if the app is closed cannot be met with a web app.

8.3.3   Hybrid apps

In order to allow the development of apps that run on multiple operating system platforms, while allowing access to most of the hardware functions of the underlying device, several frameworks have been developed that combine the advantages of native apps and web apps. Apps developed with those frameworks are usually called hybrid apps. Due to the layered architecture consisting of a container and an app running in that container, performance is somewhat slower than with native apps.

The following table summarizes the main advantages and disadvantages of the approaches for mobile app development.

	Advantages	Disadvantages
Native	Best performance and features	Only for a single platform
Web	Runs on all platforms	Limited features
Hybrid	Many platforms, many features	Slower than native app

For the TUS app, the decision has been made to implement a hybrid app using the Cordova framework.[9] Cordova provides a container for each target platform that contains the almost platform-independent web app providing the functionality to the user. Plugins allow access to system functionality like storage and local notifications. The user interface has been developed using the Ionic framework[10] which allows adaptation of the app to the look and feel of native apps thus providing a uniform overall user experience. As an example, the following screenshots show the details view on Android and iOS:
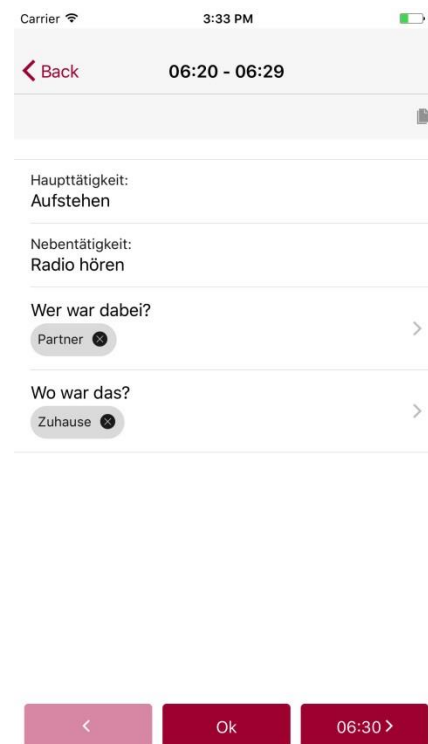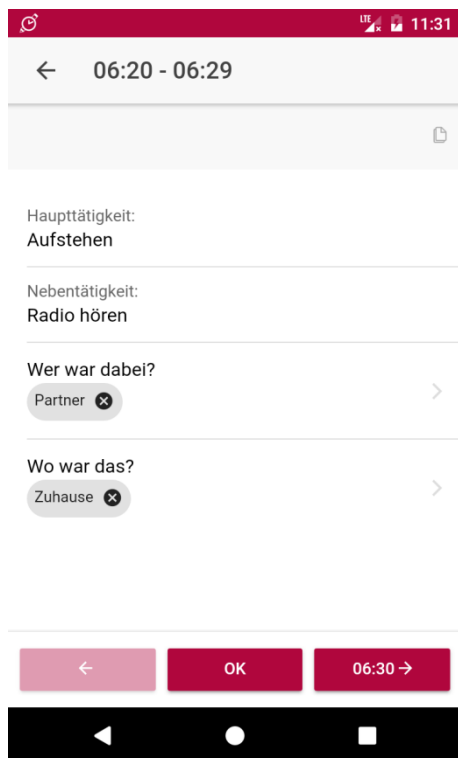
Hybrid apps run in the web browser environment provided by the operating system. Since the browser engine that is available on Android versions 4.2 and below does not fully support the requirements of Cordova, apps that need to support older versions have to ensure installation of a new browser engine. Since this would have required additional effort and resulted in a larger app package, while targeting only about 5 percent of the mobile phone users as of May 2017, the decision has been made to support Android 4.3 and higher versions only.

---

[9] http://cordova.apache.org/
[10] http://ionicframework.com/

8.4 Structure

While the app on the mobile device is the user visible part of the overall application, several other parts are needed to provide complete functionality. The server which is located at the NSI needs to validate authentication data, send the user profile and dictionary to the app and receive the respondent's entries from the app. In order to permanently store data, a relational database management system is used.
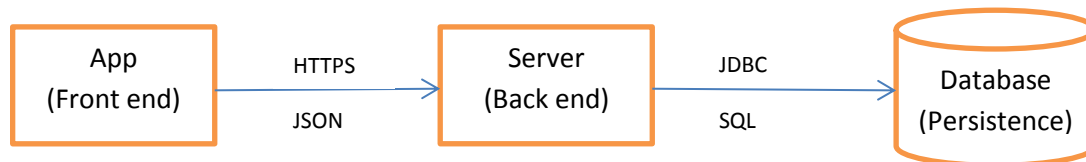
The diagram shows that the app sends data encoded in the JSON[11] format to the server over an HTTPS connection, which is an encrypted connection type to ensure privacy. After receiving data, the server which is written in the Java programming language sends SQL statements that store the data into the database over a JDBC connection. This structure comprises a three-tier-architecture, which is nowadays state of the art for designing distributed computer systems.

8.5 Distribution

Distribution comprises packaging and publishing of the app to Google's Play Store and Apple's App Store. Before publishing an app for the first time, the software developer has to register as an app developer with both companies. The following Section describes the registration and distribution process for Play Store and App Store.

Both stores require uploading an application icon, a certain number of screen shots and textual information.

8.5.1 Google Play Store

Registration to the Google Play Developer Console requires a valid Google account (i.e. an email address ending with "@gmail.com"). This may as well be a generic (non-personal) email address, which is then redirected to an internal mailing list of the NSI consisting of the app developer team. In order to upgrade that Google account to a developer account, a one-time fee of currently US$ 25 has to be paid by credit card and the developer distribution agreement needs to be accepted. Although this registration method is very straightforward, it requires the developer team to share the password of the Google account. After completion of the preceding steps, apps may be published.

Publishing an app consists of uploading the package to the Google Play Console web site, filling out additional information, uploading icons and screenshots and starting the publishing process which usually lasts from a few minutes to a few hours.

8.5.2 Apple App Store

In order to allow registration to the App Store, Apple has defined several roles in a developer team[12]. A special role is the team agent which must be a single person that has been granted authority in a formal business letter by an authorized signatory (i.e. a top level executive) to bind the NSI to a legal

---

[11] https://www.json.org/
[12] https://developer.apple.com/support/roles/

agreement with Apple. The following registration process needs to be completed using the team agent's personal Apple Id account (which may be a business account separate from their possibly existent private Apple Id account). In addition, a D-U-N-S number[13] of the NSI may be required. Upon enrolment and acceptance of several agreements, a fee of currently US$ 99 per year is due which has to be paid by credit card. Using the credit card information, the written authorization letter, the D-U-N-S number and the personal Apple Id, the team agent may complete the registration process. The registration process needs a few business days on Apple's behalf and includes verification of the data by calling back the team agent by phone.

After completion of the registration process, the team agent may invite the other members of the developer team (all of them being required to obtain a personal Apple Id first) to join the developer account, either as an admin or as a simple member. Additionally, in order to allow one or more persons of the developer team to publish new versions of the app, an appropriate iTunes Connect role has to be assigned to the respective Apple Id.

Publishing an app to the App Store is supported only by using Apple's XCode development environment running on the macOS operating system. So, the development team needs to have access to an Apple computer system (e.g. MacBook or iMac) in order to publish an app. New versions of the app undergo a review process by Apple which usually lasts a few business days per submission and is supported by detailed guidelines. Organizations publishing their first app should plan for several iterations of the review process due to rejections.

---

[13] http://www.dnb.com/duns-number.html

9 Usability Test

In this chapter the usability test of the app together with the cognitive testing of the well-being questions is analysed in detail. Therefore the used methods and the design are described and of course the results of this test.

9.1 Introduction

In the context of the "Time Use Study" (TUS) conducted by Statistics Austria, a time tracking system in the form of a mobile app was developed in the first half of the year 2017 in order to investigate everyday activities among the Austrian population. The aim of this mobile app is to assess the different activities in which respondents engage during the day in an as detailed way as possible. To that end, users are required to fill in 10 minutes-timeslots covering all 24 hours of the day by indicating their main and side activities in each slot as well as the location and other persons present there. Considering that most users have their mobile phone with them all day long, designing this time tracking system in the form of a mobile app should make a continuous protocolling of one's activities during the day easy.

In the third quarter of the year, a usability test was conducted in order to further improve the mobile app. The test aimed at answering two questions: (1) Which aspects of the current mobile app constitute sources of error in the sense that they *directly* increase the risk of measurement errors? And (2) which aspects of the process of data collection may impair the quality of the data *indirectly*, e.g. by causing strain in participants?

9.2 Methods

Design: In order to answer these questions, data was collected using two different approaches: On the one hand, nine participants were asked to use the mobile app in a real-life-setting during a period of two consecutive days and to afterwards report their experiences in a semi-structured telephone interview. Hence, while this approach can only assess reported behaviour, its benefit consists in the circumstance that participants can fill in their activities immediately as they go along. On the other hand, ten participants were invited to test the mobile app in a laboratory setting, i.e. to use it at Statistics Austria. In this scenario, participants were asked to fill in the timeslots of the current day based on their memory. Both the screens of their mobile phones and the participants themselves were videotaped, such that their actual behaviour when using the mobile app could be observed. Here, too, participants reported their experiences with the mobile app in a semi-structured interview.

Sampling: The overall sample, thus, included 19 participants. The latter were chosen using a purposeful sampling-strategy in order to obtain heterogeneous perspectives on the mobile app. The predefined sampling criteria included sex, age, educational background, occupational status, general IT- experience, as well as experience and usage habits regarding smartphones and mobile apps. Out of these 19 participants, eleven were female and eight male. In terms of age, participants ranged from 14 to 76 years and the educational background included pupils in secondary school as well as high school and university graduates. The usage habits regarding mobile apps ranged from hardly any use to an hourly use.

Data analysis: Both the video-taped observations and the interviews were analysed using qualitative content analysis (c.f. Schreier, 2012)[14]. The categories used in the coding frame were derived both deductively, i.e. based on the research questions, and inductively, i.e. based on the empirical data: The two main categories dictated by the purpose of the study were (1) sources of error, i.e. aspects of the mobile app design that directly increase the risk of measurement errors and (2) aspects that impaired the quality of the data indirectly by causing strain and decreasing participants' motivation. The first main category ("sources of error") was further divided into seven subcategories that covered the different parts of the mobile app ("download and log-in", "start", "overview", "detail view", "finalization", "well-being", and "technical problems"). The specific sources of error within each of these subcategories were then derived inductively, i.e. based on the empirical data. Similarly, the subcategories of the second main category (aspects of the process of date collection that impair the data quality indirectly) were first specified deductively, i.e. following the requirements of the study, resulting in the following subcategories: "participants' approaches to the mobile app", "aspects causing strain decrease the quality of the data", and "recommendations for the process of data collection". The specific aspects within each category were, then, again derived from the data, i.e. inductively. The following two chapters will elaborate on the specific outcomes of the data analysis in more detail.

9.3 Main sources of error

Sources of error are defined in this chapter as aspects of the mobile app that directly increase the possibility of measurement errors. Other aspects of the mobile app that may indirectly affect data quality by decreasing the motivation of the users or causing strain for the users are discussed in this chapter only as much as necessary. More detailed aspects of such indirect problems are discussed in the next chapter, "Respondents' behaviour and recommendations for the process of data collection".

An error is treated as a major error if at least two respondents encountered the same problem.

9.3.1 Technical issues

There are two major technical issues with the mobile app.

Storing data: controllability

4 out of 19 respondents were not sure how to save their input, or if their input is only saved when the diary is completed. This led to an uncertainty whether they could interrupt the process of filling in diary entries or if they had to fill in the whole day at once. The uncertainty about saving their data was more common with respondents who did not have previous experience with mobile apps.

Performance

In 13 out of 19 cases the respondents experienced the performance of the mobile app as so slow, that they considered it a problem. Issues were either the poor input responsiveness or that the display seemed to freeze. For the users, it seemed as if the mobile app did not accept any further input. By retrying, they often changed data that had been entered previously. (So the focus of the problem changed from answering the question to closing the entry record. Thereby respondents sometimes did not mind anymore which input they actually gave.) Some respondents also stated that if they were at home (as opposed to being in the laboratory), they would have aborted the test.

---

[14] Schreier, M. (2012). „*Qualitative content analysis in practice*", London: Sage.

Recommendations for improvement
It is unclear if the problem occurs due to performance issues or bad responsiveness of a functionality within the mobile app. However, both can be improved with a combination of various measures. First, by letting the user do something productive while data is stored asynchronously, in case of poor internet connectivity or connectivity loss data should be locally cached. Second, if some waiting is inevitable, a progress indicator can significantly reduce frustration. For the mobile app, an animated icon seems to be sufficient. Longer delays would be better covered with a progress bar, or even an approximation of the time that an operation is going to take before starting it. However, the functionalities of the mobile app are of low technical complexity and therefore do not justify longer delays.

9.3.2   Download and login

There are three major problems concerning the download and login process.

Information letter

In 2 out of 9 telephone interviews the respondents considered a registered letter for activating the mobile app as too cumbersome.

Name of the time tracking system

There was a discrepancy of the name of the time tracking system in the information letter and the name of the app in the app store. In 4 out of 19 cases this made the respondents doubt whether they were downloading the right app.

Code

Two teenagers had problems with the code, although they are very experienced users of mobile apps, as the last symbol of the code given in the letter was not interpreted as part of the code, but as a punctuation mark. Logins often failed, and they needed the help of the observer. So, if they had tried to install the mobile app at home, it is not sure, if they would have managed to activate it (even if they had persisted in trying longer).

A solution could be to graphically highlight the code even more. However, another problem is that the download and activation process of the Statistics Austria mobile app does not follow a very common pattern, as the users get the information on how to download it in a separate letter. It seems that this led to more problems with younger users, irrespective of their mobile app using habits.

Another problem was also that the first letter of the code word was not capitalized. This is not very suitable for the task, as many mobile phones automatically capitalise the first letter of an entry.

9.3.3   Start

When the login process is finished, the user is asked to fill in the beginning of his/her day. The only major problem is related to the meaning of the question:

When did your day start? When did you get up?

Three problems occurred in respect to this question. (1) The meaning of the question was not clear. Answers could refer to the first experience of the day (when did your day start?) or to the time of getting up, which are not necessarily the same. In one case, the respondent had gone to bed late, so she was not sure if her day started before she went to bed. In other cases, people reported activities that precede getting out of bed (such as chatting with a friend or snooze). Different users have different interpretations, if these activities should be filled in or not.

Recommendations for improvement
It should be decided to skip one of the questions. As it might be interesting to get data on activities before getting out of bed, the second question could be skipped. On the other hand, asking about activities before getting up might be too private. Thus, it might be not very appropriate as an opening question.

9.3.4 Overview

After filling in at what time the user's day started, a table is displayed which starts with the time the user had given as the beginning of the day and ends 24 hours later. The timeline is divided into 10-minute slots, such that there is a total of 144 time slots.

Task is not clear: each and every timeslot?

Some respondents think that only major activities should be entered, so if nothing extraordinary happened, they leave time slots empty. Others think they should just fill in the starting time of an activity and leave the following time slots empty, if the same activity continues. This uncertainty occurred in 6 cases.

Recommendations for improvement
It could be helpful to highlight the instruction text, as not all users read it. Besides, the instruction text ("please tap a timeslot to fill in your activities") needs to be rephrased. It is not precise enough, as it does not tell the user whether all time slots have to be filled in.

Task is not clear: which level of detail and what kind of activities?

Half of the respondents did not know in how much detail they should protocol their day. The morning activities (breakfast, taking a shower, etc.) as well as the way to school or work are mostly protocolled in a very detailed way. Regarding work or school, it is not clear if broad categories (such as "work" and "school" respectively) are suitable, or if the users should describe what they are actually doing (having a meeting, writing e-mails, different classes at school, etc.). However, most respondents decided to use broad categories such as "school" or "work". This way of describing their day is due to various problems that will be described later on. However, it is evident that the instruction text needs to clarify in how much detail the users should report their daily activities. Many respondents wanted to have examples of which activities should be protocolled. To make the instruction text more helpful, it would also be important to graphically highlight the text in a more prominent way. It might also be a problem that the time tracking is called "diary" (at the bottom of the overview page there is the field "close the diary"). The word "diary" suggests that only activities that are worth being remembered should be protocolled.

Furthermore, examples could be included in the input fields as placeholder texts.

Lacking possibilities to make inputs in a timely manner

For most users it was not possible (or they were not willing) to take the mobile phone and protocol their day every ten minutes. Being at school, at work or driving a car, one cannot fill in one's activities at such a high frequency. Most respondents made their inputs every couple of hours, when they had time. To do so, they had to remember their activities in order to reconstruct their day, which obviously is not possible on such a detailed scale. Some respondents also filled in activities in advance.

Many respondents wished to be not bound to the ten-minute slot per activity entry, but to be able to enter the duration of each entry (as it is the case for the sleeping period). However, this would change the research question, as it can be assumed that if this was possible, relatively long periods would be entered and thus no detailed information on the users' daily life would be generated. However, other measures may help:

Recommendations for improvement

- In many cases, activities cannot be reduced to smaller units but actually take longer than 10 minutes (a meeting, driving the car, a school lesson etc.) Many respondents did not find the copy function; others did find it, but still found it very time-consuming, as the user has to copy and paste for every single time slot. So, it may help if – depending on the activity filled in – a time wheel could pop up. For example if the user has a meeting, the mobile app could ask when the meeting ended. This is already the case for "sleeping" and could also be used for other activities, if the grade of detail is good enough.
- The mobile app could be more interactive. A user could start one or more activity streams and stop them by tapping a button. This might lead to a more real-time activity monitoring. Scheduled or rule-based pop-up messages (e.g. after a certain time of inactivity) could ask the user what they are doing at the moment. However, it is obvious that pop-up messages every ten minutes would be annoying and counterproductive.
- It might also help if not every time slot is visible on the overview. For example the day could be divided into four phases (morning, midday, afternoon, evening), such that the first impression of the overview would not be so overwhelming.
- Other graphical highlights could be created, so that the focus is not placed on the (high number of) timeslots.

Difficulty to fit activities into time slots

In contrast with the problems described above, in some cases, users had difficulties to make their daily activities fit into the ten-minute time slots, as the activities either took more or less time or did not correspond to the pre-set slots (e.g. one respondent went to work from 7:45 to 8:15, so the ten-minute intervals – 7:40, 7:50, 8:00, etc. did not fit). One respondent solved the problem by using the side activity input field to enter an activity, as neither activity took 10 minutes.

9.3.5   Detail View
When the user taps on one timeslot, the detail view is shown. On this page, four entries can be made: main activity, side activity, people present and the location of the activity.

In the following, problems are described referring to these fields.

9.3.5.1 Main and side activities

Task is not clear

It is not clear to the users, if side activities are a required input field.

Recommendations for improvement
The most common implementation to indicate required fields is to place the asterisk to the left of a field label.

Definition is not clear

In most of the cases (12 out of 19) the definition of main and side activities was not clear to the users. This is independent of the age or the level of education. This issue leads to a broad range of interpretations what main and side activities might be. Some respondents think that the main activity is one's occupation; others understand the main activity as the more important one and the side activity as the less important one. And yet others think that the main activity is the activity with the longer duration and the side activity the one with the shorter duration. However, respondents were not sure if their understanding is correct, which often led them to leave the side activity field blank.

Recommendations for improvement
It might be considered to rename the "main" and "side activity" fields as "first" and "second activity". Thus, the user does not have to decide on the importance of each activity and does not need to classify as main and side activities. This is often not possible or inconsistent.

Autocomplete suggestions (list of activities)

When respondents enter the first letter(s) of an activity in the main or side activity field, a list of autocomplete suggestions pops up. The users can then select one of the suggestions to speed up input. Regarding these autocomplete suggestions, several problems arise.

Task not clear: are the suggestions compulsory?

7 out of 19 respondents did not know if they had to choose one of the autocomplete suggestions or if they could also key in a different, individual text. However, in many cases the respondents found out by trial and error that an individual text is possible. (Nevertheless, they had to find out themselves, and those users who did not want to use the autocomplete suggestions and thought they had to use them were annoyed.)

Recommendations for improvement
When the autocomplete suggestions pop up, a remark is necessary that the usage of the suggested terms is optional. Otherwise, it is not seen as help, but leads to the frustration of the user.

Difficult search

Problems occurred in 13 out of 19 cases when users were browsing the list of suggestions. It is very difficult to find the matching item on the list of autocomplete suggestions, because a systematic search is not possible. Suggestions pop up that have some letters in common with the user entry or

the beginning of the entry. That does not necessarily mean that the quality of the search results is good. Some suggestions are synonyms, and some suggestions are not clearly differentiated (e.g. "reading the daily newspaper" and "reading the newspaper"). Often none of the suggestions is appropriate at all. (Missing activities in the autocomplete suggestions are for example: taking the underground (in German: U-Bahn fahren), taking a break ("Pause (machen), "sich fertig machen", which means morning routine and getting dressed, meeting, eating, writing a test, going home, going to school, going to the toilet, going to work, to blow-dry one's hair, etc.)

As a result, using the autocomplete suggestions is often much more complicated for the users than writing individual texts.

Recommendations for improvement

Eliminate the autocomplete function in the mobile app. Perhaps use speech input instead.

The process of generating activities

Six factors can be identified that have a negative impact on how users input activities. To various degrees, these lead to false or divergent entries and thus to a reduced data quality:

- the respondent's own perception of their activities: for example, it seems as if working or attending school is experienced as one homogenous unit, whereas brushing teeth, getting dressed, having breakfast and so on are seen as distinct activities. So in most cases there are many activities protocolled for the morning and the grade of detail is decreasing steadily during the day.
- Social desirability: Which activities are protocolled, is also caused by wanting to give socially desirable answers. For example activities such as chit-chat, idling, staring into space are not protocolled. Thus, protocols of a consistently productive/industrious day are created.
- Trivial or intimate activities are often left out (like going to the bathroom). This is also due to being not sure if these activities should be protocolled or left out.
- Difficulty to remember activities: As entries cannot be filled in promptly, the users often cannot remember activities in every detail but summarize activities and group them into a larger unit. (For example, users found it difficult to remember their work day in every detail but it is possible to simply protocol "working".)
- Decreasing motivation: in most cases, motivation decreased during the course of the survey. This is another cause of generalizing activities in entries that are made later during the day.
- Inputs are also determined by constraints of the mobile app. If the respondent thinks that they have to use the autocomplete suggestions, it is obvious that the inputs will be determined by them (see "autocomplete suggestions"). If they are not totally accurate, a shift of meaning is provoked. For example, one test person wanted to enter lay down ("liegen"), but could not find it and entered "brushing the teeth" instead of it. In another case, the respondent wanted to enter "going to school" but could not find an appropriate suggestion. So he entered "going". If users consider the selection of the autocomplete suggestions as mandatory, they tend to type synonyms until a suggestion fits (somehow). If this process takes too long, frustrated users just select any suggestion to fill in the time slot.

9.3.5.2 Locations and person present

In the detail view, the user is not only asked about his or her activities, but also where they were and who was with the user during the described activity. In contrast to the main and side activity, there is no possibility to write a free text or choose from a long list of autocomplete suggestions. The user can only choose from response categories given by the mobile app.

Referring to the "person present", the following problems occurred:

- Missing categories: response categories are focused on the household. Classmates, colleagues, and friends are missing. Almost all respondents had problems with this question.
- Unknown categories: two young respondents did not know the meaning of the phrase "household members".
- Members of which household? This lack of clarity occurred only once. Although it is not a major problem in a strict interpretation, we assume it is an important case: A student had problems to define if her sister was a household member or not. She was a household member of her home in Lower Austria. However, in Vienna, her secondary residence and the location where she met her sister, she was not a household member. So the respondent did not know whether to fill in whether a household member was present or not.
- Definition of being alone is not clear: 7 out of 19 respondents had problems to define when they were alone. Problems referred to the question if a person is only alone when no other person is in the flat or house, or if being alone only means nobody else is present in the same room. And whether it is possible to be alone when other people are in the room, but the user him/herself is conducting an activity on their own (such as taking a shower)? One respondent also assumed "being alone" was an inappropriate answer if many other but unknown persons were with him (for example when taking public transport).

Regarding information on locations, the following problems occurred:

- There are no possibilities to indicate locations in case of movement. Thus, in the case of mobile activities (driving a car, going to work, etc.) all respondents had difficulties in naming a location, and the responses they gave (were obliged to give) were often questionable. For example in the case of going to work, one respondent chose "work" as the location. Almost all respondents had problems with this question.
- Missing categories: In contrast to the response categories of people present, there were more possibilities to describe locations. However, nearly all users wished to be more precise on the location where they had been. Missing categories are for example: school/university, means of transport. The option to choose "other location" (in German „anderer Ort") was regarded as unsatisfactory. In one case, the user thought that she would get more response categories after tapping "other location".

9.3.5.3 Error warning

When the main activity, the location or the people present are not filled in, an error warning (a red exclamation mark) pops up next to the time slot where the entry is missing. The visibility of the icon is good. However, the exclamation mark does not indicate *what* is wrong. When the test users noticed the error warning, they opened the detail view again, but no hint was given what they did wrong, so the warning was not always regarded as helpful, but caused even more troubles. One respondent recommended to add an explanation which fields have to be filled in.

Recommendations for improvement

Errors and warnings have to be communicated to users gracefully and clearly. An effective error/warning message provides following information:

- clearly communicate what is happening
- describe how a user can resolve it
- preserve as much user-entered input as possible

The mobile app can be improved by immediate real-time inline validation which informs users about the correctness of data they provide. If the mobile app is performing in-line form validation, and the field with the error/warning is clearly marked, users can correct the errors they make faster without switching from the overview back to the detail view and searching for possible reasons for the error/warning.

9.3.5.4 Copy function

In the detail view, there is an icon and a note that says that the user can copy the previous entry. A copy function is essential for this mobile app, as many activities take longer than ten minutes. Besides, by copying entries, the high number of timeslots becomes manageable.

In regard to the copy function, four main issues occurred:

- Visibility: 4 out of 19 respondents did not see the copy icon, at least at the beginning. The problem is that the copy function is not immediately visible. The icon only pops up after the previous time slot is completely filled in (except for side activities). So it is only possible to copy a time slot when the fields of the main activity, location and people are filled in. However, quite often not all entries are filled in, as it is not clear which fields are mandatory.

Recommendations for improvement

The copy function should be visible all the time. However, a hint should be given that the user has to fill all entries before the copy function is enabled.

- Intuitive handling: The handling of the copy function was not intuitive for two respondents, as they could not copy and paste activities, but just paste one activity into the following timeslot.
- Tedious process: It is only possible to copy one time slot at a time. In the case of broader categories, this process gets tedious. For example, if a respondent wants to enter a long period of work (e.g. eight hours), they have to repeat the copy process 56 times. Many respondents wished to be able to copy an entry many times at once.

Recommendations for improvement

The idea behind the ten-minute time slots and the lack of possibilities to enter longer durations at once is to prevent the user from using broad categories. However, we assume that other possibilities (like explaining the task, giving examples, making the mobile app more interactive and attractive to use, etc.) are more suitable to increase the grade of detail. Constraints such as a limitation of the possibilities of the copy function, hardly lead to an increase of detail, but rather enormously increase the risk that users abort the test.

- In two cases, the copy function did not work or provoked a crash of the mobile app.

9.3.6 Finalization

When "sleeping" (in German „schlafen") is filled in in the main activity field, a time wheel appears and the user can indicate the duration of sleeping. So in this case, the user does not have to fill in individual ten-minute time slots. This is extremely helpful.

Two problems occur:

- **Only the word "sleeping" provides access to the time wheel function**, and the time wheel does not appear if users type in "going to bed", "going to sleep", "sleeping until", etc. This only happened four times, as the users in the test lab of Statistics Austria did not protocol their whole day (but only the time up to the point they arrived at the lab for the test), but this is likely to be one of the main problems that cause users to abort the test without finishing it.

Recommendations for improvement

There could be a "sleep" button in the detail view, to tell the mobile app that the user goes to bed. The user can push the button, which opens the time wheel function. In this way, user irritation about the fact that a certain word in a free text input prompts unexpected user interfaces could be eliminated.

- Unclear end of the diary entry: 5 out of 19 people did not know if the protocol ends at 24:00 o' clock or in the next morning. It is not obvious to the users that the protocol period is not a calendar day, but 24 hours, from the first entry in the morning until the following morning. The confusion is further intensified by the time wheel (which indicates the duration of sleeping) which ends at 24:00 o' clock.
- Handling: As the time wheel ends with 24:00, more interactions are needed (two entries have to be made: first sleeping until 24:00 o'clock, and then sleeping from 0:00 until the next morning).

Recommendations for improvement

- In the instruction text at the top of the overview, it could be explained that Statistics Austria is interested in the course of 24 hours from waking up in the morning, until the following morning.
- The time wheel should continue until the next morning.

9.3.7 Well-being

The respondent is also asked once how he/she felt at a certain time. Three questions are about the grade of happiness at that moment, the level of stress and whether they liked what they were doing at that moment. This question appears either at the end of the diary entries (if all time slots are filled in at once) or in the course of the day. In the latter case, the respondent is asked how they are feeling at that moment. If all time slots are filled in at once, the question will be how the user was feeling at a certain time (which is chosen randomly). For each Respondent the well-being notifications popped up four times during the day at different 10 minute timeslots. The fourth notification always popped up during night (24h) or in the very morning (6h). These night/morning timeslots were never answered via notification. From the remaining 3 notification timeslots, 56% were never answered, 33% were answered via notification and 11% were answered at the end of diary completion. Put in another perspective: Out of all completed diary days (12 days), we collected for 8 days (67%) answers to at least one well-being timeslot via notification and for 4 days (33%) via

end of diary completion. Although these numbers cannot be counted as representative (as this was a qualitative study), the data hints to the following assumptions: The majority of all possible timeslots cannot be reached via notification. On the other hand, if it is the goal to get an answer to at least one timeslot per day, this goal can be reached via a combination of notification and a question at the end of diary completion (asked if a person has not reacted to any notification). In this way we will probably get in at least half the days live answers to at least one well-being timeslots. Although this seems promising, we learned that simply sending notification is not enough to get live answers. Improvements should be undertaken in getting the respondent to allow the app sending notifications and checking their phone more often during the protocol-day.

Other Problems were:

- It only happened once, however it is seen as important, that only two out of the three questions where displayed on the screen. The third question was invisible due to poor screen resolution. A more responsive design would fix this issue.
- The time periods to which the questions referred were seen as inappropriate by 4 out of 19 users: the questions referred to the time when the respondent was sleeping.
- Often the question about the user's happiness was not answered referring to the time period in question, but in a general way, i.e. referring to the whole day.

9.4 Respondents' behaviour and recommendations for the process of data collection

From a more general perspective on respondents' behaviour, different approaches to the mobile app could be identified and a systematic link between strain in participants and the quality of the data become obvious. Both these aspects as well as recommendation for the process of data collection derived from these insights are delineated in the present chapter.

9.4.1 Participants' approaches to the mobile app: expectations, motivation, and styles of dealing with it

Participants' expectations toward the mobile app, as well as their general motivation to accomplish the task it requires and their way of dealing with it vary considerably. Some participants – mostly highly educated adult ones – are aware of the scientific purpose of the app and perceive their participation in the study as a legal duty. Others, on the contrary, mostly teenagers, expect a mobile app to be entertaining and are disappointed with the design of the mobile app and the possibilities it offers for gaming – or rather: which it does not offer. The design being perceived as rather plain, some conclude that the authors' of the mobile app do not care much, whether participants make an effort in responding. These participants are quickly annoyed with the task and not very motivated to accomplish it, since they feel that it has nothing to offer to them, and only do accomplish it in the laboratory setting at Statistics Austria.

Regarding the ways in which participants handle the mobile app, different styles can be observed: While some users do all they can to adapt to the requirements of the mobile app as much as possible (e.g. scroll for appropriate key words and rephrase several times until they find one that fits their activities), others part from the activities they can remember and reconstruct their day based on these reference points (i.e. they focus more on their own doing than on the requirements of the mobile app). Also, while some users click from one time slot in the detail view directly to the next, others go back to the overview after completing a slot and before going to the next slot. Lastly, the

degree of detailedness with which respondents provide information on their activities also is a matter of style, as respondents decide themselves, which activities are relevant enough to be mentioned. This particularly regards side activities in the sense that some are able to make use of this category while (most) others simply skip it – be it due to low motivation or unclarity about the definition of side activities as compared to main activities.

9.4.2 Aspects causing strain decrease participants' motivation and the quality of the data

As indicated in the previous chapter, a general link between respondents' strain and the quality of the information they provide can be observed: Aspects of the mobile app that cause uncertainty, confusion, and even irritation have a straining effect on participants. Technical problems, for instance, such as the freezing of the mobile app, entries not working, etc. or unclarity about the purpose of the mobile app, which information to provide, etc. cause strain and decrease participants' motivation to provide extensive and accurate information. With the distinction between main and side activities not being clear to number of participants, for instance, many did not provide information on their side activities at all. Others, for instance, who repeatedly experienced technical problems with the mobile app, got frustrated over time and copied previous entries in order to get the task done, at least formally. Hence, straining aspects decrease the completeness and the correctness, i.e. the quality, of the data provided.

However, the core problem of the app that causes strain in participants and, thus, impairs the quality of the data substantially, is a structural problem, namely the overall arrangement of 24-hours day into 10-minutes-timeslots. Independent of individuals' level of motivation, the top-down logic of this structure forces participants to deliver an adaptive performance in the sense that they have to adapt their day to the structure provided by the mobile app. These 10-minutes-timeslots are perceived as too long for some activities such as brushing one's teeth, while for many others they are judged as way too short. This is particularly the case for activities that last several hours and which are perceived as uniform, i.e. where participants do not further distinguish between subparts included in these activities, such as working. In such cases, the copying function is not perceived as an appropriate means of dealing with the strain involved in the task and, consequently, the majority of participants wished for the possibility to determine timeslots themselves. Hence, the structure of the 10-minutes-timeslots is highly demanding on a cognitive level.

In addition, it is also highly demanding on a motivational level: with this structure implying a high level of detail, i.e. 144 timeslots, the overview listing all the slots is perceived as so demotivating that less information is provided and in a less accurate way, i.e. previous entries are copied and activities do not necessarily match the timeslots at which these activities took place.

In practical terms it is worth mentioning, that in order for the 10-minutes slots to be filled in accurately, it is necessary to complete them immediately and continuously during the day. As this is not convenient/possible at all times, at least some slots are necessarily filled with a certain time delay, making recollection necessary. In light of the huge amount of slots, however, this may cause mental overload, which, in turn, impairs the quality of the data.

It must, thus, be noted that the task is overcharging *independent of participants' motivation*. Consequently, all participants use one strategy or another in order to deal with this insoluble task: while some confine themselves to rather broad and abstract information such as "working", others copy previous entries in order to accomplish the task at least formally, and others again leave out

some information (e.g. side activities) or entire timeslots, and still others abandon the mobile app entirely – and they even do so in a laboratory setting. Using such strategies that impair the quality of the data seems all the more likely in a) a real-life setting, i.e. were no researcher observes one's doing, and b) in the context of a study, in which users are legally obliged to participate (as is usually the case for Statistics Austria).

9.4.3   Recommendations for the process of data collection

As demonstrated in the previous section, the core problem lies in the rigid top-down structure of the 144 10-minutes-timeslots in order to capture a 24-hours day, and it can be doubted that the quality of the data generated by using this structure is better than it would be with an entirely different arrangement of the mobile app.

Consequently, it seems advisable to consider a different structure of the mobile app. In order to make it less deterring and demotivating, broader timeslots of one hour, for instance, might be a solution. This, however, does not solve the problem that predefined timeslots often do not match with the length of the actual activities.

In light of that, one may have to consider the possibility for participants to choose the timeslots freely, i.e. to allow them to make entries as a function of their activities and to adapt the time-span accordingly, rather than the other way round. While this implies a greater effort for the analysis of the data (as longer time-spans have to be broken down into 10-minutes-timeslots by the researchers), this approach offers the benefit of an increased quality of the data: First, it avoids mental overload, which impairs the quality of the data directly. And second, it avoids decreases in motivation and makes it, thus, needless to employ strategies of dealing with an insoluble task that, in the end, impair the quality of the data indirectly.

Should it be decided, however, that the rigid top-down structure of the 10-minutes-timeslots is to be maintained, it is strongly advisable to avoid any other kind of strain resulting from the app (i.e. irritations resulting from technical problems, unclarity about demands, definitions, and intentions of the mobile app, etc.). For detailed recommendations in that regard, please refer to the previous chapter on "Main sources of error".

In either case, there is an ambiguity inherent to the mobile app that seems hard to remedy when looking at the data: As most users have their smartphones with them all day long, designing a time tracking system in the form of a mobile app should make a continuous protocolling of one's activities during the day easy. It appears, however, that smartphones are mostly perceived as a means of entertainment rather than as a means of work. Consequently, many users, especially young ones, are not inclined to spend effortful time on a mobile app. Several means to remedy this issue have been suggested by the participants, such as providing some scope for design to users (e.g. changing colours or including emojis), giving colourful graphics of how one spent the day, or providing the possibility to gain points and win. A gamification of the app, however, may run contrary to the (scientific) purpose of the study and a reward-based means of data collection may yield quite different results. Hence, it appears that this fundamental ambiguity inherent to the mobile app cannot be completely dissolved.

9.5   Conclusion of the usability test

The usability test identified a number of rather minor "sources of error" that can be remedied quite easily and a core issue regarding the structure of the mobile app, which causes participants to

employ strategies that impair the quality of the data and which call for more fundamental changes in the mobile app.

Such a minor problem, for instance, is that in some cases, the task is not clear. For example: Does the user have to fill in each and every time slot? Which items are mandatory? Are the autocomplete suggestions for main and side activities compulsory? Another problem was a lack of clarity regarding the definitions of main and side activities, of household members and of "being alone". Problems also occurred in respect to the handling: The handling of the copy function and the time wheel were not intuitive for some users. Another problem was, that only the word "sleeping" (and not "going to bed", etc.) as a main activity provides access to the time wheel function. However, these problems are quite easily to solve. Instruction texts can be rephrased and highlighted, the handling can be made more intuitive (e.g. the time wheel should continue until the next morning), there might be a "sleep" button in the detail view, in order to tell the mobile app that the user goes to bed, etc. (For a detailed description of the problems and recommendations for their solutions, please refer to chapter 2).

However, the most important aspect causing strain and, hence, harming the data quality is the rigid top-down structure of 144 10-minutes-timeslots in order to capture a 24-hours day. This arrangement requires an adaptive performance on the part of the respondents, causes mental overload, and decreases motivation. Hence, the task was perceived as insoluble, and, as a result, all participants used one strategy or another to get by (copying, leaving out information, using only broad terms, etc.). All of these strategies, however, decreased the completeness and the correctness, i.e. the quality, of the data provided. Therefore, different approaches in structuring might be considered. One such means might be to use bigger timeslots (e.g. of one hour) where several activities can be indicated. Or, rather, participants might indicate their activities first and, based on that, indicate the according time-span freely. The mobile app might also be more interactive. A user could start one or more activity streams, for instance, and stop them by pressing a button, which might lead to a more real time activity monitoring. Alternatively, scheduled pop up messages or rule-based pop up messages (e.g. after a certain time of inactivity) can ask what the user is doing at the moment.

Hence, the solutions to many sources of error suggested in chapter 9.2 can only remedy rather minor problems, but this can be achieved relatively easily. Solutions to the core aspect impairing the data quality, on the other hand, require more fundamental changes in the structure of the mobile app.

10 General conclusions and recommendations

In general the project was a valuable way to learn and to get new experiences in measuring time use with innovative tools. The following recommendations come from literature review, the best practise review and the Module Study but especially from our experiences with the usability test. How usable is our prototype? What should be changed? Are look up tables something that we would recommend? What well-being item should be used? What problems might occur if we use what we constructed?

The following features worked well:

The app was working satisfyingly. All respondents received an invitation letter with the information for the download. All of them were familiar with downloading the app. The simple and consistent design of the app with its Icon led the participants through the app. All of the respondents were not cautious to fill out the diary and it was learning by doing. But they also indicated that they are already used to apps and have higher demands on the standard regarding the usability of an app.

The arrow from the main diary page led to the detail page of the diary and for all respondents it worked intuitively. Also if the warning (a triangle at the related timeslot) was shown it was regarded as a "nice reminder" - not very specific but not annoying. People were willing to fill out the three well-being questions if they are short and not open. The subjective well-being questions were also seen as a reflection about the personal life and this was a nice side effect. The last page to send the data was also working adequately.

Technical issues and login

For the next time use survey the technical impact has to be considered. The loading time of an app and different devices with different operating versions can result in performance problems - especially for older mobile phones. It is essential to have a performative code of the app. Also security reasons at the login-procedure can be in contrast with user friendliness. Very complicated but secure passwords are not convenient for the users. On the other hand the security has to be ensured. For the next wave the login-procedure should be administrated with letters including QR-Code or an e-mail with a link or even a personal contact at the beginning of the survey.

Well-being

Regarding the items the unidimensional 7-point scales can be recommended- not only for theoretical reasons mentioned in chapter 5 but also because respondents did not seem to have problems with them. Also, the questions themselves were usually easy to understand. The only exception is the happy-item, where in some cases the answer did not correspond to the time period, as intended and stated explicitly, but seemed to refer to a more stable mood over the day. That might be due to the fact, that *glücklich* (german translation for happy we used, which could also be translated as lucky) has other connotations than happy and could be perceived as a stable mood rather than a concrete emotion. Here further research is needed on the different translations of "happy" into different languages. One solution might be to use *gut gelaunt* (in a good mood) instead of *glücklich.* In any

case, new items and also designs should be tested not only in an usability test but also in a (separate) cognitive test.

Concerning the time periods chosen for the well-being-beeps, some respondents mentioned that it made no sense to them to answer well-being questions for a period where they were asleep. One solution for this problem might be to include a "sleep-button" (and maybe also a "wake up button"). Then the diary could automatically be filled with "sleep" for the time in between, push-messages could be turned off and respondents would not be asked for well-being information about this time period at the end of the diary.

In general, respondents did not report to be overburdened or annoyed by the push-messages but reacted to them not very often (see 9.3.7), either because they didn't notice them or the function was turned off or people decided that they did not want to react to them. One of the reasons could be that people did not expect them and /or were not prepared to answer them. Another reason could be that some respondents did not take their phones with them all the time or kept them in a bag where they could not hear them. One conclusion could be that one should inform respondents through a little tutorial at the beginning or in the process of the data collection about what to expect and ask them to keep their mobile phone ready for use. The better way however would be to build the app in a way, that it is used during the day constantly anyway (for example asking the respondents to press a button at the beginning and at the end of an activity). In some cases it could even be useful to switch to another device, which is worn at the body and always at hand (e.g. a smart watch). Finally, one could motivate responses to push messages by gamifications (e.g comparisons of individual respondents' data to the average respondents, analyses at the end of the period about mood curves during the day).

Our solution to present non- answered push-messages at the end of the period again for the concerning time period worked quite well and is strongly recommended. Respondents did not report recall problems. However, to foster well-being measurement on time, one should take measures to encourage respondents to answer as many push messages as possible immediately.

Time slots

A challenge the respondents faced were the 144 time slots. The aim to get more precise activities is hard to achieve. The test shows that people tend to drop out - especially busy people gain higher drop outs rates with this structure of the diary. Even if people are highly committed and answer all time slots they tend to be more imprecise and this leads to washy results. The respondents' abstract or "forget" activities or leave activities out that seemed not important to them. Sometimes they reported the change of the activities but were less detailed in between this time slots.

Activity list

The activity list has the requirement that suggestions should be „complete", in a language the respondents use and in a consistent level of abstraction ("brushing the teeth," "dress up", "going to work"). The language of a person displays often the personal life-style and can differ between e.g. the younger and the older generations and in-between these groups. The dictionary with this activity-list should also be able to learn and the searching function should be developed furthermore. This complete activity list with these additional features would need a lot of time to work it out and we do not see that this is possible in a sustainable and efficient way. We advise to use the build-in-

dictionary of the mobile device in connection with the autocomplete function of the mobile phone and not like in the test a given activity list. In terms of data preparation we suggest not spending too much time in developing the activity list and invest a lot more time in automatic text recognition.

Types of filling out

The usability test showed there are different types of time diary users. The one who fills out the activities afterwards, others report their activities 'live' during the day and some even in advance. The app has to cover all needs. For the implementation of apps it is necessary to get the results of the "Task force on Innovative Tools and Sources for TUS" and other awarded Grants if they have experience with online-diaries. We advise also to get in touch and use ideas from open source projects i.e. "Tabi project" or ESSNET Serv.

Calendar-like-version of the diary

For the respondents it is very difficult and for some even impossible to differentiate between main and side activity. We suggest giving them the possibility to protocol every activity they do parallel without necessarily having to judge which one is more important. With the calendar-like-function this challenge should be solved. We advise to use a feature like we had for sleeping at this prototype-app for all activities. It is possible to press a button that starts and ends the activities (familiar to the calendar function of android or apple calendars). This would push the people to protocol live. In order to be close to the respondents` reality, questions could be adopted upon the done activities (e.g. asking pupils if they breaks during a long day in school). Future research should cover this issue more in depth.

Persons who protocol at the evening could have a help with the GPS-option. A list or a map of the places the person has been could be displayed and aid the recalling of activities. Some apps offer the possibility to track the time someone has used other on apps i.e. social media apps, mail-programs, or time for surfing on the internet. Of course this has to be an option someone can choose. More features mean also a higher power consumption and leads to a performance issue of the app and it is a privacy issue.

Concerning the results of the usability test – persons' showing a lot of abstract activities after a few hours could be remembered to mention their other or parallel activities such as taking a break, having a meeting. These notifications could be sent like the well-being-messages.

Instant gratification of the app

For the next time use survey we suggest that the app is more grounded in the reality of the user's behaviour. The app should adapt to the user's personal lifestyle and not force an artificial reality upon them. Especially the ten minute slots, although they are meant to help the respondents to be as precise as possible, cannot reach this aim. Not every activity, (in fact only a few) lasts exactly for 10, 20 or 30 minutes.

The app should give the participants immediate feedback. This is essential and it serves the need of the respondent. The detailed reconstruction of the day is burdensome and the users of an app are used to get an instant gratification. This gratification can be a graphic demonstration of the personal time for the different activities. But also the keyword "gamification" of apps should be taken into account without being childish or suspect. The app should make fun and give you some extra value immediately.

Concluding it was a big step forward to develop this time use survey application. In future it will be necessary to go more into the deep and concentrate more on the usability of the app. And before the fieldwork starts a new test phase should be definitely considered. Enough testing time is really necessary. Unforeseen bugs and challenges that can occur in the field phase need extra time for getting new solutions.

11 List of References

Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual review of psychology*, 54(1), 579-616.

Cummins, R. A., & Nistico, H. (2002). Maintaining life satisfaction: The role of positive cognitive bias. *Journal of Happiness Studies*, 3(1), 37–69.

Diener, E. (2000). Subjective well-being: The science of happiness and a proposal for a national index. *American Psychologist*, 55(1), 34–43.

EUROSTAT (Hrsg.) (2009): Harmonised European time use surveys; 2008 guidelines. Luxembourg: Office for Official Publications of the European Communities.

EUROSTAT (Hrsg.) (2004): How Europeans spend their time; Everyday life of women and men. Luxembourg: Office forOfficial Publications of the European Communities.

Franz, A. (Ed.). (1996). *Familienarbeit und Frauen-BIP*. Vienna: Statistics Austria. [Family work and women's GDP]

Gershuny, J. (2011). *Time-use surveys and the measurement of national well-being*. Centre for Time Use Research, University of Oxford, Swansea, UK, Office for National Statistics. http://www.timeuse.org/sites/ctur/files/public/ctur_report/4486/timeusesurveysandwellbein_tcm77-232153.pdf

Kahneman, D. (1999). Objective happiness. In: D. Kahneman, E. Diener, & N. Schwarz (Eds.) (1999), *Well-Being: The foundations of hedonic psychology* (pp. 3-25). New York: Russell 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. (2004). Toward national well-being accounts. *The American Economic Review*, 94(2), 429-434. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.320.928&rep=rep1&type=pdf

Keuning, S., J., & Timmermann, G.J. (1995). An information system forEconomic, Environmental and Social Statistics: Integrating Environmental Data intoSESAME". In: *U.S. Bureau of Economic Analysis (Eds.), Second Meeting of theLondon Group on Natural resources and environmental accounting, Conference papers,* Washington D.C., S. 378-369.

Kimball, M., & Willis, R. (2006). *Utility and happiness*. University of Michigan, 1-67. http://www.ifs.org.uk/conferences/kimball_0611_1.pdf

Krueger, AB 2009 (ed) *Measuring the Subjective Well-Being of Nations: National Accounts of Time-use and Well-Being*, Chicago: University of Chicago Press/NBER.

Minnen, J., I. Glorieux, T.P. van Tienoven, S. Daniels, D. Weenas, J. Deyaert, S. Van den Bogaert, S. Rymenants (2014): Modular Online Time Use Survey (MOTUS) – Translating an existing method in the 21st century. Electronic International Journal of Time Use Research, 11, 1: 73-93.

Reis, H. T. (2012). Why researchers should think "real-world": A conceptual rationale. In: Mehl, M.R., & Conner, T.S. (Eds.), *Handbook of research methods for studying daily life* (pp. 3-21). New York: The Guilford Press.

Schaffer, A., & Stahmer, C. (2006). Women's GDP-a time-based input-output analysis. *Schweizerische Zeitschrift für Volkswirtschaft und Statistik*, *142*(4), 367.

Schwartz, J. E., & Stone, A. A. (1998). Strategies for analyzing ecological momentary assessment data. *Health Psychology*, 17(1), 6.

Schwarz, N. (2012). Why researchers should think „real-time". A cognitive rationale. In: Mehl, M.R., & Conner, T.S. (Eds.), *Handbook of research methods for studying daily life* (pp. 22-42). New York: The Guilford 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.) (1999), *Well-Being: The foundations of hedonic psychology* (pp. 61-84). New York: Russell Sage Foundation.

Sonck N. and Fernee H. (2013): Using smartphones in survey research: a multifunctional tool. Implementation of a time use app; a feasibility study. The Netherlands Institute for Social Research (

Staudinger, U. M. (2000). Viele Gründe sprechen dagegen, und trotzdem geht es vielen Menschen gut: Das Paradox des subjektiven Wohlbefindens. [In spite of many causes, many people are well off: The paradox of subjective well-being].*Psychologische Rundschau*, 51(4), 185–197.

Stiglitz, J. E., Sen, A., & Fitoussi, J. P. (2010). *Report by the commission on the measurement of economic performance and social progress.* Paris: Commission on the Measurement of Economic Performance and Social Progress.

United Nations (2005). *Guide to producing statistics on time use: measuring paid an unpaid work*. United Nations: New York. https://unstats.un.org/unsd/publication/SeriesF/SeriesF_93e.pdf

United Nations Secretariat (2012). *Report of the United Nations Expert Group Meeting on the Revision of the United Nations Trial International Classification of Activities for Time Use Statistics (ICATUS)*, 11-13 June, 2012. United Nations: New York.

Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American psychologist*, 35(2), 151-175.

12 Annex

12.1 Summaries of the ten videos

12.1.1 Video no. u00031

32 years old male in employment, IT experience 7/10; smartphone experience 7/10; app- utilization 9/10; phone: Samsung Galaxy S6

1) Summarizing account of the test procedure

The respondent is a quite experienced app and smartphone user, yet generally overwhelmed and even annoyed with the task: he feels that the overview is demotivating and that filling in every 10-minutes-slot is very tedious. The copy-function helps somewhat in that regard, but he discovers it only after five minutes and also thinks that it is not sufficiently helpful for activities that last several hours. Also, the copying takes too long, which makes it annoying for him to copy many slots. He would prefer it, if he could define the time-slots himself (as is the case for the activity "sleeping"). Also, the respondent dislikes the scrolling for the time he got up and would prefer entering the time on the keyboard.

He perceives the **suggestions from the drop-down menu** regarding activities as partly heavily inappropriate, but amusing. This bears the risk of distraction, as being entertained by absurd suggestions is more attractive than filling in the app. He suggests **indicating the time slot in the detail view** as well as a means for orientation. At the same time, numerous key-words that he is looking for, are missing in the suggestions, namely "pinkeln, U-Bahn fahren, arbeiten, Pause (machen), Feierabend, unterwegs sein, zu Fuß gehen, sich (morgens) fertig machen, Hemd bügeln, Schicht (-übergabe)".

The options regarding "**other persons present**" repeatedly seem inappropriate to him and, thus, cause confusion (e.g. "alone" when in the underground with many other, yet unfamiliar, persons). Likewise, the fact that public areas (such as means of transport) are not available as "**locations**" also seems inappropriate and causes confusion.

A **FAQ-section or a button explaining the functioning of the app** would be helpful to him, since it takes him a while to realize that he can enter activities in his own words (and doesn't have to chose from the suggestions) as well as to discover the copy-functions. These two possibilities, however, made filling-in the app more tolerable and, thus, improved his motivation.

In order to **make the app more attractive** to him, he would find it useful if one could save certain activities, i.e. personal routines that occur regularly.

2) Sources of error

In the respondent's view, the **aviso letter** should contain a **QR-code** rather than login-data in order to avoid problems with typos.

The **download** took too long considering the data volume of only 6MB.

Log-in and start: Minuscule as the first letter of the log-in and/or password are problematic, as the software often starts with capital letters automatically.

The **overview** is perceived as deterrent and demotivating and he feels that filling in every one of the 144 10-minutes-slot is very tedious. Also, the 10-minutes-slots are too short to be recalled precisely hours later, but can only be completed accurately when doing it continuously throughout the day. This, however, is not feasible at work. As a consequence, the activities he enters only roughly reflect what he has done during the day (e.g. "working" rather than precise tasks involved at work).

Detail view: He falsely assumed that clicking on the search-button would yield suggestions hence the search-button initially implied a risk of misunderstanding. It took him 15 minutes to find out, that he can write activities in his own words. Consequently, searching for predefined key words in these first 15 minutes decreased his motivation. It is not clear to him how **main and side activities** are defined.

Suggestions from the drop-down menu are perceived as very inappropriate, but entertaining. This, however, constitutes a risk of distraction when scrolling for appropriate suggestions for too long, as the key words that he is looking for are missing in the drop-down menu and inappropriate key words (e.g. Ofen ausräumen; Reisig sammeln; Heidelbeeren pflücken) are more entertaining to look at than filling out the app. His suggestion for improvement in that case would be to indicate the time-slot one is working on in the detail view as a means of orientation.

Entering locations and side-activities by pressing the **OK-button does not work repeatedly**. As a consequence, he tries entering other, less precise key words as a side-activity (e.g. doing laundry instead of ironing shirts) in order to make progress with the app. Yet, the OK-button still does not work. Yet, the app did save the information after all, as becomes clear after rebooting the app. Likewise, the copying function does not work repeatedly and finally only does so when clicking several times on the OK-button.

The fact, that the time up to which one sleeps can only be entered in 10-minutes-steps based on the current time (e.g. sleeping until 6:19/29/39/… a.m., when the app is completed at 22:19 p.m.) causes **irritation** and is **misinterpreted** as relating to the time that he usually gets up.

He perceives the questions concerning his **well-being** as inappropriate for the chosen time-slot (going by underground).

18 years old female, Matura (high school graduation), smartphone experience: 7/10; app-utilization: 8/10, daily usage, phone: HUAWEI P8 Lite ALE-121

1) Summarizing account of the test procedure

The test was made in the lab. Therefore, all activities were filled in at once. After the tests, the test person summarizes that the mobile app worked well. However, during the test, she often seems annoyed. She thinks that using the mobile app is too time-consuming.

Reasons and causes: (1) she considers response time to be poor, the mobile app seems to freeze and it leads to several attempts until dairy entries seem to be stored; (2) the number of time slots; (3) She did not use the copy function. The test person thinks that she would not have finished the survey if she would had done it at home. She believes the aim of the mobile app is that the users get information on their day, so if she does not benefit from the mobile app, it is evident that she is not motivated to continue the survey. (-> The aim of the app should be more transparent.)

Download: The download was fast. However, the icon of the mobile app was not recognizable at the first instant. It was also confusing that the name of the mobile app was different in the app store to the information letter.

Overview: At first, the task was not clear. The high number of time slots confused her and it was not clear if she has to fill in every single time slot. Corresponding to that, it was not clear how detailed she should protocol her activities. (-> The task should be more specified.) In solving the tasks, she does not follow the process suggested by the app (filling in one activity after the other in chronological order), but she follows her recollection (first trying to remember what she did during that day, and then considering when each of these activities happened), scrolling up and down to fill in the slots. In the end, some time slots are left free, as she only marks the beginning of an activity. As with many other respondents, the amount of information given is decreasing. For the morning activities, every time slot is filled in. Later in the day, only extraordinary events are mentioned (like testing the mobile app at the Statistics Austria). When the error warning came up (red exclamaiton mark), she did not correct anything but deleted the input. (-> It would be helpful not only to mark that a mistake happened, but which mistake happened.)

Detail View: The test person does not fill in side activities but sometimes she fills in two activities as main activities. She tries to choose autocomplete suggestions: e.g.: she searches for „fertig machen", which is a colloquial phrase for the morning routine and getting dressed. As she cannot find it, it is changed into "putting on make-up". However, if no suggestion fitted, she wrote her own text. She was not sure if this was the correct way. The suggestions – or not to know if one has to take them – irritated her.

Finalization: Similar to many other respondents, she thinks that the dairy entries end with 24:00.

2) Sources of error
In general:

(1) The process of generating activities; performance: The test person considers response time to be poor, the mobile app seems to freeze which leads to user retries until the dairy entries seem to be stored: She tries to solve the problem by pressing the back button and answering the questions once more. Therefore one of the questions was answered

differently the second time it came up. It seems that the test person was annoyed and wanted to fill in anything just to be able to continue with the process.

Download

(2) There was a discrepancy of the *name* of the time tracking system in the information letter and the name of the mobile app in the app store.

Overview

(3) Task is not clear: each and every timeslot? At the beginning, the task was not clear. The user focused on the high number of timeslots and not on the instruction text. (-> The manual should be more highlighted.), It is not clear if every time slot should be filled in.

(4) Task is not clear: which level of detail and what kind of activities? It is not clear how detailed activities should be described.

Detail view:

(5) Are the autocomplete suggestions compulsory? The test person is not sure if she is allowed to write her own text or if she has to choose from the auto suggestions.

(6) Copy function: the function was not clear

(7) Location and people present: being at home: She is not sure if being at home means in the village or in the own house.

Finalization:

(8) Unclear end of the diary entry: She thinks the day ends with 24:00 and not on the next morning

Well Being:

(9) The question about well being referred to a time slot when she was sleeping and thus could not be answered.

23 years old female in employment, Matura (high school graduation), IT experience 6/10, mobile app-utilization 3/10, mobile app experience 5/10, phone: Samsung Galaxy A3 2016

1) Summarizing account of the test procedure

The test was made in the lab. Therefore, all activities were filled in at once. The test person thought that the mobile app testing procedure was quite laborious overall. That was caused by technical problems (low responsiveness) but also by the high number of time slots. In the feedback interview, she said that the high number of time slots made her give less detailed input. She would like to define the time intervals herself. However, it can also be seen that the test person is very diligent in her responses, e.g. she corrected typos. The test person says she is used to track activities at work and it seems as if she would use the system for orientation.

She would like to be given an overview after finishing the protocol showing her day, telling her how long she was alone, how long in the car, how long at home and so on.

Download: As it was the case for many others, the download confused the test person, that there were two different names for the same mobile app in the app store and in the information letter.

Overview: As with many other respondents, the grade of detail is relatively high for the morning activities (putting on make-up, getting dressed, waking up the partner, ...), but when it comes to describing her workday, she uses the more general category "working at the office" to describe her day. However, when she changes locations, makes breaks or has a meeting, this is filled in as well. Durations of activities are shown by leaving the following time slots empty, so she only marks the beginning of an activity (the first time slot). She does not use the copy function. Only after getting the error message indicating that she has to fill in the time slots that were left empty, she starts to copy activities. However, sometimes the test person also puts in new entries (e.g. tidy up the workplace, cooking.)

Detail view: In contrast to many other respondents, she uses the autocomplete suggestions only for the first time slot. She first thinks that she has to use them but tries to write a text of her own which was also accepted by the mobile app. From the second entry she does not choose from the suggestions any more.

She does not know if she has to fill in side activities. Describing her spare time, she does not enter any side activities, as they would not be important enough, as she claims. (eg. driving the car and calling her mother). In contrast to that, for her workday, side activities are described in two cases.

Naming locations and people present was sometimes difficult. This was the case for filling in a location in the case of the time slot "driving". (There are no possibilities to indicate locations in case of movement) She also had problems to answer which people were present during her workday, as the person categories focus on household members.

Finalization: It is of major importance, that in her case, the time wheel to fill in a longer sleeping period did not come up. A time wheel only pops up if "schlafen" ("sleeping") is filled in. However, the test person filled in "going to bed ("schlafen gehen"). So it is possible that respondents have to fill in every single time slot for the sleeping period. Less important but also remarkable, as it occurs quite

often, was that she had to enter "sleeping" twice (after the observer told her to fill in "sleeping"), as the time wheel ends at 24:00.

2) Sources of error

In general:

    (1) Performance: low responsiveness

Download

    (1) Two different names for the mobile app in the information letter and the app store

Overview

    (2) task unclear: At the beginning it was not clear if she should or can fill in her activities in the overview or if she should skip that page.

Detail View

    (3) Obligatory side activities? She is not sure if she has to fill in side activities.
    (4) Places: In the case of mobile activities: Should she fill in the place of departure or of arrival?
    (5) People: Missing categories: The category "colleagues" was missing.

Finalization:

    (6) The time wheel to fill in a longer period of sleep was not shown. A time wheel only pops up if "schlafen" ("sleeping") is filled in. However, the respondent wrote "going to bed ("schlafen gehen"). So it is possible that respondents have to fill in every single time slot for the sleeping period.
    (7) Less important but also remarkable, as it occurs quite often, was that she had to enter "sleeping" twice (after the observer told her to fill in "sleeping", as the time wheel ended at 24:00).

12.1.4 Video no. u00049

32 years old female university graduate in employment, IT experience 4/10; smartphone experience 2/10; app- utilization: rarely; phone: Sony Xperia (bought in December 2016)

1) Summarizing account of the test procedure

The **aviso letter** was perceived as appropriate regarding the amount of information it provides and both the **download** and the **log-in** went smoothly.

Generally the respondent showed great willingness to comply with the task and very patiently filled in the slots. She first did not realize that it is actually possible to enter activities in her own words, and even after she has found out that this is possible, she still stuck to suggestions from the drop-down menu. Although finding a more or less appropriate suggestion from the drop-down menu often required several attempts of rephrasing and searching for quite long in the drop-down menu, she indicated in the interview that it seemed less of an effort to her to search for an suggestion, than to type in an activity in her own words. Also she assumed that picking suggestions would make the data analysis easier. Hence, it was not clear to her, that phrasing activities in her own words was (at least) equally legitimate, which led to less precise information (e.g. "talking with colleagues" instead of "having a meeting") – and that despite her impressive motivation to comply with the task.

The respondent used the possibility to go to the detail view of the next time-slot directly from the previous one only after having filled in 3,5 hours of slots. Before that she keeps going back to the overview and continuous to do so after having discovered the alternative. In the interview she states that the overview served her as means of orientation that reminded her that the task was to fill in her activities in slots of 10 minutes each.

She perceived the copy-function as a "temptation" to take a short cut, i.e. to provide less detailed information.

Regarding the suggestions from the drop-down menu, she (falsely) thought that these were derived from texts she had written before. Hence, an explanation regarding the origin and purpose of the suggestions might be helpful, also with regard to issues of confidentiality.

The respondent did not enter activities that she considered all too trivial such as going to the toilette. Also, she revised down the length of her lunch break ("It was not actually an hour"), which also indicates that social desirability may have played a role. Moreover, this points to the fact that entries based on one's memory may be biased.

Happiness as part of the well-being questions was interpreted as "being in a good mood" and stress was interpreted as extraordinary events that cause additional chores (e.g. a family member being ill).

The respondent generally filled in the slot as if she had told a friend "what she had done today" and appreciated the app as an opportunity to reflect on how she spent the day. A means to make the mobile app more attractive might be, in her view, to provide an overview of one's day or a comparison of one's day with the one of other persons.

2) Sources of error

The **overview** is perceived as deterrent and demotivating. Yet, in light of her outstanding motivation to comply with the tasks she comes to grips with it and even appreciates it as a means of orientation (e.g. how long her break actually lasted and that the mobile app is about 10-minutes-slots, not half-hour-slots.)

Detail view: She falsely assumed that **main activities** referred to her occupational activity. As a consequence, she only fills in side activities at first and erroneously interprets the warning sign in the overview as indicating that the time-slot was filled in completely. Hence, the warning sign should not only indicate *that* there is a problem, but also *what* exactly it is. Also, it was not clear to her, how to distinguish main and side activities. Here, too, an explanation (e.g. in the form of a "?"- button) might be helpful.

Regarding the **suggestions from the drop-down menu** it is noteworthy that while the respondent showed an abundance of patience in order to find a suggestion that matches her actual activity at least remotely, she indicates in the interview that the large number of inappropriate suggestions constituted a risk of distraction.

When, using the zoom, she discovers that it is possible at the end of the detail view to freely enter a time in the "go to 'next time-slot'"-button, the zoom of her phone (Sony Xperia) stops working. Consequently, she can only guess where to click. The area of the entry fields, in which clicking is accepted as a command, however, is sufficiently large, such that her entries were registered even though she did not see the entry fields on the screen. The return button finally makes the app (resp. the zoom) function normally again. Here, however, she was unsure, whether the return button would take her back to the overview or end the app entirely with all data being lost.

Following this experience (the zoom not working when trying to freely choose the time span of her activity), she uses the **copy-function** in order to fill in activities that lasted several hours. Interestingly, she not only copies but also adds specific activities, i.e. the malfunctioning of the phone (and, hence, of the app) makes the information she provides more detailed.

Regarding the "**persons present**", the respondent felt that entering "alone" when she was with lots of other, yet unfamiliar, persons seemed inappropriate. Here too, an explanation regarding the question, which persons are relevant for the purpose of the study might help decrease unclarity and disconcertment.

The fact, that the **time up to which one sleeps** can only be entered in 10-minutes-steps based on the current time (e.g. sleeping until 6:19/29/39/… a.m., when the app is completed at 22:19 p.m.) causes **irritation**.

She perceives the questions concerning her **well-being** as inappropriate for the chosen time-slot (sleeping).

Female university student in her early 20s (presumably); no information available on IT and app experience

1) Summarizing account of the test procedure

The respondent generally filled in the slot as if she had told a friend "what she had done today" and appreciated the app as an opportunity to reflect on how she spent the day. What made filling in all the slots easier for her was the fact that she had planned the day that she now had to remember beforehand.

She goes to Google Play before even having read the **aviso letter**. When entering "time" (German "Zeit") there is no suggestion for the mobile app yet, such that she has to enter the full name of the mobile app in order to find it. The **download** only works after she has clicked on "install" three times, the loading takes 3 minutes (after the first minute she shows signs of impatience, which may also be due to her worries of insufficient storage space).

Two aspects regarding her general approach to mobile app seem noteworthy: The respondent keeps going back to the overview rather than going to the detail view of the next time-slot directly from the previous one. Also, once she has discovered that she can enter activities in her own words, she exclusively does so, even though she sometimes is unsure, whether these entries can be processed in data analysis.

The respondent feels that filling in all the slots is straining, but the copying function makes it acceptable.

If the mobile app provided interesting graphics of the day (e.g. summarizing activities under categories in bar charts), the respondent would feel that the mobile app also provided a personal benefit to her.

2) Sources of error

During the **download** of the mobile app, which takes relatively long (3 minutes), the respondent is worried she might not have enough storage space for the mobile app. When **starting** the mobile app, i.e. when entering the time she gets up, the digits of the time wheel slide to far up and down two times, such that it takes her three attempts until she gets it right.

Overview with the 144 10-minutes-slots is perceived as deterrent and demotivating.

In the **detail view**, she misunderstood "**main activities**" as occupational activities. Back in the overview, however, the real meaning of activities became clear. At first, the respondent assumed that it was mandatory to use one of the **suggestions from the drop-down menu** to fill in main and side activities. She also falsely assumed that clicking on the search-button would yield suggestions hence the search-button initially implied a risk of misunderstanding. Missing key words in the suggestion drop-down menu result in entries that may not be clear ("fertigmachen" as a synonym for "getting ready") and may, thus, cause problems for data analysis. Moreover, the respondent repeatedly enters (in her own words) two activities as main activities, rather than entering one as a main and one as a side activity. In the interview she indicates that the distinction between side and main activities was not clear to her. Similarly, the fact that means of transport and public places (e.g. university, park, train) are not available as "locations" leads to questionable entries (e.g. location "at home" when the

activities is "going to university by tram"). Likewise, regarding the "**persons present**", the respondent felt that entering "alone" when she was with lots of other, yet unfamiliar, persons seemed inappropriate e.g. taking a test at university). Here too, an explanation regarding the question, which persons are relevant for the purpose of the study might help decrease unclarity and disconcertment. The **definition of who is a member of the household** may require further precision or explanation, as the respondent indicates her sister as a member of the household in Vienna, while she do not share the household there but in Lower Austria. Similarly, the **location** "at home" ("zuhause") may require specification for respondents with two households, as is the case for her.

The respondent repeatedly experiences technical problems, i.e. where the touch screen does not accept entries for main activities and locations. After rebooting the app, it functions properly.

The respondent perceives the questions concerning her **well-being** as inappropriate for the chosen time-slot (sleeping).

Regarding the activity "sleeping", it is noteworthy that if in the first of these slots one field of entry is missing, in *all* of the slots appears a warning sign in the overview. And one can only make it disappear in *all* these slots at once, if the first sleeping-slot is corrected. Finding this mechanism out took the respondent quite a while and one may wonder, whether she would have shown the same level of motivation, had she filled in the mobile app at home, rather than in a laboratory setting.

Male in his mid- to late-40s, working (office at home); no information available on IT and app experience

1) Summarizing account of the test procedure

Generally, the respondent demonstrates a high level of reflexivity, weighing the pros and cons of the design of the mobile app. While he finds it tedious having to fill in many 10-minutes slots when the overall activity is the same for several hours (e.g. working), he thinks that, on the other hand, such short time-slots may make respondents give more detailed information.

Regarding the aviso letter, the respondent suggests that it might be helpful if it contained information on Statistics Austria, their work and that the study is a legitimate enterprise, as well as how to obtain the 30€s.

Both finding the mobile app in the app store, downloading, and installing it go smoothly.

Regarding his general approach to mobile app it seems noteworthy that he mostly tries picking activities from the suggestions. Since these often do not math his needs, he has to rephrase repeatedly – such rephrasing, however, being dependent on the respondent's motivation and language skills.

After 20 minutes, the respondent says that it became exhausting to concentrate and that it would decrease strain if one could indicate time-spans freely (e.g. from 8:45 a.m. until 11:50 a.m.) for activities other than sleeping.

2) Sources of error

Log-in and start: Minuscule as the first letter of the log-in and/or password are problematic, as the software often starts with capital letters automatically. The start is somewhat unclear with the respondent wondering about the definition of "getting up" as he had breakfast in bed.

Detail view: Moreover, the respondent repeatedly enters (in his own words) two activities as **main activities**, rather than entering one as a main and one as a side activity. In the interview he indicates that the distinction between side and main activities was not clear to him. Also, the is not sure about how detailed the information he gives should be (e.g. whether it is enough to enter "working at the office" is too general, i.e. whether it makes a difference to the researches if he is working on the computer or talking on the phone).

The respondent feels that having to fill in (or copy) many time slots, when an activity lasts several hours (e.g. working) is tedious and it would be better if one had the possibility to indicate a time-span freely (e.g. from 8:45 a.m. until 11:50 a.m.) for activities other than sleeping.

Persons present and locations: The respondent felt that it was difficult to decide whether to indicate "at home" or "at work" when he works at home. Also, he thought it was inappropriate that one can only choose "other location" for when in a means of transport.

Once, the respondent clicks on the **return-button**, which takes him back to the start instead of the overview (as expected). This causes a little shock, as it is unclear at first, whether the data he had entered already were saved or lost.

Regarding the **time up to which one sleeps**, the respondent does not realize immediately that the mobile app is based on a 24-hours-rhythem and that he can enter a time the next morning rather than midnight. Hence, this unclarity causes the respondent additional work with the mobile app.

15 years old female high school student; experienced IT and mobile app user

1) Summarizing account of the test procedure

She is an experienced mobile phone user and familiar with app stores. She uses the mobile phone primarily for entertainment, and criteria for the evaluation of the mobile apps are similar to chat apps and other games she is using on the mobile phone – e.g. she misses emojis. The graphic design as well as the lack of possibilities for creative personalisation (choosing colours but also time intervals herself) are demotivating. Moreover, she thinks that the mobile app programmers did not make an effort to make it more fun because they were not interested in the mobile app and in her entries. The test person thinks the mobile app is so boring that she wants to abort the test. The presentation and mode of operation do not fit her needs.

The test was made in the lab. Therefore, all activities were filled in at once.

The **Download** of the mobile app works without problems but slowly. She does not blame the mobile app but her limited data volume. While waiting, she looks at the preview of the mobile app. This is the first time she gets the notion that the mobile app is probably boring (the first picture is the Login Page presented).

Entering the code has to be repeated often and she needs help. For her, the question mark at the end of the password was not part of the code but part of the sentence, so the process is tedious. (-> highlighting it graphically)

Start: The question does not match her daily routines, and the test person is irritated that she cannot enter the time of getting up herself but has to choose from the times suggested (10-minute intervals). Activities done before getting up (in her case chatting with her friends and listening to music) get lost.

Overview: The high number of timeslots is demotivating, the problem is intensified because she does not find the copy function. (-> It would be helpful if the copy function would be visible at the overview page.) The test person does not start filling in for a while but is scrolling up and down and looks at all the time slots. She is not sure but fears that she has to fill in each single time slot. She decides not to fill in each slot but to enter "breakfast", "on the bus" and "lunch break", leaving free slots in-between. As the lunch break was one hour, she looks for a way to fill in a longer duration, but cannot find one and decides to only fill in the first 10 minutes of lunch break. The last input is testing the mobile app at the Statistics Austria. Remarkable is also which entries she fills in. Attending school for example is left out. It seems as only activities that do not take too long are filled in to avoid being obliged to fill in one activity repeatedly (as she did not find the copy function). On the other hand, the test person thinks that she should only enter activities which seem important. Suggestions of activities like clearing the table ("Essen wegräumen") are irritating.

Detail View: Other causes for not filling in time slots are caused by the functionality of the detail view. She thinks that she has to select one of the autocomplete suggestions. If an activity she fills in does not pop up as a suggestion ("take the underground", "mathematics"), she is annoyed and deletes the activity without trying other words (and leaves the time slot free).

She fills in the location as well as the people present, but is annoyed about the limited possibilities to choose. "Friends" and "park" are missing.

Finalization: She wants to finish the diary entry and the error message occurs that 136 time slots are not filled in yet. She is annoyed and wants to abort the test. As proposed by the observer, she enters "sleeping" for the whole day (using the time wheel) however ending with 24:00, as she assumes the day is finished by then.

2) Sources of error

Download:

1) Download is slow (-> decreasing motivation)

Login:

2) Code is entered wrongly for several times: the question mark at the end of the password is understood as part of the sentence) (-> highlighting the password graphically)

Overview:

3) Error warning (exclamation mark): She notices the exclamation mark but does not know what exactly was wrong – the exclamation mark *does not indicate what was wrong*. (-> the error message should give hints about the error)

Detail View:

4) The difference between the "ok" button and the "arrow" button is not clear. She is not sure how to return from the detail view to the overview.
5) Missing autocomplete suggestions: "take the underground", "mathematics" (she means the school subject mathematics)
6) Autocomplete Suggestions do sometimes mean the same (no clear distinctions): taking the bus (Bus fahren und Autobus fahren), reading the papers (Zeitung lesen, Tageszeitung lesen), talking with the child
7) Suggestions for locations and people are missing: friends and park
8) Household members: she does not know the definition

Finalization:

9) Unclear end of the diary entry: As proposed by the observer, she enters "sleeping" for the whole day (using the time wheel) however ending with 24:00, as she assumes the day is finished by then.

Well-being

10) It is problematic to ask this question at the end, because she is slightly annoyed that she has to think back to remember how she felt at that point.

12.1.8 Video no. u00144

18 years old male high school student; experienced IT and mobile app user

1) Summarizing account of the test procedure

The respondent first tries to find activity suggestions from the drop-down menu that fit his needs. But as these are not appropriate, he quickly starts entering activities in his own words. After ten time-slots (i.e. before he finds the copying function where his motivation is decreasing) he stops entering side activities, possibly as a result of strain. He does not go back to the overview but directly goes from the detail view of time-slot to the next. Potentially embarrassing activities (e.g. going to the toilette) or trivial/implicit activities (e.g. on the smartphone) are not indicated by the respondent.

2) Sources of error

Both loading the app store and **downloading the app** take quite a while (approx. five minutes)

The respondent repeatedly enters (in his own words) two activities as **main activities**, rather than entering one as a main and one as a side activity. In the interview he indicates that the distinction between side and main activities was not clear to him. Not finding the copying function during the first twelve entries caused him strain.

The respondent (falsely) assumes that the end of the day in the mobile app was 12 pm, such that he enters midnight as the **time up to which one sleeps**. Hence, the 24-hours-rhythm of the mobile app was not clear to him.

15 years old male high school student; experienced IT and mobile app user

1) Summarizing account of the test procedure

The test was made in the lab. Therefore, all activities were filled in at once. He works fast, concentrated and thoroughly, despite that he thinks that it is boring. He also thinks he does not benefit from the mobile app. He would welcome gamification, where one can collect points if she/he fills in a defined amount of time slots. He also thinks that it is strange to give information about very private affairs. However, he also thinks that the survey is useful. (->The guarantee of anonymity could be emphasized.)

Overview: The task is intuitively understandable although he apparently does not read the instructions. He fills in every time slot. The grade of detail is higher for the morning activities (breakfast, brushing teeth, leaving the house, waiting for the tram, …) and is decreasing when school starts (the broad category "school" is chosen). However, breaks between classes were described as well. (*The process of generating activities*) To fill in each class, he uses the copy function (as the duration is 50 minutes). He also uses the copy function when activities are changing but people or locations remain the same.

Detail View: Not only each time slot, but also each question in the detail view is answered without hesitation, however sometimes not in a very congruent way. He uses the different inputs in a flexible way. Sometimes side activities are chosen to describe an activity conducted parallel to a main activity ("breakfast" and "watching TV"), sometimes he chooses side activities to describe two activities conducted consecutively.

In the case of autocomplete suggestions in the detail view he at first sticks to the suggestions. He does not want to fill in activities that do not pop up in the drop-down menu. Thus problems occur because some activities are missing in the suggestions. Therefore the user changes the input to a totally different one. (He wants to enter lay down ("liegen"), but cannot find it and enters "brushing the teeth" instead of it.) In another case, the entries become meaningless. He wants to enter "going to school" but cannot find an appropriate suggestion. So he enters "going". Towards the end of the protocol, he does not use the suggestions but once in the case of an intimate activity. He writes "aufs Klo gehen" (which is a colloquial phrase for going to the toilet) and searches the suggestions for the appropriate phrase. As he cannot find it, he chooses "toilet" instead of it.

People present and locations: For the time slot "being at home" he first chooses "alone", then "other household members", thus the definition of being alone is not clear. In the case of movement from one location to another, it is difficult to specify the location. For taking the tram he chooses "another location" which does not indicate a movement.

Finalization: According to his understanding, the diary ends at 24:00 (which is also indicated by the time wheel), so for the sleeping period he fills in until 24:00. For the first time, he does not enter locations and people present. (-> Maybe he is irritated, as the process works differently this time?) He notices the error warning, but does not correct the missing information.

It is disturbing for him that his effort is not honoured when he finishes the diary.

Well-being: As only two questions are visible on the display of his mobile phone, he does not notice the third question and wants to finish the diary entry. However, he cannot finalize this step of the process. Only then he notices the third question.

2) Sources of error

Download and Login

1) Password: The star at the end of the code is not recognized as part of it (->highlighting of the password graphically)
2) Meaning of "getting up": he understands it as waking up

Overview

3) Activities do not fill the allotted time slot: sometimes he has to use the side activity input field to enter an activity, as both activities do not take 10 minutes. In that case he had to delete the side activity that he had entered before: first: main activity: taking the bus, side activity: playing with the mobile phone – then he changed the side activity into: waiting for the tram.
4) It is unclear if he reads the instruction texts. (the text is not very conspicuous)

Detail view

Main and side activities:

5) His definitions of side activities vary: sometimes they indicate two consecutive activities, sometimes two parallel activities. In the second case, he differentiates between more and less important activities (e.g.: classes at school: main activity: school, side activity: writing; during breaks: main activity: break, side activity: school)
6) In general, he has a very broad definition of activities: e.g. "school" as an activity

Autocomplete suggestions in the drop-down menu

7) It is unclear if one is obliged to use the autocomplete suggestions – this makes the process much more complicated (searching for an appropriate word) and leads to the input of different or vague activities ("brushing teeth" instead of "going to bed", "going" instead of "going to school")
8) Missing suggestions: leaving the house ("außer Haus gehen"), going to school ("in die Schule gehen"), lay down ("liegen"), going to the toilet (sowohl aufs Klo als auch auf die Toilette)
9) Inappropriate suggestions: sometimes inappropriate suggestions pop up, e.g. he enters "taking the tram" ("mit der Straßenbahn fahren") – the suggestions "band rehearsal" (Bandprobe) and "car wash" (Waschstraße) pop up and he erroneously chooses "band rehearsal" from the suggestions.

Locations and other people present

10) Being alone: he is not sure what being alone means (->Does it mean to conduct an activity alone? Does it mean that no other person is present in the flat/room?)
11) In the case of mobile activities it is not possible to correctly name a location
12) school is missing
13) schoolmate is missing
14) Unclear definition: household member

Finalization

15) Unclear end of the diary entry – 24:00 or in the next morning? He finishes the dairy entry with 24:00. This is also suggested by the time wheel, which does stop with 24:00.

Well Being

16) Only the first two questions are visible on the mobile phone display

32 years old female, university degree, full time employee, mobile app utilization: 20 mobile apps per day

1) Summarizing account of the test procedure

The test person is a Statistic Austria staff member. Therefore, she was highly motivated and interested to detect problems. However, also in her case, the number of time slots was too high.

The test was made in the lab. Therefore, all activities were filled in at once. She said it was hard to remember them in detail.

Overview: Every time slot is filled in. The grade of detail is high when she describes her morning routines, however the grade of detail is decreasing when the test person describes her workday. For her morning activities she lists taking a shower, drying her hair, getting dressed, going to work. Her workday is not subdivided into particular activities. Only the lunch break is described as a different activity. In the interview after the test she explains that it was difficult to remember all activities during her workday. To sum them up as "work" made it possible to protocol her day. (This is contrary to her morning activities, which she could remember much better.) It seems that the amount of details entered depends on the kind of activities. In this case, she also used the copy function. She wishes to have the possibility to copy an activity for a longer duration at once, as there are "extremely many slots".

Detail view: Not only the difficulty to remember activities, but also the restrictions set by the mobile app are limiting her possibilities of protocolling the day. (1) The input of activities is difficult because she thinks she has to select one of the autocomplete suggestions. However, to find the right suggestion is quite laborious because there is no possibility to search systematically and many suggestions are not appropriate. (2.) Locations and person present: As there is no possibility to fill in "colleagues" she chooses to fill in that she works alone. There is no possibility to protocol a location change, so she as the location for the activity "going to work", she enters "work".

The user tries to prevent errors which consequently leads to a limitation of the detail grade of the protocol. The test person is not sure about the meaning of side activities; therefore she only fills out one activity per slot.

Finalization: To finish the protocol, she fills in the duration of sleeping. Intuitively she wants to roll down the time wheel until the next morning, but the time wheel ends at 24:00. Changing dates from one day to the next is not intuitive.

Questions about her **Well-being**: Answering the question about stress level, the respondent refers to the specific point in time, but the other questions are answered in a more general way. Since she could not remember her stress level for the time period, she chose a middle answer.

2) Sources of error

In general:

(1) Font size in mobile app instructions is too small
(2) Difficulties of recollection (activities are filled in in the test lab)

Overview:

(3) Time periods do not add up: e.g. she went to work from 7:45 to 8:15

(4) Task is not clear: Level of detail: she does not know how detailed she should record her activities (e.g.: is "work" sufficient, or should it be writing e-mails, making phone calls, …)

(5) Difficulty of remembrance: if activities are not recorded immediately, it is hard to remember them later

Detail view:

Main activity and side activity:

(6) Task is unclear: is it required to enter these?

(7) Definition is not clear

Autocomplete Suggestions:

(8) *Performance*: Poor input responsiveness

(9) Difficult search: No logical order in suggestions, sometimes suggestions pop up that have nothing to do with the term that is required

(10) Does the user have to choose from the suggestions in the drop-down menu, or is it possible to enter an individual text (*Task not clear: are the suggestions compulsory?*)

(11) Missing suggestions: going to work, to blow-dry one's hair

(12) Sometimes, several suggestions have the same meaning: working at the office, working at the company

Location and other people present

(13) Other place: she thinks that more suggestions pop up if she selects "other place" (which does not happen)

(14) Other familiar person: definition is unclear: familiar to whom? To the respondent or the mobile app?

(15) Missing category: Colleagues

(16) Being alone: This option is chosen even though colleagues are present at work and her partner is present in the morning. She decides to choose this option, because the category colleagues are missing.

Copy-paste functionality

(17) Visibility: Icon to small

(18) Handling is unclear: she does not tap the copy button but the ok button

Finalization

Time wheel

(19) Handling is not intuitive: She cannot scroll down until the next morning, the time wheel ends at 24:00.

12.2 Summaries of the nine telephone interviews

12.2.1 Telephone interview no. u00089/u00093

27 years old female high school graduate in full-time employment; smartphone experience 9.5/10; app-utilization: daily, every hour; 10 mobile apps installed; phone: Samsung Galaxy A5 (2017)

3) Summarizing account of the test procedure

The respondent reports that both the beginning and the end (i.e. the download, log-in, start and the finalization) went smoothly and without any problems or misunderstandings.

During a work day, however, it was not possible for her to continuously fill in the time-slots. As a result, she filled them in afterwards – or even beforehand. Her dealing with the mobile app differed on the two days in the sense that on the first day she still tried to fill in the time-slots punctually. On the second day she entered up the information every two hours. In general, her discipline and accuracy decreased over time.

The respondent only used suggestions from the drop-down menu, when these were appropriate, otherwise she wrote activities in her own words.

In order for the mobile app to be useful to the respondent, she suggests a time line indicating how far she has made progress with filling in the time-slots. Also, she would appreciate it if it was possible to upload photos of what one is doing (like on instagram or facebook).

4) Sources of error

The respondent felt that the fact that the **aviso letter** was registered was tedious.

The **start** of the mobile app was confusing for the respondent: she found it hard to say when she "got up" since she was snoozing.

The respondent feels that having to fill in many time slots, when an activity lasts several hours (e.g. working) is tedious – despite the copy function. Consequently, it would be better if one had the possibility to indicate a time-span freely (e.g. from 8:45 a.m. until 11:50 a.m.) for activities other than sleeping and all activities should be organized according to fields (eating, sleeping, working, leisure time).

In the **detail view**, the respondent thinks that it would be useful if mandatory fields were marked as such. She thinks that the search for **suggestions from the drop-down menu** that fit her activities takes too long. Also, she is unsure how detailed the information she provides for main and side activities should be.

Regarding the **persons present**, the respondent suggests that it should differentiate between "friends" and "colleagues".

Moreover, the respondent reports a number of **technical problems**, which were annoying. First, that when choosing an activities X from the suggestions provided, the mobile app accepted the activities Y. Also, it was necessary all the time to click very slowly in order for the entries to be accepted by the mobile app.

The respondent is surprised and a little annoyed about the well-being questions as she misinterprets them as reminders to fill in the time-slots.

12.2.2 Telephone interview no. u00094

61 years old female, employed as a teacher; IT-experience: 9/10; smartphone experience 9/10; app-utilization: 9/10; phone: Sony Experia

5) Summarizing account of the test procedure

The respondent reports that the beginning (i.e. the download, log-in, and start) went smoothly and without any problems or misunderstandings. She finds the mobile app generally interesting as it brings her to mind what she is doing and what she is not doing. In her view, the overview of the 10-minutes time-slots is a good point of reference.

Regarding the activities, the respondent tries to choose terms from the drop-down menu, but finds them not appropriate. However, as soon as a suggestion roughly fits her activities she enters it (e.g. "coffee break" instead of "break"). Otherwise, she mostly uses broad terms to describe her activities (e.g. working) as it is not relevant to her, which task this actually involved in a given time-slot and it did not occur to her, that more detailed might be of interest.

The respondent reports having filled in the time-slots with different degrees of delay: while she completed some of them immediately, she filled in others one or two hours later.

As entering the **time up to which one sleeps** causes misunderstanding, it would be useful if the time wheel did not stop at 23 p.m. but continued until the next morning.

6) Sources of error (incl. socio-demographic variables)

In the **detail view**, it is not obvious to the respondent how **main and side activities** are defined. Also, she thinks it is exhausting to enter side activities and, consequently, leaves them out. Activities that last less than ten minutes as well as activities that last long but are perceived as comprehensive (e.g. "working") are entered using broad terms (e.g. "getting up" rather than "brushing teeth" and "going to the toilette").

In terms of the **suggestions from the drop-down menu**, the respondent reports that many of the suggestions did not fit with her activities.

The respondent experiences some **technical issues**, namely that her entries were delayed. Secondly, when using the **copy function**, the mobile app repeatedly crashed. As a result, she uses this function less and makes her entries manually.

Regarding the questions on her **well-being** she states that her entries were much influenced by her mood in this very situation and might have been quite different 30 minutes later.

Regarding the **time up to which one sleeps**, the respondent does not realize immediately that the mobile app is based on a 24-hours-rhythem, but enters midnight on the first day. The next day, however, she becomes aware that she can enter an hour of the next day.

12.2.3 Telephone interview no u00182 and u00191

- Only available in German - 55 Jahre, weiblich, Matura, ca. 10 Apps am Handy

7) Zusammenfassende Darstellung des Testvorgangs

Die Respondentin ist hoch motiviert. Ihre Probleme liegen darin, dass sie in der App Limitationen sieht, ihren Tag zu protokollieren. Sie fand die App gut zu bedienen, interessant und auch lustig. So freute sie sich etwa darüber, dass „Katze streicheln" als Tätigkeit vorgeschlagen wurde. Sie würde bei einer Befragung mitmachen.

Download und Login: Installationen von Apps – also auch von diesem App - führt ihr Ehemann durch.

Die Eintragungsfrequenz erfolgte vergleichsweise zeitnah, wenn sie die Möglichkeit dazu hatte – als maximale zeitliche Abstände gab sie 1,5 Stunden an. In der Arbeitszeit hatte sie öfter die Möglichkeit als in der Freizeit. Eitragungen konnten auch nebenbei erfolgen, wenn sie „eh grad aufs Handy schaue".

Overview: Die Aufgabenstellung auf der Übersichtsseite beschreibt sie als sofort verständlich. Sie meinte alle Felder (auch Orte und Personen) ausfüllen zu müssen. So glaubte sie, dass sie das Tagebuch nicht abschließen könne, wenn in einem Feld noch ein Rufzeichen aufschien. Das Rufzeichen befand sie dabei als nützlich.

Detail View: Tätigkeiten wurden aus den vorgeschlagenen autocomplete suggestions ausgewählt. Sie meint, es wäre nicht möglich gewesen, einen Freitext einzugeben. Dies bereitete Probleme, da Schlagwörter fehlten. Fehlende Schlagwörter waren (sie nennt sie beispielhaft): „vom Einkauf nach Hause fahren", „Einkäufe wegräumen". Sie beschrieb sowohl Haupt- als auch Nebentätigkeiten. Als Haupttätigkeit definierte sie die aufwendigere. Sie hätte gerne die Möglichkeit gehabt, drei Tätigkeiten pro Timeslot protokollieren zu können. Da sie dies nicht konnte (sie versuchte es mehrmals), konnte sie teilweise Tätigkeiten nicht festhalten. Meistens wurden die Ortsangaben ausgefüllt. Auch Personenangaben wurden meisten ausgefüllt. Für sie ist es logisch nichts einzutragen, wenn man Tätigkeiten ausführt, bei denen man ganz offensichtlich alleine ist, wie z.B. „duschen". Die Kopierfunktion wurde von Anfang an genutzt. Sie nutzte sie auch, wenn die Tätigkeit gleich blieb, aber sich mitanwesende Personen und Orte änderten. In diesem Fall wurden letztere nach dem Kopiervorgang korrigiert.

Well Being: Die Well-Being-Notifications wurden immer bemerkt. Das blinken ist ihr aufgefallen. Die Fragen wurden immer beantwortet. Die Frage wie glücklich sie sich fühle hat sie generell beantwortet.

8) Fehlerquellen

Download

    (1) Name der App im Brief anders als im Store

Detailansicht – autocomplete suggestions:

    (2) Verpflichtende autocomplete suggestions? Sie meint sie könne nur aus den Schlagwörtern wählen und keinen Freitext eingeben.
    (3) Schlagwörter fehlten: „vom Einkauf nach Hause fahren". „Einkäufe wegräumen"

Detailansicht Nebentätigkeiten

(4) Anzahl Nebentätigkeiten: Kann nicht mehr als eine Nebentätigkeit eingeben, versuchte es mehrmals

Detailansicht – Personen und Orte

(5) Bedeutung „alleine sein": Wenn sie Tätigkeiten alleine ausführt, ist es für sie logisch bei „mit wem" nichts einzutragen.

(6) bei mobilen Tätigkeiten (Autofahren, spazieren gehen) Ortsangaben unpassend. Wünscht sich „nach Hause gefahren", „in der Umgebung"

Well-Being

(7) Fragenbezug: Frage zum Glücklich sein hat sie eher ganz generell beantwortet.

- Only available in German - weiblich, sehr wenige Apps am Handy

9) Zusammenfassende Darstellung des Testvorgangs

Die App wird als zeitaufwendig empfunden. Besonders stört sie, dass der Tag in 10 Minuten Einheiten protokolliert werden soll. Hier geht es ihr vor allem darum, dass sie nicht die Möglichkeit hatte, längere Zeitdauern eingeben zu können. Auch fühlt sie sich persönlich angegriffen, als sie die Fehlermeldung kam, dass noch Einträge ausgefüllt werden sollten, da sie sich Mühe gegeben hatte. Sie will aber auch generell nicht an Befragungen teilnehmen, da sie sich dann überwacht fühlt. Sie weiß nicht, was mit ihren Daten passiert und auch, ob sie dann immer wieder wegen weiteren Befragungen angerufen wird, was sie als Belästigung empfindet. (Daher wollte sie auch keine Mittheilungen erhalten).

Dies führt dazu, dass sie den Vorgang bereits am ersten Tag beinahe abbricht und nur mit Hilfe der Kopierfunktion die Aufgabenstellung möglichst schnell beendet. Der zweite Tag wurde nicht protokolliert.

Bezüglich **Download des Apps, Login und Start** werden keine Vorgänge beschrieben, es wird aber als problemlos bezeichnet.

Overview und Detail View: Es wurden nur als wesentlich empfundene Tätigkeiten eingetragen. Kurze „Zwischensachen", wie Geschirrspüler ausräumen wurden weggelassen bzw. verallgemeinert. Damit zusammenhängend wurden nur Haupttätigkeiten ausgefüllt. Nebentätigkeiten sowie Angaben über Orte und Personen wurden nicht gemacht. Nebentätigkeiten wurden weiterhin nicht eingetragen. (Da die Möglichkeit bestand, das Feld frei zu lassen meinte sie, dass es nicht wichtig wäre eine Nebentätigkeit einzutragen.) Eine erste Umstellung ihrer Strategie erfolgte, als sie das Rufzeichen sah, nachdem sie den ersten Eintrag ausgefüllt hatte. Sie ergänzte Orte und Personen. Es wurde der Beginn einer Tätigkeit eingetragen, die im Folgenden frei gelassenen Timeslots sollten die Fortdauer dieser Tätigkeit suggerieren. (Beispiel: Eine Busfahrt dauerte eine Stunde, hier trug sie nur für die ersten 10 Minuten „Bus fahren" ein). Als sie das Tagebuch beenden wollte (nach 10 Einträgen), kam die Meldung, dass noch 127 Einträge auszufüllen wären. Hier wollte sie abbrechen, löste die Aufgabe aber, indem sie bei den fehlenden Einträgen *die Kopierfunktion* verwendete (die sie zu diesem Zeitpunkt entdeckte). Der zweite Tag wurde aber nicht protokolliert.

10) Fehlerquellen

Allgemein:

1) Speichern: Sie meint, dass sie das Tagebuch abschließen muss, damit die Einträge gespeichert werden

Overview:

2) Unklare Aufgabenstellung: Sie trägt nur den Beginn einer Tätigkeit ein, die im Folgenden frei gelassenen Timeslots sollten die Fortdauer dieser Tätigkeit suggerieren.
3) Unklare Aufgabenstellung: Es wurden nur als wesentlich empfundene Tätigkeiten eingetragen. Kurze „Zwischensachen", wie Geschirrspüler ausräumen wurden weggelassen bzw. verallgemeinert.

Detailansicht:

4) Kopierfunktion: Symbol nicht verständlich: Ihr ist das Zeichen schon zu Beginn aufgefallen, aber sie hat es für einen Inhaltsverzeichnis-Button gehalten.

5) Kopierfunktion: Intuitivität: Kopieren bedeutet für sie Copy und Paste. Der in der App gewählte Vorgang unterscheidet sich von diesem. Der Vorgang ist für sie nicht intuitiv.

6) Haupt- und Nebentätigkeiten: Sie weiß nicht was es bedeutet und hätte darüber nachdenken müssen.

7) Bedeutung der Kategorie „allein sein": In manchen Fällen fand sie es seltsam, gefragt zu werden, mit wem diese Tätigkeit ausgeführt wurde. „Es gibt ja Tätigkeiten da kann man nur allein sein. Z.B. schlafen, lesen etc."

Empfehlungen von ihr:

1) Stärkeres Solidaritätsgefühl erzeugen – suggerieren, dass „wir" etwas gemeinsam machen

2) Lebenshilfe: Zum Abschluss könnte die Tagesstruktur dargestellt werden und Empfehlungen gegeben werden.: Bsp.: „im Schnitt braucht der Mensch zwischen 6 und 9 Std. Schlaf am Tag. Sie haben im Schnitt der letzten beiden Tage 5 Stunden geschlafen".

12.2.5 Telephone interview no u00245 and u00250

- Only available in German - 31 Jahre, weiblich, Uni-Abschluss, 50 Apps am Handy, tägliche App-Nutzung.

11) Zusammenfassende Darstellung des Testvorgangs

Vor der Befragung war sie auf die App gespannt. Sie empfand die Tätigkeit dann aber als nervend und bricht die Befragung nach einiger Zeit ab. Sie findet die Anzahl der vorgeschriebenen Einträge zu hoch und den Vorgang diese zu tätigen zu kompliziert. Die Vorgänge beschreibt sie als user-unfreundlich und sinnlos. Da sie in ihrer Arbeit mit dem PC arbeitet, hat sie keine Lust sich auch in ihrer Freizeit mit einer "schlecht gemachten Technik" "herumzuschlagen". Müsste sie an der Befragung teilnehmen, würde sie die Tätigkeiten auf einem Blatt Papier dokumentieren und jemanden anderen dafür bezahlen, dies in das App einzutippen. Auch die Kopierfunktion wird als mühsam beschrieben. Sie würde die App mit 3 von 5 Sternen bewerten, da sie abschließend dennoch meint, dass die App dennoch einfach zu bedienen und logisch sei.

Der **Download** und das Installieren bereiteten keine Probleme. Das App war einfach zu finden, Icon und Brief werden als nützlich beschrieben. Die Frage, ob Mittheilungen geschickt werden können, wurde bejaht. Sie hat angenommen, dass Mittheilungen zur Erinnerung dienten.

Die Eintragungsfrequenz erfolgten in größeren zeitlichen Abständen: Am Protokoll Tag wurde zu Mittag mit angefangen, dann einmal um 17 Uhr und einmal um 22 Uhr Einträge getätigt. Sie wollte zu einem späteren Zeitpunkt noch weitere Einträge tätigen, hat aber darauf vergessen.

Overview: Die Aufgabenstellung auf der Übersichtsseite beschreibt sie als sofort verständlich. Die Aufgabenstellung wurde so verstanden, dass Zeitdauern angegeben werden können, also die Zeitspannen selbst eingegeben werden können und nicht die 10 Minuten Intervalle nicht bindend waren.

Detailview: Auf der Detailseite erkannte sie, dass sie jedes Feld einzeln ausfüllen musste. Dies wird als „Schreck" wahrgenommen. Sie meinte aber weiterhin, dass es genüge, den Beginnzeitpunkt einer Tätigkeit einzutragen. Tätigkeiten, meint sie, müssten aus den Schlagwörtern gewählt werden und es könne kein eigener Text eingegeben werden. Dies bereitet Probleme, da die Vorschläge nicht passend sind. Bis zum Ende des Protokolltages bemerkte sie nicht den Fehler, dass sie alle Einträge ausfüllen musste. Sie bemerkte erst, dass sie das Tagebuch nicht abgeschlossen hatte, als sie am nächsten Protokolltag für den zweiten Tag keine Aufzeichnungen machen konnte, da sie den ersten Tag noch abschließen musste. Als sie abschließen wollte kam die Fehlermeldung, dass sie noch weitere Einträge auszufüllen hatte. Diese Aufgabe löste sie, in dem sie kopierte. (Sie wusste aber bereits zuvor, dass es eine Kopierfunktion gab.) Dieser Vorgang dauerte 10 Minuten. Die Speicherung jedes Kopierbefehls dauerte lange und nach jedem Kopierbefehl wechselte sie wieder auf die Übersichtsseite.

Sie gestaltet die Vorgänge also jeweils sehr kompliziert (es ist für sie nicht ersichtlich, wie es einfacher geht), wodurch der Eindruck der User Unfreundlichkeit und der Zeitaufwendigkeit verstärkt wird.

12) Fehlerquellen

Download:

1) Die Namen des Apps im Store und im Brief waren unterschiedlich. Dies verunsicherte. Hierbei war die Detailbeschreibung im Store hilfreich, da die Statistik Austria genannt wurde.
2) Konnte Benutzername und Passwort nicht aus dem PDF kopieren, wurde als störend empfunden (Anm MP: Aviso ausnahmsweise per Email verschickt)

Übersichtsseite:

3) Aufgabenstellung: Unklar, wie detailliert die Einträge sein sollen.

Detailansicht – Orte und Personen

4) „andere bekannte Person" nur im Singular eintragen zu können war störend (hatte zwei Freundinnen zu Besuch, konnte das nicht angeben)

Detailansicht – autocomplete suggestions:

5) Fehlendes Schlagwort: Morgentoilette
6) Sie meint, sie darf nur Schlagwörter auswählen und keine eigenen Eingaben machen.
7) Keine passenden Schlagwörter: Sie findet keine Schlagwörter, die dem entsprechen, was sie eingeben will. (Sie findet keinen Ersatz für „Morgentoilette")

Detailansicht – Haupt- und Nebentätigkeit

8) Unklare Definition: Sie meint sie solle entweder eine Haupt- oder eine Nebentätigkeit eingeben. Bei der Bezeichnung als Haupt- oder Nebentätigkeit ginge es ihrer Ansicht nach um eine Klassifizierung: Sie meint:„Evtl. zum Beispiel wenn ich den ganzen Tag „arbeite" dann ist „Arbeiten" die Haupttätigkeit. Und wenn ich dann aber mal 10 Minuten damit verbringe vorm Kopierer zu stehen dann würde ich Hauptaktivität leer lassen und „Kopierer bearbeiten" in Nebentätigkeit eingeben." Da sie es aber nicht genau wusste, hat sie keine Nebentätigkeiten angegeben.

Empfehlungen von ihr:

1) Sie hätte gerne Erinnerungen bekommen, dass sie Einträge machen soll. (Am liebsten eine Erinnerung in der Früh, eine am Abend und eine am nächsten Tag.)
2) Persönlicher Nutzen: Erinnerungen mit Tätigkeiten verknüpfen: z.B.: Wenn die Tätigkeit Buch ausleihen ist, könnte daran eine Erinnerung geschaltet werden, wann das Buch wieder zurückgegeben werden muss. Es könnten – wie in einem Haushaltsbuch – auch Ausgaben mit Tätigkeiten verknüpft werden.
3) Da sie meint, sie müsse Tätigkeiten aus der Vorschlagsliste wählen, schlägt sie vor, in die Auswahlliste „kein passender Begriff gefunden" aufzunehmen um im Anschluss selbst einen Text eingeben zu können.
4) Unklarheit über Detailliertheit der Tätigkeitsangaben: Beispiel wäre hilfreich

12.2.6 Telephone interview no u00246 and u00252

- Only available in German - 33 Jahre, weiblich, Uni-Abschluss, mehr als 30 Apps am Handy, tägliche App-Nutzungen

13)Zusammenfassende Darstellung des Testvorgangs

Im Allgemeinen bewertet die Respondentin den Vorgang als gut und die App als selbsterklärend, sie gab der App aber nur 3 von 5 Sternen. Sie empfand aber auch einige Aspekte als mühsam und die Ladevorgänge waren teilweise verlangsamt.

Die Respondentin hatte das Bedürfnis sich auf die Befragung vorzubereiten. Sie versuchte daher sich am Vortag einzuloggen. Dies war nicht möglich, aber durch die Bilder im App Store war die Funktionsweise bereits gut vorstellbar.

Der **Download und das Login** am Testtag verlief ohne Probleme, sie weiß nicht mehr, ob sie eine Anfrage bekommen hatte, ob sie zukünftig Mitteilungen zugesandt bekommen will. Sie lehnt aber generell derartige Anfragen bei Apps ab und glaubt, dass sie automatisch auf „Nein" geklickt hat.

An den Testtagen hat die Respondentin nicht sofort, aber in regelmäßigen Abständen ihre Tätigkeiten eingetragen. Hierbei loggte sie sich jeweils neu ein. In der Arbeitszeit waren die Abstände mit jeweils zwei Stunden größer als in der Freizeit, dies wurde durch die Arbeitsintensität begründet. Es wurden sowohl vergangene als auch teilweise zukünftige, geplante Tätigkeiten eingetragen. Eingaben wurden im Nachhinein überprüft. In der Arbeitszeit wurden auch vermehrt geplante/zukünftige Aktivitäten eingetragen. Sie musste sich die ausgeführten Tätigkeiten (und den Zeitpunkt dieser Tätigkeiten) merken, um sie (in zwei Stunden Intervallen) eintragen zu können.

Overview: Die vorgegebenen 10 Minuten Timeslots beschrieb sie als nervend. Die Tätigkeiten selbst wurden eher genau eingetragen, sie wusste aber nicht, ob dies der Aufgabenstellung entsprach. Die Höhe des Detailgrads schloss sie aus den vorgegebenen kleinen Timeslots, die einen hohen Detailgrad suggerieren würden. Als Beispiel nennt sie, dass sie nicht wusste, ob „arbeiten" als Tätigkeit eingetragen werden soll, oder aber einzelne Aktionen wie „Meetings", „E-Mail schreiben" usw. Kaffeepausen wurden nicht dokumentiert. Generell wurde „arbeiten" weniger detailliert eingegeben als Tätigkeiten während der Freizeit.

Detail View: Meistens wurden nur Haupttätigkeiten eingetragen und keine Nebentätigkeiten angeführt. Dies war durch die Schwierigkeit begründet, Haupt- und Nebentätigkeiten voneinander abzugrenzen. Wurden mehrere Tätigkeiten während des vorgegebenen Zeitraums ausgeführt, versuchte sie eine Überkategorie über beide Tätigkeiten zu finden (wie Haarpflege statt föhnen und kämmen). War dies nicht möglich, hat sie manchmal die zweite Aktivität als Nebentätigkeit angegeben, meistens hat sie aber keine Nebentätigkeiten angeführt. Wenn möglich, hat sie sich der Schlagwortliste bedient um Tätigkeiten anzugeben. Dies beschleunigte den Vorgang. Wenn keine Vorschläge zutrafen, hat sie einen eigenen Text eingegeben und gab in dem Fall die Tätigkeiten weniger detailliert an. Die Kopierfunktion entdeckte sie beim dritten Timeslot. Über diese Funktion meinte sie, sie war dankbar, dass es sie gibt und dass sie gut funktioniert habe, aber dass der Vorgang langwierig war, da sie nach dem Kopieren immer auf „ok" tippen musste um den Eintrag zu speichern.

Zweck und Ziel der Fragen zum **Well-Being** sind für sie nicht klar.

14) Fehlerquellen

Beginn:

1) Bezug der Frage: „Damit wir das Tagebuch an ihren Tag anpassen können, sagen Sie uns bitte, wann ihr Tag begonnen hat. Wann sind sie heute aufgestanden?"
Frage ist uneindeutig. Wird das erste Erlebnis am Tag erfragt oder der Zeitpunkt des Aufstehens? Da sie erst nach 0 Uhr ins Bett gegangen ist, war sie sich zu Anfang nicht sicher, ob sie diesen Zeitpunkt (dies würde sich eher auf „wann hat Tag begonnen" beziehen) oder auf Zeitpunkt des Aufstehens nach dem Schlafen angeben soll. Sie hat sich aber für zweiteres entschieden.

Overview

2) Aufgabenstellung: Unklar wie detailliert Tätigkeiten angeben werden müssen. (Bsp. Reicht die Angabe „arbeiten" aus? Oder soll E-Mails schrieben, mit KollegInnen reden usw. angegeben werden.

Detail View

Haupt- und Nebentätigkeiten

3) Definition unklar: Schwierigkeit der Unterscheidung zwischen Haupt- und Nebentätigkeiten (Bsp.: Sie hat sich die Haare gekämmt und geföhnt. Was ist die Haupt- und was die Nebentätigkeit?
4) Wie verfahren, wenn mehr als zwei Tätigkeiten ausgeführt werden?

Autovervollständigungsfunktion

5) Dieselbe Tätigkeit wurde in unterschiedlichen Varianten aufgelistet
6) Schlagwörter fehlen: Bsp. U-Bahn fahren

Well Being

7) Bezug der Frage: Frage wie glücklich/unglücklich sich die befragte Person auf einer 7 stufigen Skala fühle:
Den Bezug der Frage auf einen bestimmten Zeitpunkt findet sie unverständlich. Sie hat die Frage daher eher allgemein auf die Situation bezogen.

Empfehlungen von ihr:

1) Wichtig war, dass man eine Pause von einem Tag zwischen den beiden Befragungstagen hatte. Wenn mehr als zwei Tage protokolliert werden müssten, dann wünscht sich die Respondentin mindestens eine Woche Pause dazwischen.
2) Schlägt vor, Nebentätigkeit zu "weitere Tätigkeiten" umzutaufen und optional n-viele hinzuzufügen.

12.2.7 Telephone interview X (x00301)

34 years old male university graduate working part-time; smartphone experience 8/10; mobile apps used all the time (58 mobile app in total installed); smartphone: Xiaomie Redmi 3

15) Summarizing account of the test procedure

The respondent reports that the download, log-in, and start went smoothly and, in particular, that it was convenient that one had to do the log-in only once a day. However, having to do the log-in everyday meant that on the second day, when he did not have the log-in date with him, he could only fill in the slots in the evening, i.e. in retrospect drawing on his memory. As a consequence, he may not have entered all the activities in the right the time-slots and the information he provided on the first day were generally more detailed.

Furthermore, the respondent reports that filling in so many time-slots was exhausting and that he generally thought that the mobile app was not great. However, due to the overview, the mobile app was generally self-explaining. Regarding side activities, he reports having written more in the more in the beginning and less over time.

The respondent said he experiences technical problems, i.e. entries not being accepted by the mobile app, which he made (probably) worse by clicking several times. Closing the mobile app and rebooting it helped in such cases.

Regarding the activities, he tried hard in order to find short, concise terms himself. He would appreciate it, if the database for suggestions from the drop-down menu were more comprehensive.

Moreover, the respondent feels that it would be helpful if a pop-up reminded him of the mobile app after 120 minutes of inactivity on the mobile app in order to avoid filling it in from memory.

Joy is irrelevant to the respondent when filling in the mobile app. Instead, what matters to him is the instruction by Statistics Austria to take part in the study as he trusts this institution and complies with its demands.

16) Sources of error

The respondent felt that the fact that the **aviso letter** was registered was tedious. He also thought that the **overview** with the mere quantity of the144 10-minutes time-slots was deterring.

Regarding the **detail** view, the respondent was unsure about which activities to fill in, in particular how **main and side activities** were defined. He interpreted the activity that lasted the longest as the main activity. Here, an instruction would be helpful.

The **suggestions from the drop-down menu** were helpful only to a very limited extend, as certain activities do not exist in there (e.g. activities concerning his cats). Hence, the respondent had to improvise a lot and feels that at bigger database would be better.

The respondent feels that the copying function is helpful – provided that it does not crash the mobile app. Another technical problem was that the mobile app repeatedly did not accept his entries, which he thought was straining.

The **locations** provided were highly unsatisfying, especially when being on the road. This option ("on the road", German "unterwegs sein") should be added, according to the respondent. Likewise, he thinks that the **persons present**-section was weak, too, since, regarding his colleagues from work he had to enter "other known persons" as they are not his friends despite the fact that he knows them for a long time.

The respondent perceives the scale on which he has to rate his **well-being** as too long, i.e. he cannot rate his well-being on a scale ranging from 1 to 10. He suggests that a shorter scale would be better – or a Kunin- rather than a Likert-scale, i.e. using emojis instead of numbers.

The error message at the end was not perceived as helpful, since the mobile app crashed at that point and some of the information was lost.

76 years old male, retired, apprenticeship and master craftsman; smartphone experience 4.5/10; app-utilization: daily, but not every hour

17) Summarizing account of the test procedure

The respondent reports that both the beginning and the end (i.e. the download, log-in, start and the finalization) went smoothly and without any problems or misunderstandings.

Making one's daily routines explicit was unusual and a little challenging (not aware of what she does normally)

The respondent seems quite motivated, as he indicates that one really has to make an effort in order to provide accurate information. To that end, he took notes whenever he could not enter the information directly into the mobile app and entered it up later on. In his view, without such discipline, there is a risk of inaccuracy.

18) Sources of error

The respondent feels that having to fill in many time slots, when an activity lasts several hours (e.g. working) is tedious – despite the copy function. Consequently, it would be better if one had the possibility to indicate a time-span freely (e.g. from 8:45 a.m. until 11:50 a.m.) for activities other than sleeping.

Regarding the **suggestions from the drop-down menu**, the respondent feels that looking for a key word takes longer than writing an activity in his own words. Also, these suggestions constitute a risk of choosing an activity that does not really fit, as it is more convenient than writing in one's own words. Moreover, he reports, that sometimes the activities he clicked on, were not accepted by the mobile app. In such cases, no sign appeared indicating to be patient. Typing in other activities did not help either.

Regarding the **time up to which one sleeps**, the respondent suggest that the day should end at midnight, in order to avoid confusion: For the date indicated in the mobile app stays the same as the day before when entering a morning hour.

The respondent is surprised and a little annoyed about the **well-being** questions and erroneously assumes that these questions apply to the present well-being (rather than to the time-slot indicated in the question).

12.2.9 Telephone interview Z (x00303)

14 years old male middle school student; smartphone experience 9/10; app-utilization: daily, every hour; 40 mobile apps installed; phone: Huawei, T8 light (on the 1st day), Sony Experia Xa (on the 2nd day)

19) Summarizing account of the test procedure

The respondent reports that both the beginning and the end (i.e. the download, log-in, start and the finalization) went smoothly and without any problems or misunderstandings. Also the overview page indicated the task clearly, such that he knew immediately what he had to do. Likewise, he reports that he found what he was looking for among the suggestions in the drop-down menu and that the copy function made the task easier, especially in case of activities that last longer.

Regarding his general approach to the mobile app, it is noteworthy that the respondent entered up much of the information in hindsight.

In order for the mobile app to be useful to the respondent, he suggests a graphic overview of the day (and, later, of the whole week), i.e. of what he has done.

20) Sources of error

In the **detail view**, the respondent first is unsure, whether he must choose from the **suggestions in the drop-down menu** or if he may enter activities in his own words.

Regarding the **persons present**, he feels that the category "other familiar persons" should differentiate between friends and parents. In terms of **locations**, the respondent thinks that one of the "other locations" should be "school" and "on the road".

The respondent considers the fact, that one can enter activities only the same da problematic as it implies the risk of loosing data. Another technical issue in his view is the circumstance that the mobile app decelerates over time and sometimes freezes.