Department of Electrical Engineering and Information Technology
Distributed Multimodal Information Processing Group
Prof. Dr. Matthias Kranz

An Investigation of Activity Logging Methods in User Studies

Eine Analyse von Methoden zur Aktivitätsaufzeichnung in Benutzerstudien

Barbara Elisabeth Schmid

Diploma Thesis

Author:	Barbara Elisabeth Schmid
Address:	███████████
Matriculation Number:	██████
Professor:	Prof. Dr. Matthias Kranz
Advisor:	Dipl.-Medieninf. Andreas Möller
Begin:	24.04.2012
End:	18.10.2012

Department of Electrical Engineering and Information Technology
Distributed Multimodal Information Processing Group
Prof. Dr. Matthias Kranz

Declaration

I declare under penalty of perjury that I wrote this Diploma Thesis entitled

An Investigation of Activity Logging Methods in User Studies
Eine Analyse von Methoden zur Aktivitätsaufzeichnung in Benutzerstudien

by myself and that I used no other than the specified sources and tools.

Munich, October 17, 2012

_____

Barbara Elisabeth Schmid

Barbara Elisabeth Schmid

Kurzfassung

Heutzutage steigt der Marktanteil für Smartphones im Vergleich zu einfachen Handys aufgrund ihrer größeren Vielfalt von Funktionen. Diese Funktionen ermöglichen es Forschern, verschiedene Studien mit Smartphones durchzuführen. Genauso wie Handys begleiten Smartphones ihre Besitzer meist den ganzen Tag und sind zusätzlich in der Lage verschiedenste Informationen zu sammeln, zum Beispiel durch die Nutzung ihrer Sensoren. Es wurden bereits Applikationen (Apps) entwickelt die Fragebögen anzeigen und die Antworten verarbeiten. Verschiedene Studien wurden mit Smartphones im Bereich der Medizin, des Sports, von Multimedia-Anwendungen oder der Smartphone-Nutzung an sich durchgeführt. Es wurden jedoch noch keine allgemeinen Richtlinien für die Forschung mit Smartphones formuliert.

Für diese Diplomarbeit wurde eine sechswöchige Studie durchgeführt, um allgemeine Richtlinien aufzustellen. Die Studie bestand aus 30 Teilnehmern, welche drei verschiedenen Gruppen mit jeweils einer anderen Forschungsmethode zugeteilt wurden. Die Forschungsmethoden waren: das Feedback-Tagebuch, die Intervall-gesteuerte Experience Sampling Methode und die Ereignis-gesteuerte Experience Sampling Methode. Um die subjektiven Einträge der Teilnehmer mit objektiven Daten zu vervollständigen wurde die Smartphone-Nutzung geloggt. Während der sechs Wochen mussten die Teilnehmer Fragebögen über Ihre Nutzung der Facebook- und der Mail-App direkt nach dem Schließen der Apps beantworten. Die Facebook-App wurde dabei als Repräsentant einer Anwendung aus der Freizeit gewählt, während die Mail-App eine Anwendung repräsentiert, die hauptsächlich während der Arbeitszeit genutzt wird. An das Beantworten der entsprechenden Fragebögen musste sich die Feedback-Tagebuch-Gruppe selbstständig erinnern. Die Intervall-gesteuerte Gruppe wurde zusätzlich einmal täglich benachrichtigt, wenn sie einen Fragebogen vergessen hatte. Für die Ereignis-gesteuerte Gruppe erschienen die Fragebögen automatisch nach der Nutzung der Facebook oder der Mail App.

Um die Forschung mit den drei verschiedenen Methoden und dem Logging zu ermöglichen, wurde die Questionnaire App für Android entwickelt. Die App präsentiert die Fragebögen und lädt die Antworten automatisch zu einem eigens implementierten Backend-Server hoch. Der Server speichert diese Antworten. Er verfügt außerdem über Werkzeuge zum Erstellen von Fragebögen und zur Evaluierung der gegebenen Antworten und der geloggten Daten der Teilnehmer.

Mit Hilfe dieser Software wurden die Daten während der Studie gesammelt und nach der Studie analysiert. Basierend auf der Analyse wurden allgemeine Richtlinien aufgestellt. Diese Richtlinien beinhalten, welche Forschungsmethode in den beiden Anwendungsfällen die höchste Antwortrate hat, sowie Empfehlungen für die Dauer von Studien mit den drei Methoden. Darüber hinaus wurden weitere Schlussfolgerungen aus dem Vergleich der angegebenen und der geloggten Daten gezogen.

Abstract

Nowadays, the market share of smartphones is increasing compared to common mobile phones due to their better appeal and greater variety of features. These features enable researchers to conduct various studies using smartphones. Like mobile phones, smartphones accompany the owner most of the day and are additionally able to collect real-life information with, for example, their sensors. Applications (apps) have been built that gather information with questionnaires. Various studies in the field of medicine, sports, multimedia consumption or the smartphone usage itself have already been conducted with them. However, no common guidelines have yet been investigated for scientific studies with smartphones.

For this diploma thesis, a six-week study was conducted in order to establish common guidelines. The study involved 30 participants that were equally distributed in three groups, each of them using a different research technique. The research techniques were: a feedback diary, an interval-triggered experience sampling method and an event-triggered experience sampling method. In order to complete the subjective entries of participants with objective data, data logging was added. During the six weeks the participants had to answer questionnaires on their Facebook and mail app usage directly after they closed the apps. The Facebook app, thereby, has been chosen as a representative of a leisure time object, while the mail app represents an object mostly used during work time. After an app usage, the feedback diary group had to remember the task of answering the corresponding questionnaires. The interval-triggered group got a daily notification in case they missed an entry and for the event-triggered group the forms appeared automatically on the screen.

An Android application, named Questionnaire app, has been written to enable the research with these three different techniques together with data logging. The application presents the questionnaires and uploads the answers automatically to the implemented backend server. The server saves the answers and offers tools to edit forms and evaluate the given answers and the logged data of the participants.

With this software, the data was collected during the study and analyzed after the study. Common guidelines were extracted from the analysis. These guidelines identify which research technique has the highest response rate in the two contexts and give recommendations for study durations with the three techniques. Additionally, further conclusions from the comparison of self-reported and logged data have been drawn.

Contents

Chapter 1.

Introduction

Mobile phones have become ubiquitous. Today, about 80% of the world's population owns one or more mobile phones[1]. Besides simple mobile phones, so called smartphones, manufactured by e.g. Nokia or RIM, entered the market in 2001[2]. But with Apple's iPhone, the trend to buy a smartphone started in 2007. Today, there are about five billion mobile phones used worldwide of which 1.08 billion are smartphones (see Figure 1.1). In 2011, about 491.4 million smartphones were shipped which meant an increase of 61.3% to the previous year[3]. The most common operating systems in 2011 were Android with 46.9% and iOS with 28.7% market share (see Figure 1.1). All these statistics show that the number of smartphones is rapidly increasing and that more and more people have direct access to some sort of mobile phone.

Due to their widespread availability, mobile phones and in particular smartphones offer interesting possibilities to companies as well as researching institutions. Obviously, the development of market-leading smartphones and applications requires a profound understanding of the costumers' needs and wishes. To improve their products, companies conduct studies by evaluating the usage of the phones and applications in real life. Some applications automatically collect usage statistics and ask the user for feedback. These statistics and the feedback are used to further improve the products.

Besides the studies for product optimization, the smartphone has also become an interesting tool for general research studies. A smartphone accompanies its owner all day, can automatically collect different kinds of data with its sensors and enables communication between the researcher and the subject. Using these functionalities, smartphones can, for example, replace a classic paper questionnaire in studies. In addition to writing down answers, a smartphone allows the user to take pictures and record audio and video content that offer more information for the researcher. By using electronic surveys, the researcher does not have to print and distribute paper questionnaires and is able to faster analyze the electronic answers.

A lot of researchers already conducted studies with the help of smartphones that investigate

---

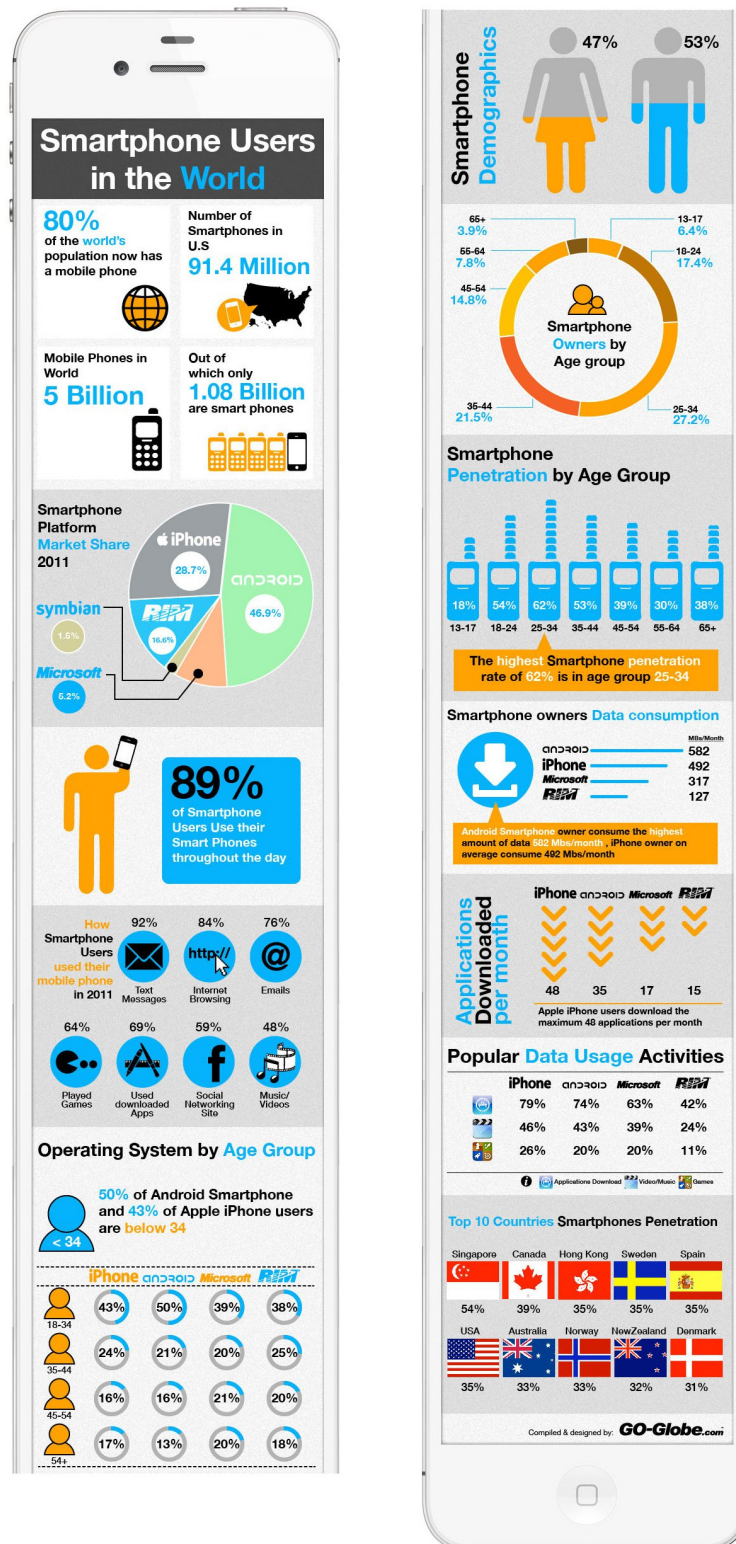[1] http://www.go-gulf.com/blog/smartphone
[2] http://en.wikipedia.org/wiki/Smartphone
[3] http://www.idc.com/getdoc.jsp?containerId=prUS23299912

Figure 1.1.: Statistics about smartphones.
Taken from: *http://www.go-gulf.com/blog/smartphone*

all fields of human life; monitoring and assessing levels of diseases, investigating friendships as well as researching the smartphone usage itself. Although various studies have been conducted, no research has been made to establish common guidelines for studies with smartphones. Such guidelines could aid the conductor of the study in making the necessary decisions prior to the execution of the study. The most important parameters for a study are the duration and the evaluation method. Depending on the way a study is conducted, it can pose a minor or major burden to the participants. Every additional day, week, month or year decreases their motivation and interest to further take part in the study. At some point in time, they drop out. Therefore, two important questions are:

1. Which way of asking the participants questions or requesting information is the best one without being too intrusive?

2. How long should the study last in order to gather as much information as possible without the majority of participants dropping out?

The aim of this thesis is to set up common guidelines for the choice of the method and the duration of a study conducted with smartphones.

A six-week study was conducted with 30 participants to investigate the response rate of three different modes in comparison to logged data. In order to conduct the study, an application was implemented for Android. The implemented Questionnaire app provides three modes to answer a questionnaire and enables to log the usage of other apps on the smartphone. With the help of this data, common guidelines are formulated. These guidelines indicate in which context which mode has the best response rates and after what period of time the response rates are too low to continue the execution. Furthermore, the comparison of the self-reports and the logged data will provide information about the reliability of self-estimated data from participants. The Questionnaire app and the established guidelines can be used for future studies with smartphones.

The following chapter will illustrate some common evaluation techniques. After that, a brief overview of some studies, which were conducted with smartphones, will be given. The second chapter will finish with some already existent survey tools for smartphones. In the third chapter, the concept of the planned study to set up guidelines will be formulated. The fourth chapter will describe the implemented Questionnaire app and its functionality. Thereafter, the pilot study as well as the main study, which were conducted with the Questionnaire app, will be presented in the fifth and sixth chapter, respectively. Chapter seven will discuss the combination of all data gathered in the main study. Chapter eight describes the results of the long-term effects study that was conducted after the main study. Finally, some conclusions will be drawn and future research directions will be proposed.

Chapter 2.

Research with Smartphones

This chapter lists some common research methods and explains them briefly. However, only some of them are applicable in the context of smartphones. Afterwards, notable studies from literature that were conducted with smartphones are presented. Finally, an overview of already existing survey apps, which enable to run a study, is given.

2.1. Research Methods

There are a variety of research methods for conducting studies. Some of them were designed for usability evaluations, others for medical science or psychology. In the following, the common methods

1. survey,

2. case study,

3. observation,

4. interview,

5. focus group,

6. experience sampling method,

7. diary and

8. logging

are outlined.

2.1.1. Survey

An often used research method is a *survey*. A survey consists of one or more questions that participants have to answer [1]. Usually, the answering process takes place in absence of a researcher. Nowadays, surveys are conducted, for example, with pen and paper, by a phone call, via email, on a web page or with a smartphone [2]. There are different types of surveys which can be conducted [3]. The first one is called pre-survey and is used before the actual study to get information, like demographic data, about the participants. These information are used after the actual study in order to understand the collected data by comparing the different participants with each other. The second type is a survey during the study. It is conducted after changes or after a period of time. The third type is the post-survey at the end of the study. The post-survey catches the overall impression of the participants. Besides these main three types, there are other specialized ones.

Surveys have the advantage that they are in general more time-saving and cost-effective than interviews. The survey can get a sample of data from many users in a short period of time [1, 4]. These users can even be distributed all over the world. Another advantage is that no interviewer has to be present. An interviewer can unconsciously influence the answers of the participants, for example, by asking questions in a specific way. This phenomenon is called the interviewer effect. Surveys also have some drawbacks. First, the response rate is lower than in an interview. Participants who should answer questions on their own have the choice to just drop out. These participants can cause a non-response error because often people with a similar background withdraw. Moreover, sampling or measurement errors can occur. A sampling error is caused by an undersized sample of test subjects; a measurement error by poorly formulated questions. Another drawback of a survey is that the answers do not have to be spontaneous. The researchers do not know under what conditions the questions were answered and if they were answered seriously. Finally, in case a researcher does not understand an answer, he is not able to dig deeper like in an interview [1, 4].

After the investigator chooses the survey as research method, he has to decide which kind of survey he wants to conduct. The advantage of a study with paper forms is that almost everybody is able to answer them, while internet, email and smartphone surveys require electronic devices. Surveys with electronic devices, in return, allow for multimedia content like audio or video records. The collected answers are saved in a database and do not have to be entered manually. There are high setup costs, but the overall costs to run a study are often lower than with paper forms because electronic surveys have to be set up once and can then be distributed for free to a large number of participants. An important drawback of electronic surveys is that the results are possibly not generalizable to the whole population as not everybody owns an electronic device. In case participants who do not own a device should take part at the study, researchers have to equip them with electronic devices, what can cause a lot of costs.

2.1.2. Case Study

Case studies are conducted to investigate a situation, an object or a person in real-life context [1]. In contrast to a survey, a case study is conducted to gather in-depth information. Therefore, the biggest drawback of this method is that only a small sample of participants can be investigated. The number of participants has to be chosen carefully. On the one hand, the amount of time spent for the in-depth investigation should be considered. On the other hand, the sample size should be big enough to secure the validity of the data.

The goal of a case study is to formulate or test hypotheses from the results. There are four different uses for a case study. A case study can be used to explore a new situation or to describe an existing object. Additionally, it is able to show the designers of a studied object the way their creation is used or explain the reason for this way of usage. Information for the case study is gathered by observing or interviewing the subjects.

2.1.3. Observation

A technique that is often used in case studies is the *observation*. With the observation of the participant, in-depth information about him can be collected [5]. There are three different ways to observe a participant [2, 4]. In an overt observation, the user knows that he is being observed; in a covert observation, on the contrary, the researcher does not identify himself and the participant does not know that he is being watched. For the third type of observation, the participant is instructed to self-observe himself.

Similar to a case study, observations can be conducted only with a small sample size as they are time-consuming. Additionally, the observer introduces a bias to the data. The bias is caused by the fact that people behave differently when they feel that they are being observed. Therefore, it is important that investigators are quiet and do not interrupt the subjects.

2.1.4. Interview

An *interview* is conducted by an interviewer, who asks the participants questions to gather opinions and other information. The questions are asked directly, via telephone or by chatting online [1]. There are three ways to conduct an interview: unstructured, semi-structured and fully structured [2]. An unstructured interview is like a common conversation. The researcher has no prepared questions and just listens to the participant talking about the topics that need to be discussed. The second type is a semi-structured interview, which requires some prepared questions. These questions are asked in order to initiate a conversation. The last type, a fully structured interview, has some sort of script the interviewer has to follow. All questions are prepared and have to be asked in the specified order. In comparison, a fully structured interview is easy to

conduct and to analyze. However, it does not offer as much possibilities to research the topics in-depth as an unstructured interview.

An interview has many advantages as a research technique. First, it provides detailed data that other techniques, like a questionnaire, cannot. Second, depending on the type of interview conducted, it enables to spontaneously formulate new answers. Third, it is able to ask follow-up questions in case an answer is not clear to the interviewer. Finally, an interview enables to notice the body language of the participant which is a good indicator whether the given answer was a lie.

Contrary to this, an interview is time-consuming [4, 6] because each participant has to be interviewed individually. Additionally, the interviewer has to be experienced. The way an interviewer formulates and asks the questions influences the interviewee. Furthermore, the interviewer has to write everything down and enter the results manually in a data sheet which can be time-consuming. Therefore, unlike an electronic survey, the analysis cannot be made immediately after the data is gathered. In case the interview is conducted with the aid of electronic devices, the data lacks of the body language clues of an one-to-one interview. Finally, the analysis of the qualitative data makes it more difficult to compare the participants.

2.1.5. Focus Group

A variation of an interview is a focus group [1, 4]. A focus group is conducted with a group of interviewees in order to gather several opinions in one meeting. Usually, a focus group includes six to nine participants and one session lasts up to two hours.

An advantage of a focus group is the easy data collection of multiple answers at once. In case the group is made up of diverse participants, their contrary opinions can lead to an informative discussion that provides insights in various opinions. Moreover, a focus group often gathers more data than an one-to-one interview because the participants feel more comfortable in a group.

A disadvantage of a focus group is that some people may lead the discussion and others stay silent. If that happens, the interviewer has to motivate these silent people to share their opinion. Hence, an experienced moderator is required.

2.1.6. Experience Sampling Method

The experience sampling method (ESM) is used to gather in-situ information from participants. In order to collect data in-situ, subjects are, for example, beeped whenever they have to fill out a form [7, 8]. This form asks them questions about the current situation or their current feelings [9]. The participants can either be interrupted randomly, after a defined interval, at specified times of the day or after an event [3, 10]. The event can be triggered by any sensor, for example, by a motion, location or audio sensor of a mobile device.

The advantage of the experience sampling method is that the participant does not have to recall events some time later, but can answer questions immediately [11]. In case the ESM is conducted with a smartphone, the phone triggers an alarm and records how long it took the subject to start answering the questions and how long answering took him. Additionally, it can timestamp data and automatically save it in a database. The participant is also able to take photographs, shoot videos or make audio records of a situation. The smartphone's biggest advantage, however, is that almost all situations can be captured because most people carry their mobile device with them all of the day [12].

The drawbacks of an experience-sampled study are the higher costs for setting up and maintaining such a system. Moreover, the automatic alarm may disturb in some situations and annoy the participant. The participant could also see the system as a burden because the sampling rate is too high or the duration of the study is too long. Therefore, careful considerations about the sampling intervals and the length of the study have to be made.

2.1.7. Diary

Similar to the experience sampling method, a diary catches the in-situ opinions and feelings of the participant over a specified period of time. Unlike ESM, a diary enables the participants to decide on their own, when to capture which information [1, 7–9, 13, 14]. Therefore, the user answers questions or takes notes either immediately after an event happened or at some later point of time. This indicates whether an event is important enough for a participant to write it down directly after it happened or if it is not so urgent, but still important enough to be noted. The diary may be written down on paper, with a computer or with a smartphone. A phone also enables the participant to call a special number in order to record her information as voice mail [8].

There are two kinds of diaries: an elicitation or a feedback study [1, 7]. On the one hand, an elicitation study requires the participant to capture something like a picture or a short note as a prompt for an interview [14]. The interviewer uses these prompts to talk with the participant about the events that led to them. A feedback study, on the other hand, asks the participant predefined questions. The participant may also capture a picture as information prompt, but the main goal of a feedback study is to collect all information at once. The researcher should be able to understand the situation and the participant's actions or feelings with the help of the answers. The structure of the diary has to be decided, as well [1, 8]. An unstructured diary provides the possibility to record activities in the way the participant wants to write them down. A structured diary, on the contrary, has pre-defined categories or options to describe an event with the given options.

Diary studies have many advantages. First, the recording of the information takes place in a real-life context and the data is therefore more reliable than data which is recorded at a later time [1]. Second, the user is not in a formal setting like in an interview and hence feels more

comfortable and gives more honest answers. Third, participants from all over the world can be acquired in a feedback study [1]. Additionally, the data is collected with low cost because no interviewer or researcher has to be present [7]. The usage of electronic devices such as computers or smartphones enables the capturing of videos or photographs and the immediate upload of the data [8].

A drawback of diary studies is that subjects could be overburdened and get annoyed in case the number of events or situations to be captured is too high. This can lead to a higher drop out rate than with interviews [1]. Furthermore, a diary entry is not always clear for the researcher and he would have to ask the participant further questions in order to get detailed information.

2.1.8. Logging

Logging is a technique that is used in combination with an electronic device, which automatically collects data by a system or a server. A participant uses the device without any interruption like filling out a questionnaire [5]. The collected data provides insight into usage patterns or into the problems a user has with the system. Data can either be extracted from web logs or stored application data. It can also be collected by web proxies or custom-built-software [1].

With this technique, no researcher has to be present and the participants use the tools without feeling observed [15]. Another advantage is that logging can be conducted with a great number of participants because the data collection is efficient in terms of costs and time [4]. Obviously, a big sample size increases the validity of the collected data as the study can be conducted over a long period of time. Due to the passive collection of data, it is not a burden to test subjects. Finally, the data logging technique can be applied to various parts of the system usage, like the clicked areas of a website, which are not easily collectable with other techniques.

The drawbacks of logging are that considerations about what to log have to be taken seriously. Logging everything creates giant databases that are hard to analyze [1]. The analysis can provide usage patterns, but data logging misses information about the intention of the user [5]. Therefore, data logging is often conducted in combination with other methods. Deane et al. [16] conducted a study to compare the results of data logging and a self-report. They found a high correlation between the two samples of data and a trend to slightly overestimate the duration and frequency in self-assessment. Therefore, they propose to use both techniques to complement each other.

2.2. Studies Conducted with Smartphones

In multiple research studies, some of the aforementioned techniques have been used in combination with smartphones to gather information. In the following, an overview of these studies will be given. Table 2.1 shows a summary of all mentioned studies.

2.2.1. Medical Science

Studies play an important role in the field of medical science. For example, medicines are tested before their release or a sick person's course of disease is being researched. Often, these studies cannot be conducted in laboratories because the long-term effect of a medicine or the daily routine of a sick person has to be considered. For studies conducted in the field, smartphones offer an easy way to get information from the patients.

For diabetes research, Jensen et al. [17] created a personalized mobile service that runs with a Bluetooth-enabled blood glucose meter and a mobile phone. The patient measures his blood glucose level and uploads the values with the phone to the server. The phone also enables the user to report the insulin intake or the physical exercises during the day and logs sensed data. The DiasNet service then predicts the future blood glucose level from the DIAS (Diabetes Advisory System) decision support system. A pilot study, which has been conducted with one user over the course of three months, was very successful. The participant welcomed the mobility as well as the ability to enter measurements during work without the need of a computer.

The MONARCA self-assessment system [18] is a monitoring system for bipolar disorder patients. The patient is able to self-assess her mood and enter other personal information into an Android app which also logs data of the sensors. The clinical staff is then able to view the data and suggest actions in risky situations. Before smartphones were used, patients had to write on paper forms and often forgot to answer them. A preliminary study revealed that patients prefer the usage of smartphones due to the remembering function and the ability to get a quick overview.

Jamison et al. [19] investigated the chronic low back pain of their 36 patients with palmtops and paper diaries in comparison. After a year, patients preferred the electronic way to monitor their pain, activity, mood and medication instead of using paper forms. A comparison between the two methods proved that the data from palmtops is valid and reliable and that patients tend to monitor their pain more often with a palmtop than with paper forms.

Patients with a chronic kidney disease (stage 5) have to monitor their intake. Hence, Siek et al. [20] built a PDA application to easily enter eaten food by scanning the barcode of the items or voice recording the intake. A three-week study with the app revealed that the six participants with low literacy skills had problems with voice recording the items. Apart from that, the app provided a convenient way to improve the awareness of the patients about their intakes.

2.2.2. Everyday Life

Hussain et al. [21, 22] implemented an app for content-based multimedia search in archives in order to stream and play the data on the phone. They investigated the app usage and acceptance in an one-week trial with 16 participants and an additional six-month field trial with 150 users. Subjects had to keep a diary, attend contextual interviews and their app usage was logged.

Table 2.1.: Studies with smartphones.

study	method	participants	duration	paper
Medical Science				
diabetes patients: monitoring life, forecast	data logging, diary	1	3 months	[17]
MONARCA self-assessment system for bipolar patients	data logging, diary	1	3 months	[18]
chronic low back pain diary	diary	36	1 year	[19]
chronic kidney disease: monitor intake	diary (bar code scanning, voice recording)	6	3 weeks	[20]
Everyday life				
multimedia search and usage	data logging, diary, interview	16	1 week	[21]
multimedia search and usage	data logging, diary, focus group, questionnaire, laboratory usability test	150	6 months	[22]
structure of friendship network	data logging, diary	94	9 months	[23]
life logging	diary	1	once	[24]
life logging: usage and privacy concerns	data logging, diary, semi-structured interview, questionnaire	13	11 weeks	[25]
shopping: new technologies for stores	diary, interview, questionnaire	13	2 weeks	[9]
navigational support for shopping	interview, questionnaire	20	once	[26]
mobile tourist guide	interview, questionnaire	17	once	[27]
personal awareness of fitness level	data logging, diary	13	3 weeks	[10]
level of daily fitness	data logging, diary, interview	9	10 days	[28]
job search for migrant workers in China	data logging, diary	82	2 months	[29]

continued from previous page

study	method	participants	duration	paper
Experience Clip	data logging, multimedia diary, interview	36	2 hours	[30]
information search, reading, producing	multimedia diary, interview, questionnaire	11	1 day	[7]
transit decisions	data logging, diary, interview	4	2 weeks	[7]
festival experience	multimedia diary and tangible objects, interview, questionnaire	7	1 day	[7]
voting with your feet	data logging, diary	16	4 weeks	[10]
Smartphone Usage				
users adaption of phone to their needs	data logging, diary, interview	21	3 weeks	[31]
user segmentation according to usage	data logging, diary, interview	50	2 weeks	[32]
app usage	data logging	11	2 days	[33]
app usage	data logging	255	7-28 weeks	[34]
app usage	data logging	4125	163 days	[35]
information search with phone	diary, interview, questionnaire	20	2 weeks	[36]
internet access: location and reason	diary, interview	19	7 days	[37]
videos on phone: reason and context	diary, interview	28	3 weeks	[38]
Going Wireless Study: integration of phone in user's life by novice owners	voice mail diary	19	6 weeks	[8]
Wireless Life-Cycle Panel Study: problems caused by new device	diary, focus group, interview, questionnaire	200	1 year	[8]

Friendships are an important part of life and, therefore, Eagle et al. [23] investigated the structure of friendship networks by using self-reported data as well as behavioural data taken from the phone. 94 subjects showed in a nine-month trial that the self-assessment of relationships is different compared to the data from the phone. A smartphone provides the functionality to collect both types of data and to complement the results with each other.

Smartphones also offer the possibility for a new kind of diary called life logging. Minamikawa et al. [24] and Kärkkäinen et al. [25] both developed such a life log system for smartphones to capture daily activities and review them later. Locations, images, emails, phone calls, text messages and played music tracks can be monitored. Kärkkäinen et al. [25] ran a study with 13 participants over the course of eleven weeks. They used semi-structured interviews twice during the study and a questionnaire at the end of the study. In case the users were not able to attend an interview because of the distance, a smartphone enables to show a form that has to be answered.
A similar self-documenting tool is called "Mobile Probes" [9]. It was tested in a shopping case study, which aimed at researching new technologies for a new kind of clothing sales point including automatic body measurements. The 13 participants, which were selected from sales staff or frequent buyers, had to document their shopping behavior during two weeks and attend an interview at the end of the study. Some results could not be explained without any help of the participants, while other analyzable data provided new insights for the salesmen.
Another shopping-related app was developed by Bohnenberger et al. [26]. The system provides a navigational support for shoppers who are looking for a specific item. A field trial with 20 participants in a mock-up mall compared the usage of PDA navigation to a paper map. In order to get the users' feedback, the participants had to fill out a questionnaire and were interviewed. Due to the system being used on the mobile phone, the questionnaire could have also been executed with the phone. The analysis revealed that the users were in average 11% faster with the electronic system than with the paper map.
Rukzio et al. [27] created a mobile tourist guide to present information about exhibits. The 17 participants tested three interaction techniques: user-mediated object selection, pointing and scanning. Questionnaires showed that the users preferred scanning and liked the system. In order to improve the efficiency, the questionnaires should be presented on the mobile phone itself.

A study conducted by Consolvo et al. [10], examined the personal awareness of the fitness level. Participants got a pedometer and had to enter their value together with a note into the app. The thirteen participants used the app for three weeks and experienced an increase in physical activity, especially because of the sharing of data with fitness buddies.
Another example for an app in the field of sport is "Shakra" [28]. The goal of the app was to increase the daily level of activity by logging the physical exercise and offering the ability to share this information with others. The app was able to log whether the test subject is sitting, walking or driving. Additionally, participants had to write a diary about their physical activity. The comparison of the logs and the one-week diary proved the app to be satisfyingly accurate.

The participants enjoyed using the app and the data revealed that they increased their activity level.

"Leho" also made use of the data logging and the diary method [29]. Leho was designed to help migrant workers find a job in China by using their mobile phones. Participants were recommended jobs and were able to browse through a list of all jobs. The two-month field trial with 82 users proved that the combination of an e-diary and logging is very efficient.

There are different methods to conduct studies in the field. Isomursu et al. [30] developed "Experience Clip". This technique requests to collect user experience by shooting video clips about the usage of, for example, an app. The app in the study was location-aware and made advertising proposals for shops or objects of interest. 36 participant teams consisting of friends used the app. One team member filmed the other team member while using the app. The method achieved a high subject compliance and the recorded clips represented a natural behavior with the app because a familiar person was filming.

Carter et al. [7] also investigated new methods for capturing information about participants with phones. There were three case studies with different methods. The first one investigated how people deal with information. Eleven users had to take a photo and write some notes for one day whenever they searched, read or produced information. This first case study revealed that photographs are triggers to remember objects. The second study investigated journey decisions. Four participants had to call a number whenever they used a public transport to give information about their journey. Additionally, their location was captured during the course of the two weeks. Interviews at the end of the study revealed that the locations were not very helpful to remember their intention for the journey. The third case study researched the experience of attending a festival. Seven visitors captured new events by shooting pictures, making recordings or collecting tangible objects. An interview proved that a photography is most helpful to remember events, while an audio recording can also be very useful. Tangible objects may be helpful recalling events, but the capacity to remember the location and persons involved was very poor.

Consolvo et al. [10] researched different in-situ evaluation techniques. One of their case studies investigated "voting with your feet". They hypothesized that the number and duration of a visit shows how much the person likes this place. 16 subjects had to use the "MyExperience" smartphone app and got questionnaires whenever they were longer than ten minutes at one place. The study revealed that the participants could not remember the places they have been in the past four weeks. The provision of the logged data at the end-interview helped to remind them of their activities.

2.2.3. Studying Smartphone Usage

Barkhuus et al. [31] investigated how users adapt their phone to their needs. They interviewed 21 participants and assigned them the task of writing a diary for three weeks. The participants had

to write down examples of their daily phone usage like the daily frequency of checking or writing emails with the phone in comparison to the computer. The researchers also used logs about the number of messages and calls in order to check the diary entries for their validity. Their results proved that each participant used her phone in a specific way even if another participant had the same phone model.

Jeon et al. [32] also investigated the usage of mobile phones with the diary method and by logging, for example, the frequency of menu access and usage context. Jeon et al. conducted their study with 50 participants for two weeks and compared the two data sets. Their goal was to segment the participants according to usage. Thereby, they discovered three types. The communicative-use type in general uses the phone to make telephone calls or write short messages. The restricted-use type does not use the phone very often at all and the entertainment-use type usually listens to music or watches clips. For this reason, Jeon et al. suggest new designs to fit these groups' needs. Demumieux et al. [33], Falaki et al. [34] and Böhmer et al. [35] investigated smart phone usage with a focus on the different applications. Demumieux et al. [33] logged the number, duration and name of the apps used. The data of eleven participants showed that a lot of time is spent on navigating phone menus.

Falaki et al. [34] further investigated the specific app usage of 33 Android and 222 Windows Mobile smartphone owners for seven to 28 weeks. With the data collected, they wanted to improve mechanisms to learn and adapt to user behavior. According to the data, the frequency and duration of an interaction are not correlated. Another result was that in average the number of used apps is 50. In general, only one app was used in one session with a median session length of under one minute.

Similar to this, Böhmer et al. [35] also researched app usage and found out that the session lengths of their 4,125 participants were on average less than one minute. Additionally, they specified that communication apps are used all over the day, news apps in the morning and games at night. The logged data of 163 days showed that the participants spent about one hour daily on their devices and used them especially in the afternoon and the evening.

A two-week study about mobile information searches was conducted by Sohn et al. [36]. They asked 20 participants to note whenever they needed an information, what information they needed and how they searched for it. Three supplementary interviews helped the researchers to clarify entries. Sohn et al. investigated, how the participants assessed an information not being important enough to be searched immediately with the mobile phone. The data showed four types of information, which differ in their urgency to be looked up in consideration of the mobile network costs caused.

Nylander et al. [37] gathered information about the location and reason for internet access. A seven-days diary and an additional interview of nineteen participants revealed that they often used the mobile phone for browsing even though they had access to a computer.

The diary and interview combination was also successfully employed to investigate the reasons for

watching videos on mobile devices like mobile phones or media players [38]. The collected data showed that most often clips were watched to kill time while a boring or unpleasant task, like waiting or eating alone, and were also used in groups to socialize. In this study, 28 participants took part for three weeks.

The diary study method was extended by Palen et al. [8]. They conducted two studies in which participants had to make their entries with a voice-mail system. The first study, named "Going wireless study" was conducted with 19 participants over a six-week period and investigated along with interviews how the first time owners of phones integrated them into their lifes. The following study, the Wireless Life-Cycle Panel Study, ran for a year with 200 participants. It focused on the problems that were caused by the new device. In comparison to the first study, the participation rate was lower due to the smaller number of personal contacts with the investigators.

In summary, electronic diaries have the advantage of a higher subject compliance than paper diaries [39]. Even users with low computer experience often have no problems in using electronic devices. Additionally, the data of electronic diaries is more reliable and of a higher quality. It is also faster analyzable than data gathered with other research methods. Therefore, electronic diaries are a good method for conducting a study.

2.3. Research Apps

So far, we know that studies with electronic devices combine several advantages compared to others with, for example, pen and paper. Besides the aforementioned research apps for the data collection from participants in section 2.2, this section introduces some others in the following. Table 2.2 lists the research apps and some key parameters.

"MyExperience" [5] is made for studies, where participants should be asked with the experience sampling method. It triggers questionnaires whenever the context requires it. Additionally, the context, the environment and the usage of the phone is logged. The logging is performed by sensors, which can be configured by the researchers. An important design specification was that participants should not be annoyed by fast battery drain, interruptions at inappropriate moments or a bad user interface. The questions asked are either of open or closed form. A closed question has to be answered with a short phrase or by using a given answer option, while a open question requests for a long text answer. In addition, multimedia content can be recorded. Researchers who used this tool found it very useful for conducting a study. The drawback of this system is the fact that it is designed only for Windows phones.

A system that is made by Carter et al. for iPhones is "Momento" [40]. It is also designed for situated evaluation with the data logging, the experience sampling, the diary and other qualitative methods. Additionally, it provides notifications for the researchers according to specific sensor information about the participant. Carter et al. analyzed the feedback of researchers who used

Momento and were able to identify future improvements that are going to be made. Overall, Momento was seen as a useful tool for conducting studies.

"ContextPhone" [41] is a platform to develop new context-aware applications for Symbian OS. It is built up of four modules (communications, customizable applications, sensors and system services) to help new developers gather all information they need and to use pre-existing tools for their app. The platform got positive reviews from researchers who used it.

Table 2.2.: Research apps for smartphones.

name	method	operating system	costs	paper
MyExperience	data logging, multimedia diary, ESM	Windows	free	[5]
Momento	data logging, diary, ESM	iOS	free	[40]
ContextPhone	platform to develop context aware app	Symbian OS	free	[41]
LifePod	data logging, diary	KDDI phone	free	[24]
Experience Clip	data logging, multimedia diary	PDA phone	free	[30]
App Sensor (in appazaar)	data logging	Android	free	[35]
droidSurvey	multimedia survey, GPS logging	Android	with costs	-
mQuest Survey	multimedia survey, GPS logging	Android, iOS	with costs	-
SurveyToGo	multimedia survey, GPS logging	Android, Windows mobile	with costs	-
EpiCollect	data logging, multimedia survey, GPS logging	Android, iOS	free	-

There are also research apps for Android smartphones available in the Google Play store. In the following, a closer look will be taken on four of them.

The first one is called "droid Survey"[1] and can be used offline. The collected data is available for download or can be viewed online with the help of graphical tools. It offers ten question types and the possibility to cancel questions based on the previous answers. Due to the ability to use multiple languages and GPS, studies can be conducted all over the world. Unfortunately, this app cannot be used for free, but requires the researchers to pay for a monthly subscription. The subscription allows to perform studies with an unlimited amount of users and devices and a limited amount of results.

Similar to that app, "mQuest Survey"[2] offers comparable features to conduct an offline study. Additionally, mQuest Survey provides the diary study method, taking photographs and recording

---

[1]https://play.google.com/store/apps/details?id=com.contact.droidSURVEY, https://www.droidsurvey.com/
[2]https://play.google.com/store/apps/details?id=de.cluetec.mQuestSurvey, http://www.mquest.eu/

audio. Another advantage of the app is the iPhone compatibility. However, like droid Survey, the service is not free.

A third survey app is called "SurveyToGo"[3]. It offers the additional functionality to record videos and ask 13 different question types. It is offered for Android and Windows Mobile. Similar to the other two apps, it has to be paid per use.

In comparison to these three apps, EpiCollect[4] is a free, data collection tool for Android and iOS. It offers to gather data with questionnaires and view it online or on the phone. The tool offers the GPS functionality and four question types. The entries made can be reviewed on the phone. In case they are uploaded, the GPS data will be exposed on a map and can be sorted according to time. Moreover, a possibility to communicate with participants via Google Talk was added.

---

[3]https://play.google.com/store/apps/details?id=dooblo.surveytogo, http://www.dooblo.net/stgi/surveytogo.aspx
[4]https://play.google.com/store/apps/details?id=uk.ac.imperial.epi_collect, http://www.epicollect.net/

Chapter 3.

Concept

The aim of this diploma thesis is to set up common guidelines for conducting a study with smartphones. Smartphones offer different ways to conduct a study as seen in Section 2.2. It provides the possibility to run a survey, a telephone interview, a diary study, use the experience sampling or the data logging method. In order to choose the appropriate evaluation technique, the pros and cons have to be weighted carefully. A common method to investigate a hypothesis is to combine a diary study with data logging. The combination yields personal insights and opinions of the participant as well as objective data from the smartphone usage. The objective data helps to verify the self-reported data of the participant according to the time the entry was made and other loggable information. Therefore, different research techniques are going to be compared in this thesis. These are

1. a feedback diary,

2. an interval-triggered, experience-sampled diary,

3. an event-triggered, experience-sampled diary and

4. data logging.

All three diary methods are evaluated together with the data logging to complete the subjective data entries of the participants with objective data. The data entries include open and closed questions. Closed questions offer options to the participant and ask him to choose one or more of them. Open questions request the user to make a free text entry to gather his opinion in his own words. The first technique is a simple feedback diary. The feedback diary requires the test subject to self-assess whether his current situation affords to make a note. The researcher gives the participant some common instructions when he has to create an entry. The participant, however, can decide when she wants to write something down. The second technique is similar to the first one and adds the ability to trigger notifications. The interval in which these notifications are sent is definable. After the defined interval has passed, the smartphone triggers a notification in case the subject forgot to make an entry after an event occurred. Finally, the third technique

eases the burden for the participant to remember writing something down. This technique automatically triggers the corresponding entry request which the user has to complete. The second and third technique are also called experience sampling method (see Section 2.1). Additionally to the notifications and the automatically triggered questionnaires, participants are invited to make a missed entry at some later point in time.

Besides the app on the smartphone, a server is required to provide the questions that are prompted. A server is also necessary to save the data supplied by the phone. The set of questions should be easily creatable and editable in case some changes have to be made. Therefore, tools to create and edit a set of questions as well as to take a look at the collected data are necessary.

The already existing apps, which were discussed in Section 2.2, are not appropriate for the requirements of this work. The app has to provide all four aforementioned techniques in order to compare them with each other. Additionally, the preferred operating system for the app is Android. As explained in the introduction, the majority of smartphones used today have Android as operating system and it enables to gain a maximum number of participants.

The research techniques should be compared in two different contexts because a context can influence the response rate. In order to gain information about two different contexts, the study investigates the usage of a mail and the Facebook app. Mail apps are used by most smartphone owners. If the phone is used for work purposes, participants have to start the app many times during one day to read or write an email. Therefore, making an entry about every usage might be sometimes inappropriate. Contrary to a mail app, the Facebook app is not pre-installed on all smartphones and is normally used during leisure. The Facebook app enables to connect with friends by, for example, writing or reading posts. This app is rather not used in a hurry, but in spare time or to take a short break. The Facebook app is chosen as contrast to the mail app and because it is one of the most frequently downloaded apps.

As can be seen by the briefly illustrated studies in Section 2.2, the duration of the conducted studies varies from two hours up to one year. The reasonable duration of a study depends on the topic that is investigated and the number of samples a researcher needs to test hypotheses. Additionally to the required data sample, the researcher has to choose the evaluation method in a way that the participant is not overburdened and drops out. Therefore, the study conducted for this thesis runs for six weeks, which is a rather long time period for daily entries. It is assumed that motivated users will make entries for the whole time and unmotivated users are likely to drop out after a few weeks.

In summary, an app has to be implemented, which offers the possibility to conduct a study with the three diary techniques as well as data logging. A server has to be set up, to provide a set of questions and to collect the supplied data. Then, a study has to be prepared that runs the different techniques. Therefore, three groups are required. In order to gather a significant data sample, ten participants per group are preferred. Each participant will have to answer questions after she used the Facebook or one of the various mail apps. After the study, the collected responses of

the three groups are analyzed. First, the combination of the response rates and the logged data will reveal, which technique has the highest number of answered questionnaires for the different apps. With this information, a first guideline for the choice of the evaluation technique in a special context can be made. The collected data is then analyzed with regard to the response rates over time. The group-specific response rates of each week will be compared to each other. The first week will most likely have the highest response rates and with each week this rate will probably decrease. This data provides information to establish another guideline. For each technique, the rates should indicate a reasonable duration of a study before the participants get unmotivated or even drop out.

Chapter 4.

Questionnaire App

This chapter presents the implemented Questionnaire app. First, the individual components of the app are named and briefly described. Second, the backend server and the supplementary tools to create and edit questionnaires and analyze the data are illustrated. Third, the communication of all these components is explained. Finally, the single steps to conduct a study with the Questionnaire app are listed.

4.1. Structure of the Questionnaire App

The Questionnaire app was implemented for Android. Android was chosen as the operating system due to having the biggest market share. In order to support most Android phones, API level 10 was chosen which corresponds to the Android version 2.3.3. The user interface of the Questionnaire app was translated to English and German.

Figure 4.1 shows the components of the Questionnaire app that are explained in the following.

4.1.1. QuestionnaireService

QuestionnaireService is the background service of the app. Its main functionality is the communication with the backend server and the data transfer inside the app.

The first task of communication with the backend server is the download of available questionnaires. In case the study is conducted over a longer period of time, it is useful to be able to add new or updated questionnaires to the Questionnaire app. QuestionnaireService downloads these and saves them on the smartphone. Newly downloaded questionnaires are not shown to the user immediately, but have a start date. QuestionnaireService checks if the questionnaires' start dates are already reached and whether the questionnaires should be displayed to the participant. It also has to make sure that all expired questionnaires are removed from the list of available questionnaires.
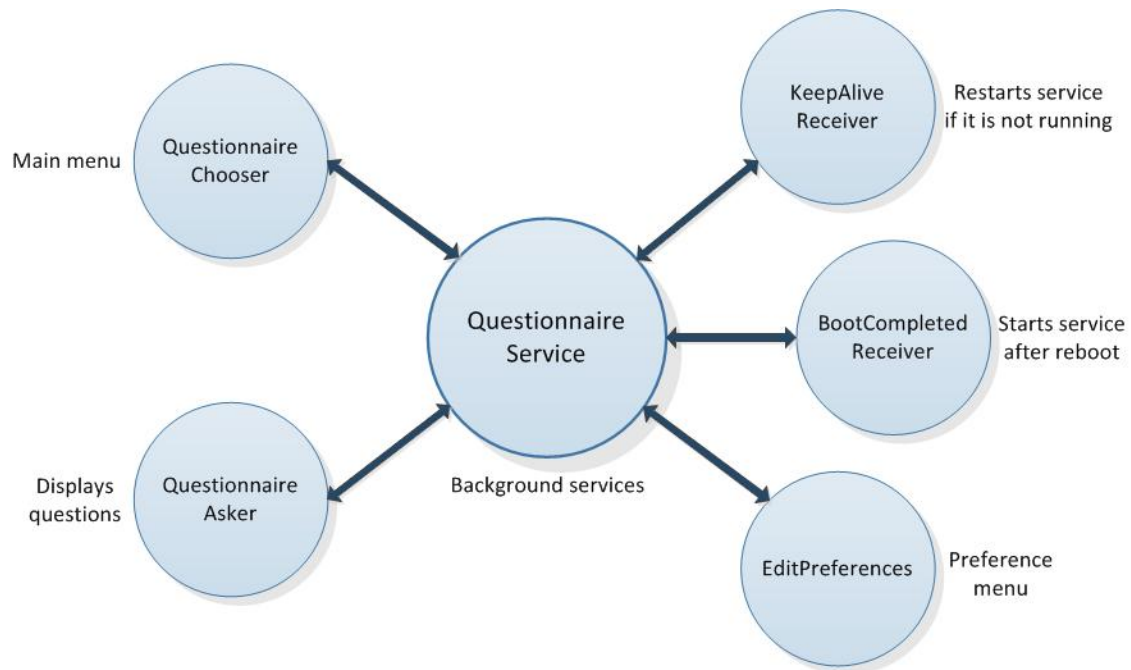
Figure 4.1.: Questionnaire app components. QuestionnaireService provides the background services of the app. QuestionnaierChooser and QuestionnaireAsker present a list of available questionnaires and the questions of the chosen questionnaire, respectively. The two receivers are responsible for keeping QuestionnaireService running and Edit-Preferences is the preference menu of the Questionnaire app.

The second aspect of server communication takes place after QuestionnaireService received an answered questionnaire from QuestionnaireAsker. These answers are uploaded to the server as well as saved on the SD card of the smartphone. After the installation of the app, QuestionnaireService creates a random user ID for each participant for the upload of his data to ensure privacy. The given answers and the logged data are matched to a user ID, but the conductor of the study does not know who this person is. On its first start, QuestionnaireService retrieves technical information about the smartphone and sends them to the server. These information include, for example, the device name and the Android version as well as available features of the phone and the questionnaire mode.

QuestionnaireService is responsible for checking which questionnaire mode was chosen in the configuration variable section of its code. There are three possible questionnaire modes: *voluntary*, *interval-triggered* and *event triggered*. In the *voluntary* questionnaire mode, the participants have to remember to fill out the questionnaires. The *interval-triggered* mode activates a notification at a definable interval. This notification is only triggered if the participant used a logged app and forgot to answer the corresponding questionnaire. The third questionnaire mode, *event-triggered*, opens the appropriate questionnaire automatically as soon as the logged app has been closed and the phone is going to switch to the home screen.

Another functionality of QuestionnaireService are the different alarms. There are

- the manual upload reminder alarm (triggers every murTime),

- the upload manual alarm,

- the scheduled questionnaire alarm,

- the periodic downloader alarm (triggers every pdTime),

- the save log alarm (triggers every slTime) and

- the questionnaire mover alarm.

The manual upload reminder alarm informs the participant to upload answered questionnaires with a notification. Usually, the data is uploaded automatically. However, if the participant chooses to upload the data manually in the settings menu of the Questionnaire app, a notification will be shown after *murTime* has passed. Clicking the notification field triggers the *upload manual* signal, which causes QuestionnaireService to send the data to the server. The next alarm, the scheduled questionnaire alarm, is activated in case the questionnaire mode is interval-triggered. This alarm triggers a notification to answer a questionnaire that has been previously forgotten. The periodic downloader alarm triggers after *pdTime* passed and causes QuestionnaireService to try to download new or updated questionnaires from the server. The save log alarm has a definable period and is activated after the passing of *slTime*. QuestionnaireService then uploads the current log to the server and additionally saves it on the SD card. The last alarm, the questionnaire mover alarm, is triggered in two cases. The first case is when a questionnaire reaches its start date and should be answered by the participants from that moment on. The other possibility is when a questionnaire should not be filled out any more because it expired. After the questionnaire mover alarm is triggered, QuestionnaireService updates the list of currently available questionnaires for QuestionnaireChooser.

Another task of QuestionnaireService is to log the usage of apps with an interval of one second. The specification which apps will be logged depends on the active questionnaires. QuestionnaireService logs the date and time, the duration of the used app and the GPS location. In order to better analyze the sequence of the used apps, the name of the subsequent app is also logged. Like the answers, the logs are uploaded to the server whenever a user finishes a questionnaire and are additionally saved on the SD card. In case the questionnaires are updated by downloading new ones from the server, QuestionnaireService also updates the activities and package names which are being logged.

The last important task of QuestionnaireService is to prepare the data for QuestionnaireChooser. It has to check whether all logged apps are already installed on the smartphone. If this is not the case, the names of all missing apps are sent to QuestionnaireChooser. In case all required apps are already installed, QuestionnaireService sends the names of all available questionnaires to QuestionnaireChooser.

4.1.2. QuestionnaireChooser

QuestionnaireChooser is the main menu of the Questionnaire app. Its first task after the installation is to display an information screen in a dialog window. This information screen asks the user to regularly fill out the questionnaires and indicates which data will be logged. Additionally, it specifies which app usages will be logged. The next screen of the QuestionnaireChooser is shown in case the user has not already installed all apps which are going to be logged. The screen indicates which app is missing and provides a button to automatically open the Google Play store. The Google Play store automatically searches for the missing app and opens the respective page in order to download it. After the download is completed or in case the user misses no required app, the main menu will be shown.

The main menu of the Questionnaire app presents all available questionnaires and Figure 4.2(a) illustrates an exemplary QuestionnaireChooser screen. In order to open a questionnaire, the user has to push the respective button. QuestionnaireChooser then sends QuestionnaireAsker the information which questionnaire was chosen. Another task of the QuestionnaireChooser is creating a connection to EditPreferences in order to show the options menu (see Figure 4.2(b)), which will be explained later in section 4.1.4.

(a) Chooser screen
(b) "Upload answers" and settings menu button

Figure 4.2.: Screenshots of QuestionnaireChooser, the main menu of the Questionnaire app.

4.1.3. QuestionnaireAsker

QuestionnaireAsker is started by QuestionnaireChooser as soon as a questionnaire has been chosen. Its first action is to load the questionnaire and extract the relevant data. QuestionnaireAsker then displays the questions one by one. A questionnaire can contain up to seven different question types that are given in Table 4.1.

Common to all question types is a text field at the beginning of each question. This text field is reserved for a question or a request that the user has to react to with the shown control elements. The first three question types illustrated in the table offer the possibility to present the answers as radio buttons, check boxes or simple buttons. The answer options to these question types should be limited to seven due to the limited smartphone screen size. *Dropdown* can handle more than seven answer options because the answers are presented in a scrollable drop-down menu. The question type *text* offers a common text field. It can be used to show some information or an image. *Edittext* has, additionally to the common text field, another smaller text field to show a prompt, a text input field and a "Done" button. The participant is able to enter a free formulated answer and confirm it with the "Done" button. The last question type is called *scaleedit* and offers a scale bar to graphically enter a value by dragging the slider. In case a participant rather wants to enter the value by typing the number, an input field is provided. Figure 4.3 shows a few possible QuestionnaireAsker screens.

Table 4.1.: The seven possible question types of the Questionnaire app.

Question type	Elements
radio	text, radio button
check	text, check box
likert	text, button
dropdown	text, drop-down menu
text	text
edittext	text, text, input field, "Done" button
scaleedit	text, scale bar, input field

QuestionnaireAsker checks each question for its type and generates the appropriate screens for it. If the user pushes the back button of the smartphone while a question is shown, a dialog window is opened (Figure 4.4(a)). The window asks the user if she really wants to quit the questionnaire and indicates that she is not returning to the last question. This dialog window was implemented to avoid frustration in case the user misinterprets the functionality of the back button. In order to gather spontaneous answers, the participant cannot return to the last question. After the user, however, has given an answer to a shown question, the "next" button is activated. By pushing the "next" button, he skips to the next question of the form and his answer is saved.

After the last question of the questionnaire, a "Thank You" screen is shown and the user is able to return to the main menu by pushing the corresponding button. All given answers of this questionnaire are assembled to a message and sent to the QuestionnaireService.
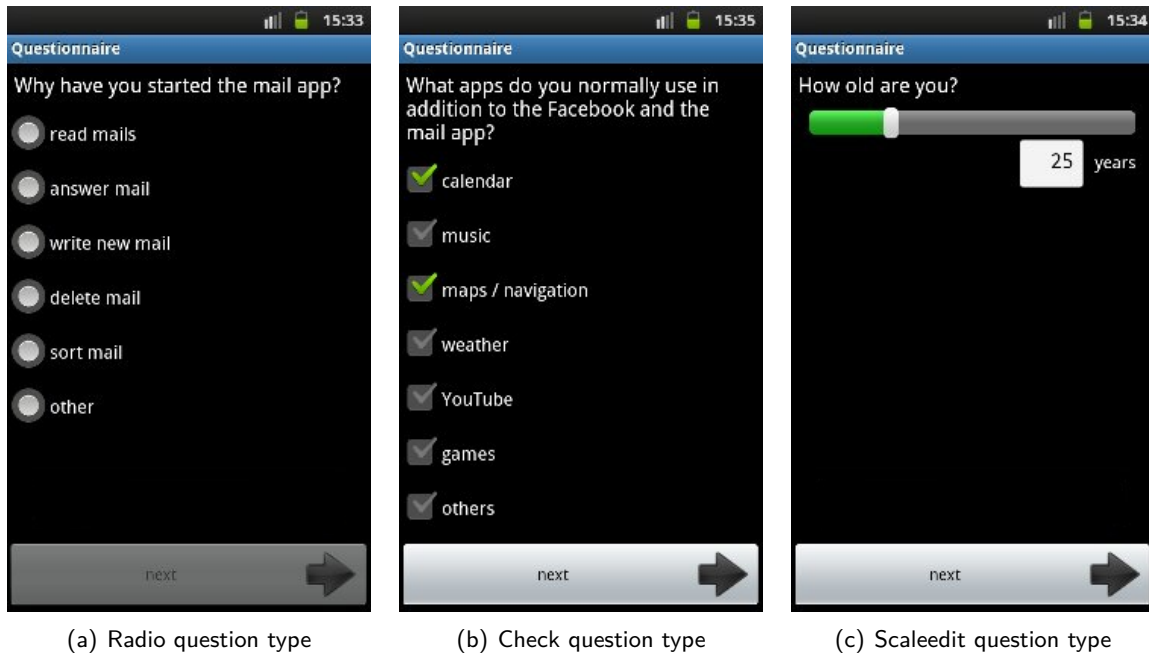

(a) Radio question type   (b) Check question type   (c) Scaleedit question type

Figure 4.3.: Screenshots of the QuestionnaireAsker. The QuestionnaireAsker presents the questions of the chosen questionnaire one after another.

4.1.4. KeepAliveReceiver, BootCompletedReceiver and EditPreferences

The KeepAliveReceiver is started by the QuestionnaireService as soon as the app is installed and started for the first time. Its purpose is to check whether QuestionnaireService is running or not. In case QuestionnaireService has stopped, the KeepAliveReceiver starts it again. In addition, Android sometimes kills services to free resources.

The BootCompletedReceiver also starts the QuestionnaireService. However, the BootCompleted-Receiver is responsible for starting the QuestionnaireService after the smartphone is rebooted for some reason.

The EditPreferences activity can be opened from the QuestionnaireChooser. On the one hand, the options menu enables the user to upload all questionnaires and logs which have not been uploaded yet by pushing the "Upload Answers" button. On the other hand, the user is able to switch to the settings menu (see Figure 4.4(b)). In the settings menu the user can define whether answers should be uploaded automatically or not. Additionally, the smartphone user can choose if only WLAN or also the mobile network is allowed to be used by the app. Finally, some information about the app can be viewed.
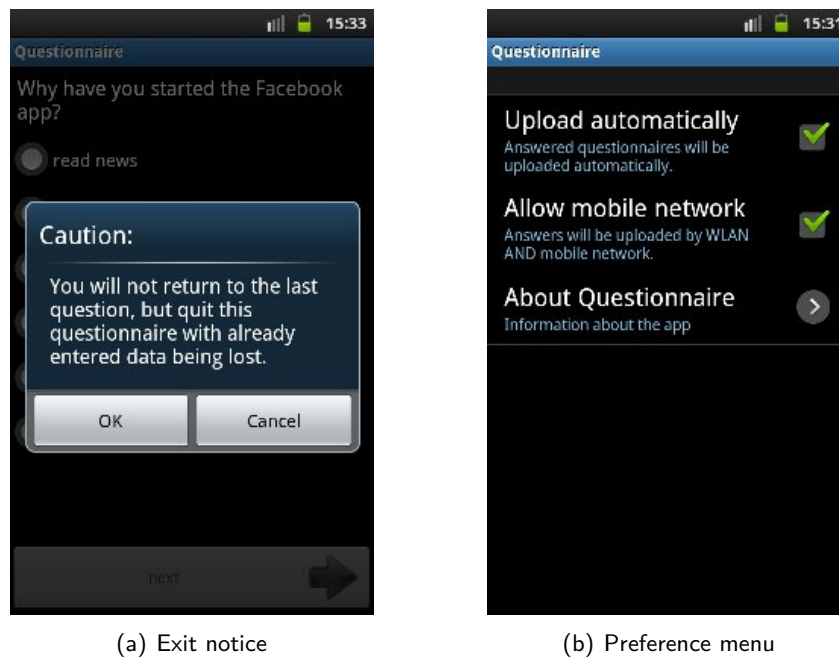
(a) Exit notice   (b) Preference menu

Figure 4.4.: Screenshots of the abort dialog of QuestionnaireAsker and the settings menu of Edit-Preferences.

4.2. Structure of the Backend Server

4.2.1. Backend Server for the Questionnaire App

The Questionnaire app requires a server in order to download new or edited questionnaires and to upload the given answers and saved log data of the participants. This server is implemented with Python Pyramid, an open source web framework. Python Pyramid was chosen because of its many advantages. One advantage of Python is, that it runs on all popular platforms and there should, consequently, be no problems for any conductor of a study to use the server. Another important fact about Python Pyramid is its easy installation. A third advantage is the integration of an SQL database that is used to save the data from the Questionnaire app.

In addition to the serving and saving function, three tools were created to be used with a web browser: QuestionnaireAdmin, DatabaseAdmin and the Evaluation framework. They are written in HTML and JavaScript. Also, Ember.js and jQuery are used to improve the usability. These JavaScript libraries improve the design of and the interaction with the website. The main advantage of Ember.js is that it runs on the client-side and does not need to send HTTP requests to the server in case the user interacts with the HTML page. These HTML pages are further explained in the following.

4.2.2. QuestionnaireAdmin

QuestionnaireAdmin is a tool to create, edit and delete questionnaires in XML format for a study. It is available in two languages: English and German. The German version can opened with a browser under:

http://IP address of server:port number/static/QuestionnaireAdmin/QuestionnaireAdmin.html. The English translation can be found by replacing *"QuestionnaireAdmin"* with *"Questionnaire-Admin_en"* in the URL.

In case the server is running, the upper half of Figure 4.5, named "Available Questionnaires", can be seen. The first column presents all currently available questionnaires on the server. Next to the name, there are the two buttons "Edit" and "Delete". In the last column, information about the questionnaire are shown. Beneath the available questionnaires, the two buttons "Create a questionnaire" and "Upload changes" offer further functionalities. "Edit", "Delete" and "Create a questionnaire" are all performed locally. The local changes can be uploaded to the server by pushing the "Upload changes" button. All questionnaires as well as the changes made are saved on the server as XML files. Therefore, the questionnaires can be read by any text editor.

In case the user pushes the "Edit" button, the lower half of Figure 4.5, named "Questionnaire Editor", appears. It can be closed by pushing the button "Close". The first two lines of the editor can be used to alter the name and the description of the questionnaire. The next line defines the date and time, when the questionnaire should be activated. In case the date is in the future, the questionnaire is stored on the server and the smartphone, but will not appear until the time is reached. Contrary to this, the last line of this group defines the date and time when the questionnaire expires and disappears from the chooser screen. By clicking in the edit text line, a time picker automatically pops up to facilitate the input of date and time and enters the value in the format YYYY-MM-DD HH:MM.

The following line called "Repetition Interval" defines after how many minutes from the start date of the questionnaire a notification should pop up. This notification will only appear in case *interval-triggered* is chosen as the questionnaire mode and the user forgot to answer a form after using the logged app. The line beneath, "Repetition Number", determines how often this questionnaire can be filled out. In case the researcher does not want the questionnaire to be shown only for a limited number of times, he has to enter "0". A form like the start questionnaire, which has to be filled out only once, requires the value "1" in this line. To enter the *event-triggered* mode for a questionnaire, the user has to enter a "1" in the next line called "Automatic Opening". This value triggers to open the respective questionnaire after the smartphone owner closes a logged app and switches to the home screen. In case the value in this line is "0", the questionnaire will not pop up automatically and the questionnaire has to be answered by manually starting the Questionnaire app. The remaining line is called "Groups". The server is able to serve multiple surveys at once. Each survey can have a unique set of questionnaires. The group parameter defines which surveys

Figure 4.5.: QuestionnaireAdmin: Available questionnaires are displayed on top. The Questionnaire Editor shows the parameters of the Facebook questionnaire.

a questionnaire belongs to. It also defines the number of groups and their names. In order to create two or more groups, the user has to enter two or more names for these groups comma-separated from each other. The same group names have to be entered in the parameter section of the Questionnaire app code in order to create different application package files (APKs) for the different groups.

The settings for the Questionnaire Editor end with the definition of the activities and package names that should be logged on the smartphone. As can be seen in Figure 4.5, the Facebook questionnaire logs the Facebook package name "com.facebook.katana". This definition triggers the smartphone to record all usages of the Facebook app. If there are more activities or packages to be logged, they have to be comma-separated. In case multiple apps have to be logged and their names are unknown, regular expressions[1] can be used.

The next section of QuestionnaireAdmin (see Figure 4.6) enables to create and order the questions

---

[1]http://www.regular-expressions.info/

Figure 4.6.: QuestionnaireAdmin: Edit questions section.

for the form. Each question is represented by a light blue box and can be dragged and dropped to a new position. In order to delete a question, the user has to push the topmost button called "Delete question". Next to this button the specification of the question is displayed. As mentioned in section 4.1.3, there are seven different question types. The seven question types only have in common that there is a question or some sort of text. This question has to be entered in the first edit text line of each box, as can be seen in Figure 4.6. Radio is the first question type shown. This question type needs input for a question or request in the first line as well as for a definable list of answers beneath. The answers are represented by dark blue boxes. Just like the questions, the boxes of the answers can be dragged and dropped in order to switch positions. Within these boxes, there is an input line for the text on the right and a button on the left that enables the user to delete an answer option. This structure is identical for the check, likert and dropdown question types. The text question type requires only some sort of text as input and offers no options for answers or further text. The other box in Figure 4.6 is of the question type scaleedit. This question type requests inputs for the start value of the scale bar slider as well as the maximum value and the unit of the scale bar. It should be noted that the numerical inputs

should be integers. The remaining question type, edittext, requests for a question or a text and a prompt to give the participants further instructions below the first text. Furthermore, an integer value defines the number of lines which shall be displayed to the user to type in his text.

In order to add a new question to the questionnaire, the user has to push the according button (see Figure 4.7) for the desired question type. For further information on the different question types, refer to section 4.1.3.
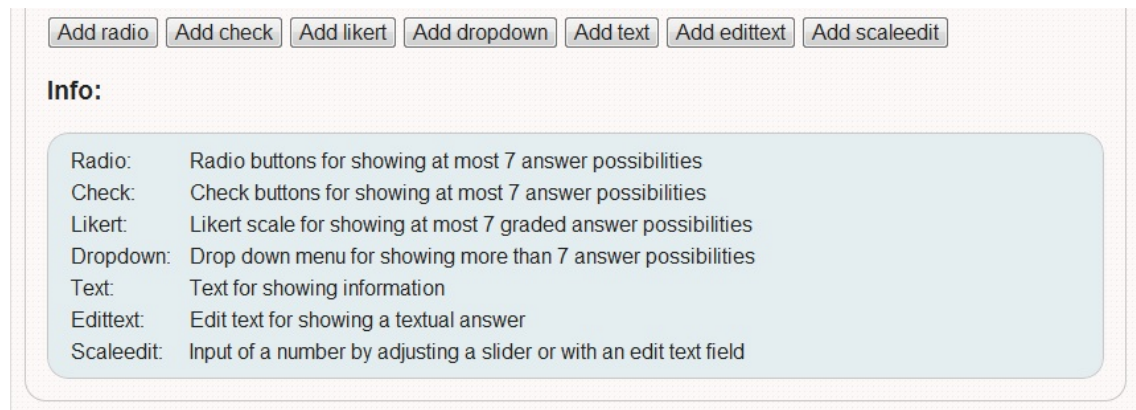
Figure 4.7.: QuestionnaireAdmin: Buttons to add a new question and information about the question types.

The server saves the questionnaires as XML files. A custom XML schema for the storage of questions and possible answers was written for that purpose. New questionnaires can also be created with a text editor by starting from scratch or using an already existing questionnaire as a template. When exporting the Questionnaire app source code, the server automatically adds all questionnaires into the app's resource folder.

4.2.3. DatabaseAdmin

After the questionnaires are successfully created and the study has started, the answered questionnaires and the logged data of all users can be viewed with DatabaseAdmin. Similar to the QuestionnaireAdmin, it is available in two languages and the German version can be opened with a browser under: *http://IP address of server:port number/static/DatabaseAdmin/DatabaseAdmin.html*. The English version can be viewed by replacing *"DatabaseAdmin"* with *"DatabaseAdmin_en"*. Figure 4.8 shows that DatabaseAdmin is divided into two sections: answers and logs. While the section *answers* lists all answered questionnaires of all users, the section *logs* displays the logged data of app usages. As described in Section 4.2.2, the activities and package names that shall be logged are entered for each questionnaire. The answers and the logs section both contain an input field to search for any sort of text in the data. Also, both sections offer information about the number of shown and available entries, respectively.

Answers:

Search: _____

Row	UserId	Time	Questionnaire	Answers
1	e2pct193btd52n76e4qh71vonm	2012-07-14 15:55:02	SystemInfo	{"nfc": false, "version": "2.3.3", "screen": "480x800", "camera": true, "device": "GT-I9001", "model": "GT-I9001 (GT-I9001)", "apiLevel": "10", "gps": true}
4	e2pct193btd52n76e4qh71vonm	2012-07-14 16:00:46	Anfangs - Fragebogen	["weiblich", "47", "Schüler", "Diplom / Master", "6", "6", "einmal in drei Tagen", "einmal in drei Tagen", "meistens UMTS", "[\"Kalender\"]", "5"]
14	e2pct193btd52n76e4qh71vonm	2012-07-15 01:36:47	Mail - Fragebogen	["25", "auf Mail antworten", "[\"Mail lesen\"]", "0"]
15	e2pct193btd52n76e4qh71vonm	2012-07-15 01:37:30	Facebook - Fragebogen	["27", "Neuigkeiten lesen", "[\"nichts weiteres\"]", "2"]

Showing 1 to 4 of 39 entries

Logs:

Search: _____

Row	UserId	Time	Duration	Activity	Packagename	To Packagename	Location
2	tj3008jgmgnhp9fs4ed34ivg1h	2012-07-14 15:52:14	0	com.questionnaire.Logger	com.questionnaire		0.0,11.5734372,48.1657606
3	tj3008jgmgnhp9fs4ed34ivg1h	2012-07-14 15:52:14	293331	com.questionnaire.QuestionnaireChooser	com.questionnaire	com.htc.launcher	0.0,11.5734163,48.1658086
1	e2pct193btd52n76e4qh71vonm	2012-07-14 15:55:02	0	com.questionnaire.Logger	com.questionnaire		provider is deactivated
5	e2pct193btd52n76e4qh71vonm	2012-07-14 15:55:02	235666	com.questionnaire.QuestionnaireChooser	com.questionnaire	com.android.vending	provider is deactivated

Showing 1 to 4 of 167 entries

Figure 4.8.: Screenshot of the DatabaseAdmin that shows the answers and logs of the participants.

The table for the answered questionnaires has columns for the number of the entry, the user ID, the timestamp and the name of the answered questionnaire as well as one column for the answers given by the user. Each answer of a questionnaire is enclosed in quotation marks and comma-separated from the next one. In case the question type was check, the given answers are additionally enclosed in square brackets to detect that these answers all belong to the same question. The first entry in Figure 4.8 presents an example for the system info. The system info is sent to the server the first time the smartphone installs the Questionnaire app and retrieves some information about the smartphone, as described in Section 4.1.1 and 4.3.2. The other entries represent answered forms for three different questionnaires.

The data table of the logged data also contains information about the number of the entry, the user ID and the timestamp of the logged data. In comparison to the answers data table, the logs data table has supplementary entries for the duration of the app activity, the name of the logged activity and the package name. Furthermore, it offers information about the package name to which the user switched after using a logged app. The last column holds the current location of the user. It displays the GPS altitude, latitude and longitude coordinates of the user in case she has not disabled the GPS functionality. If the GPS functionality is disabled, the entry in the data cell is "provider is deactivated". The duration of the app usage is logged in milliseconds and indicates how long an activity has been visible on the screen. The first and the third entry in the log data table in Figure 4.8 are special entries. These entries illustrate the first log entry after the installation of the Questionnaire app or that the phone was switched on again at that moment of time. Specific to this sort of entry, the duration is zero milliseconds and the activity is called "com.questionnaireLogger".

Overall, DatabaseAdmin presents all answers and logs in a clearly arranged way. In case there are

many participants of a study and the data tables become very large, there are two possibilities to get a better overview for each test subject. One way is to use the search input field to filter by the user ID so that only the data of the specific participant will be shown. The other way is to use the sortable column function. By clicking the field "UserId" in the first row, all entries in this data table will be shown sorted according to the user IDs. Also, all other columns can be sorted accordingly to their entries. This can be very helpful if the researcher wants to analyze, for example, just the entries of a specific day or a specific questionnaire.

4.2.4. Evaluation framework

The last tool that was built especially for the planned study was the Evaluation framework. It offers further analyzing functionality besides the DatabaseAdmin. The Evaluation framework code can be seen as a starting point and as an example for other study conductors, who want to adjust the Evaluation tool for their own purposes.

The structure of the Evaluation framework is divided into presorting tasks and the presentation of the collected answers and logs. The backend server first creates a list of all user IDs, which took part in the study. Then, a timeline of answers and logs is created by adding the entries accordingly to their timestamp. After all entries are in the timeline, multiple merging steps are performed. First, in case the string of "ToPackagename" in a row is the same as the string of "Packagename" in the next row, these rows will be merged and the single values for the duration of the activities will be summed up. For example, the Facebook app has a login activity and then switches the activity to show the news. These two activities belong to the same app usage and can be merged. Second, two or more sequenced entries will be merged if they have the same package name and the interval between the end time of the first one and the start time of the second one is less or equal than 60 seconds. This merging is done in order to treat short interruptions adequately. Short interruptions might, for example, be caused by clicking the back button too often or choosing a picture for a new post in the Facebook app. Finally, the timeline is ready and can be viewed for each user ID.

As a next step, a mapping between the user ID and the questionnaire mode is done. The mapping enables to display the start and end questionnaire entries divided into the three groups. The last step of the backend server in Python is to create the system info, the start questionnaire and the end questionnaire data tables.

Similar to the DatabaseAdmin, the prepared data is displayed using Ember.js and jQuery in a HTML page. Each answer of a questionnaire is displayed in one column. Therefore, if other questionnaires should be presented with the Evaluation tool, the files *"views.py"*, *"Evaluation.js"* and *"Evaluation.html"* have to be changed according to the requirements. The created Evaluation tool can then be viewed under:

http://IP address of server:port number/static/Evaluation/Evaluation.html.

As shown in Figure 4.9, the system info will be displayed first. Then, the start and the end questionnaires are presented. They are sorted by the three groups event, interval and voluntary. Between the data of each group, there are a few empty lines for separation purposes. At the bottom of the start and end questionnaire data tables, there are extra lines to calculate, for example, the overall average. Figure 4.9 does not include them due to the limited space.

SystemInfos:

UserId	Zeit	Modus	Geraet	Version	Screen	Kamera	NFC
29oeib4do0eautp2b7a28qto1j	2012-07-23 10:28:10	EVENTTRIGGERED	GT-I9001 (GT-I9001) (GT-I9001)	2.3.3 (10)	480x800	true	false
t5lpt1tql8s2p3q962voug4ntm	2012-07-23 10:30:57	INTERVALTRIGGERED	GT-I9001 (GT-I9001) (GT-I9001)	2.3.3 (10)	480x800	true	false
s6573qav9j6ajaoi05881017bo	2012-07-23 10:33:28	VOLUNTARY	GT-I9001 (GT-I9001) (GT-I9001)	2.3.3 (10)	480x800	true	false
9t7u960le00u9o9aqmvrnfo70v	2012-07-23 11:59:58	EVENTTRIGGERED	GT-I9100 (GT-I9100) (GT-I9100)	4.0.4 (15)	480x800	true	false
ei9ifdt9um3s7jqs6rugk2ussd	2012-07-23 12:32:50	VOLUNTARY	GT-I9001 (GT-I9001) (GT-I9001)	2.3.3 (10)	480x800	true	false

Start Questionnaire:

UserId	Zeit	Sex	Age	Occupation	Degree	Agephone	Agefbapp	durationfb	unknownfb	durationem	
9t7u960le00u9o9aqmvrnfo70v	2012-07-23 15:24:40	männlich	21	Student	Abitur	13	13	mehrmals täglich	Stimme teilweise nicht zu.	mehrmals täglich	Sti
hfsdiflctjdeei5uvdngnfrob6	2012-07-23 13:00:19	männlich	26	wissenschaftlicher Mitarbeiter	Diplom / Master	21	21	einmal pro Tag	Stimme teilweise zu.	mehrmals täglich	Sti te
6sc2n5f2mbjkirba7j6gqrumnd	2012-07-23 13:37:21	männlich	26	wissenschaftlicher Mitarbeiter	Diplom / Master	22	22	einmal pro Tag	Stimme teilweise zu.	mehrmals täglich	Sti te
4namdbuc1ulv82dllme9ggui3a	2012-08-13 00:02:54	männlich	21	Student	Bachelor	16	16	mehrmals täglich	Stimme teilweise zu.	mehrmals täglich	Sti
blank line											
		0.07692307692307693	30.23076923076923								
blank line											
		0.023809523809523808	28.476190476190474								

End Questionnaire:

UserId	Zeit	Effortfb	Performancefb	Effortem	Performanceem	Performanceweek	UlisNice	UlisInt
8brbdt14k6aocevif2f0m9vf4q	2012-09-04 19:26:56	5	4	4	4	["in der dritten Woche"]	3	5

Figure 4.9.: Screenshot of the Evaluation tool showing the system information as well as the start and end questionnaire answers.

Figure 4.10 shows the part below the end questionnaire. All user IDs are listed next to the button "Show". By pushing this button, the timelines for the specific user will be shown. First, the entire timeline with all logs and answers is displayed. Thereby, the Facebook app logs are highlighted in blue and the mail app logs in green. The logs of the Questionnaire app and the answered questionnaires are left white. This color coding facilitates a quick overview of how regular questionnaires have been answered after an app usage. For further analyses, there are two extra data tables. The first timeline presents the Facebook app logs and the answered Facebook questionnaires. The second displays the mail app logs and the answered mail questionnaires. These two data tables can be used to get an overview of the number and the sequence of logs and answers and can be easily copied into an Excel sheet for further calculations.

Users:

UserId	Anzeigen
3qofs6td8qfh4pvncc5tn2t8ho	Show
hv86vhpeuohtbdeq3n00k5p0d	Show
8m0180a79dn12ik29q14sd0lc7	Show
2vptqem48t5qt3pqah2l3anocf	Show
br1p5fppgfkvald18i0d0pct6k	Show
l4usokdd0hslg8rt2n9qsgfiod	Show

Timeline:

Zeile	Zeit	Fragebogen	Antworten	Dauer	Activity	Packagename	To Packagename	Location
0	2012-07-24 21:19:19	SystemInfo	{"nfc": true, "version": "4.1.1", "screen": "720x1184", "camera": true, "mode": "INTERVALTRIGGERED", "device": "maguro", "model": "Galaxy Nexus (yakju)", "apiLevel": "16", "gps": true}					
1	2012-07-24 21:19:20			0	com.questionnaire.Logger	com.questionnaire		0.0,11.5673674,48.1489029
2	2012-07-24 21:19:20			279	com.questionnaire.QuestionnaireChooser, (merge) com.questionnaire.QuestionnaireChooser, (merge) com.questionnaire.QuestionnaireChooser, com.questionnaire.QuestionnaireActivity	com.questionnaire	com.android.launcher, (merge) com.android.launcher, (merge) com.questionnaire	0.0,11.5691753,48.1657662
3	2012-07-24 21:23:53	Anfangs - Fragebogen	["männlich", "31", "Angestellter", "Diplom / Master", "60", "4", "einmal in drei Tagen", "Stimme teilweise zu.", "mehrmals täglich", "Stimme teilweise zu.", " [\"Kalender\",\"Musik\",\"Maps \\ Routenplaner\",\"Wetter\",\"Spiele\",\"sonstige\"]", "7", "meistens WLAN", "fiftycents"]					
4	2012-07-24 21:24:00			3	com.questionnaire.QuestionnaireChooser	com.questionnaire	com.android.launcher	0.0,11.5732381,48.1657681
5	2012-07-24 21:24:04			15	com.facebook.katana.activity.FbFragmentChromeActivity	com.facebook.katana	com.google.android.email	0.0,11.5732381,48.1657681
6	2012-07-24 21:24:19			8	com.android.email.activity.Welcome, com.android.email.activity.EmailActivity	com.google.android.email	com.android.launcher	0.0,11.5732381,48.1657681
7	2012-07-24 21:24:29			38	com.questionnaire.QuestionnaireChooser, com.questionnaire.QuestionnaireActivity	com.questionnaire	com.questionnaire	0.0,11.5731117,48.1657359

Timeline FB:

Zeile	Zeit	Fragebogen	Antworten	Dauer	Activity	Packagename	To Packagename	Location	Answer 1	Answer 2	An
0	2012-07-24 21:24:04			15	com.facebook.katana.activity.FbFragmentChromeActivity	com.facebook.katana	com.google.android.email	0.0,11.5732381,48.1657681			
1	2012-07-24 21:25:00	Facebook - Fragebogen	["Neuigkeiten lesen", " [\"nichts weiteres\"]", "1", "0"]						Neuigkeiten lesen	["nichts weiteres"]	1
2	2012-07-25 12:46:36			8	com.facebook.katana.activity.FbFragmentChromeActivity	com.facebook.katana	com.android.launcher	0.0,11.5675995,48.1489312			
3	2012-07-25 12:47:17	Facebook - Fragebogen	["Neuigkeiten lesen", " [\"nichts weiteres\"]", "1", "0"]						Neuigkeiten lesen	["nichts weiteres"]	1

Timeline EM:

Zeile	Zeit	Fragebogen	Antworten	Dauer	Activity	Packagename	To Packagename	Location	Answer 1	Answer 2
0	2012-07-24 21:24:19			8	com.android.email.activity.Welcome, com.android.email.activity.EmailActivity	com.google.android.email	com.android.launcher	0.0,11.5732381,48.1657681		
1	2012-07-24 21:25:40	Mail - Fragebogen	["Mail lesen", " [\"nichts weiteres\"]", "1", "0"]						Mail lesen	["nichts weiteres"]
2	2012-07-25 10:39:52			32	com.android.email.activity.EmailActivity	com.google.android.email	com.android.launcher	0.0,11.5732789,48.1657455		
3	2012-07-25 10:40:41	Mail - Fragebogen	["Mail lesen", "[\"Mail lesen\"]", "2", "0"]						Mail lesen	["Mail lesen"]

Figure 4.10.: Screenshot of the Evaluation tool showing all user IDs and the three timelines. The first timeline shows all answered questionnaires and all logged app usages. The second and the third timeline present only the data related to the Facebook and the mail questionnaire, respectively.

4.3. Communication of the Components

The previous sections of this chapter explained every component of the Questionnaire app. In order to better understand the communication of these components, they will be further illustrated in the following.

4.3.1. The JavaScipt Object Notation

The data format for the communication between QuestionnaireService and the backend server is JavaScript Object Notation (JSON)[2]. There are two basic data structures in JSON: an array and an object. The array is an ordered collection of values. A JSON array notation begins with a left square bracket and ends with a right square bracket. The single values are separated by a comma. The object is an unordered set of name/value pairs. In other languages it is sometimes called dictionary or struct. In JSON notation an object begins with a left curly bracket and ends with a right curly bracket. After each name a colon follows and commas separate the name/value pairs. A value, which is used in an array and an object, can be a string in double quotes, a number, an object, an array, true, false or null.

4.3.2. QuestionnaireService and the Questionnaire Backend Server

As already mentioned in the previous section, the data format used for the communication between the backend server and QuestionnaireService is JSON. Figure 4.11 illustrates how the QuestionnaireService downloads the JSON array of all available questionnaires with an HTTP GET request from the server.

In addition to the download of questionnaires, there are also two different uploads to the server. Common to both of them is that they are performed with an HTTP POST request. The first one is the upload of answers. Answers are transformed into a JSON array that consists of multiple JSON objects, built by the function *"uploadQuestionnaires()"* of the QuestionnaireService. Figure 4.11 presents the structure of such a JSON object. The JSON object includes the strings *userId*, *questionnaire*, the long variable *timestamp* and another JSON object named *answers*. The string *userId* contains the user ID of the test subject and *questionnaire* represents the name of the answered questionnaire. The *timestamp* variable is the point in time measured in milliseconds from epoch, when the questionnaire has been answered. The last element, the JSON object *answers*, contains the answers of a questionnaire. At the first start of the app, the first answer object is generated with information about the smartphone. Thereby, the answer JSON object

---

[2]http://www.json.org/
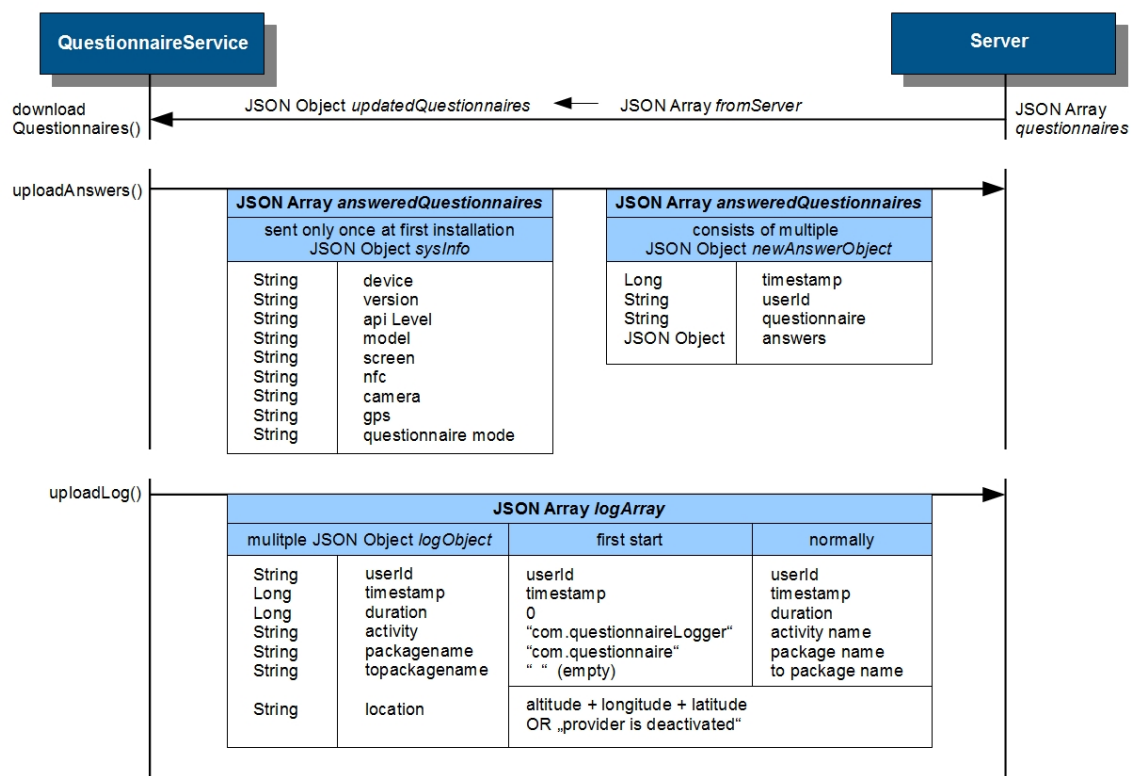
Figure 4.11.: Interaction between the QuestionnaireService and the server. The backend server sends all available questionnaires to QuestionnaireService and QuestionnaireService uploads the answers and logs to the server.

is built up of strings containing the device, the version, the API level, the model and the screen height and width. In case the smartphone has the features camera, NFC and GPS, they are supplementary specified. Additionally, the questionnaire mode (voluntary, interval-triggered or event-triggered) is sent as a string in the JSON object. This information is required for the conductor of the study to be able to identify which user ID belongs to which questionnaire mode. For regular questionnaires, the JSON object *answers* is used for the answers that are values of the format string, integer and JSON object depending on the question types in the questionnaire.

The other JSON array which is uploaded to the server is also built out of multiple JSON objects by the *LoggerRunnable* thread of QuestionnaireService. These JSON objects contain the logged data that records the current app name every second. Each JSON object contains the long variables *timestamp* and *duration* and the strings *userId*, *activity*, *packagename*, *topackagename* and *location*. Thereby, *timestamp* and *duration* indicate at what point in time and for how long the logged activity has been used. These values are given as long variables in the unit milliseconds. The strings *activity* and *packagename* contain the information which activity from which package

is currently visible on the screen. Thereby, only specified activities and packages from currently active questionnaires are logged. It should be noted that it is more reasonable to log the package name of an app. There is only one package name for an app, but multiple activity names for each possible screen. The next element *topackagename* is useful to get a better understanding about what happened after the usage of the logged app. For example, the user could have sent the smartphone to stand-by mode which would be indicated by the string "sleeping". Especially for analyzing the event-triggered mode, the conductor of the study has to know if the home screen was chosen after the app usage which would have triggered an automatic questionnaire. If no questionnaire was answered, the conductor of the study is able to see that the participant canceled the questionnaire. However, if the next app after a logged app was not the home screen, the participant was not prompted to fill out a questionnaire and therefore did not miss it. The last element named *location* is represented with the string "provider is deactivated" in case the smartphone has no GPS feature or the participant deactivated it. If the GPS feature is available and activated, the string specifies the altitude, longitude and latitude of the smartphone location at the time of the app usage.

4.3.3. QuestionnaireService, QuestionnaireChooser, QuestionnaireAsker and EditPreferences

Figure 4.12 illustrates the communication of the QuestionnaireService, QuestionnaireChooser and QuestionnaireAsker. The communication starts with a message from QuestionnaireChooser to QuestionnaireService including the string *("ACTION", "getQuestionnaire")*. The Incoming Handler of QuestionnaireService then triggers the function *"createChooserData()"* to start. This function has three tasks. First, it checks whether the information screen of the Questionnaire-Chooser has already been shown. In case it has not been shown yet, QuestionnaireService submits the package names of the monitored packages as an array to QuestionnaireChooser. The second task is to test if any of these monitored apps are missing. If there are missing apps, it sends the package names as a JSON array in a bundle to the QuestionnaireChooser. The third task is the most commonly used one, the forwarding of the JSON object of the current questionnaires. QuestionnaireChooser extracts these JSON objects and displays all questionnaires in the main menu (see Figure 4.2(a)).

A participant can choose one of these questionnaires by pushing the respective button. After pushing the button, QuestionnaireChooser starts an Intent to launch the activity QuestionnaireAsker. QuestionnaireAsker then shows the first question of the chosen questionnaire. The information which questionnaire was chosen, is sent as an extra together with the Intent. The Intent therefore gets the strings *title* and *xml* which represent the title of the questionnaire and the XML description of this questionnaire.

As further explained in Section 4.1.3, QuestionnaireAsker skips through all questions of one ques-
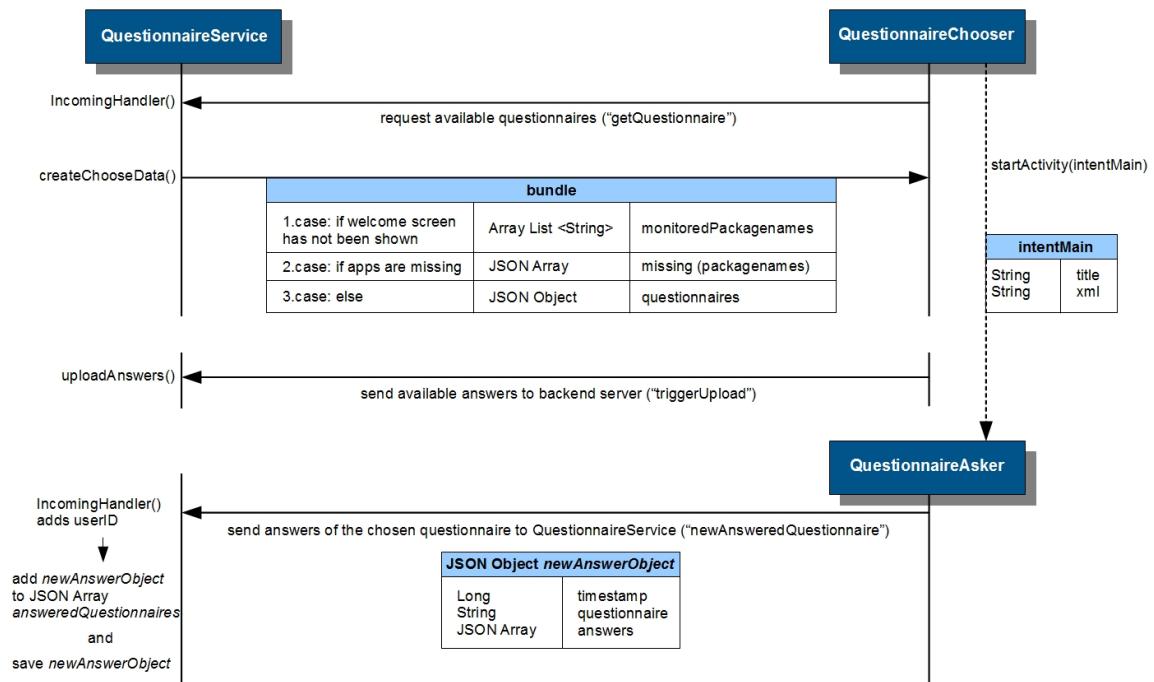
Figure 4.12.: Communication between the QuestionnaireService, QuestionnaireChooser and QuestionnaireAsker. QuestionnaireService sends all available forms on request to QuestionnaireChooser. QuestionnaireChooser can start QuestionnaireAsker, which sends the answers of a completed questionnaire to QuestionnaireService. QuestionnaireChooser can trigger QuestionnaireService to upload cached answers to the backend server.

tionnaire and saves the answers given by the participant. It then combines the answers as JSON array together with the title of the questionnaire and the long variable named *timestamp* into a JSON object. This JSON object is sent back as a bundle with the string *("ACTION", "newAnsweredQuestionnaire")* to QuestionnaireService. QuestionnaireService's incoming handler receives the message and adds the string *userId* to the object. The incoming handler then integrates the new answered questionnaire to its JSON object of answers that can be sent to the server.

QuestionnaireChooser has some additional tasks. If there were already answers given, the participant is able to upload these by pushing the "Upload answers" button in the options menu. This will trigger QuestionnaireChooser to send the string *("ACTION", "triggerUpload")* to QuestionnaireService. QuestionnaireService then uploads all data which has not been sent yet.

If the user pushes the *"Settings"* button in the options menu, QuestionnaireChooser will start the activity EditPreferences. The EditPreference activity is responsible for the layout of the settings menu. The settings menu can be left by pushing the back button of the smartphone. Then, the QuestionnaireChooser screen is shown again.

4.3.4. QuestionnaireService, BootCompletedReceiver and KeepAliveReceiver

The interactions between the QuestionnaireService, the BootCompletedReceiver and the Keep-AliveReceiver are necessary to ensure that QuestionnaireService is running.

Figure 4.13 shows that the BootCompletedReceiver starts QuestionnaireService in case the smartphone is rebooted. QuestionnaireService, on the contrary, starts the KeepAliveReceiver shortly after the installation of the Questionnaire app. In return, the KeepAliveReceiver checks every 60 seconds whether QuestionnaireService is running. In case it is not, the KeepAliveReceiver starts the service. This mechanism ensures that the logger is always running.
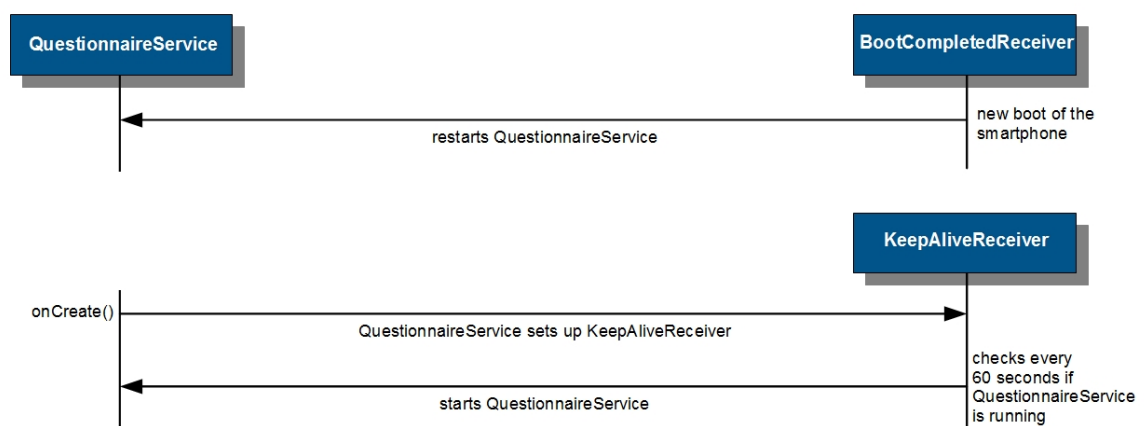
Figure 4.13.: Communication between the QuestionnaireService, BootCompletedReceiver and KeepAliveReceiver. The BootCompletedReceiver restarts QuestionnaireService after a reboot. The KeepAliveReceiver ensures that the QuestionnaireService keeps running.

4.3.5. Communication inside QuestionnaireService

QuestionnaireService has different functions that all access the same information, sometimes from different threads. In order to keep all information updated and available to every function of QuestionnaireService, it utilizes Android's *SharedPreferences*. Data values stored in the Shared-Preferences are persistent and are not lost even though the app might be stopped and started again. Table 4.2 lists all variables that are stored in the SharedPreferences together with their type and a short description of their functionality.

Table 4.2.: The data stored in the SharedPrefereces.

variable name	type	description
answeredQuestionnaires	JSON array	answered questionnaires that have not yet been sent to the backend server
logArray	JSON array	logged data that have not yet been sent to the backend server
questionnairesCache	JSON object	cache for questionnaires from the backend server
updatedQuestionnaireCache	JSON object	cache for questionnaires that are not shown yet, as their activation date lies in the future
userId	string	user ID for smartphone
answerCount	JSON object	saves the number of times how often a questionnaire has been answered
lastUsed	JSON object	variable to track the last usage of each app
lastUpdate	long	date when the questionnaires were downloaded last from the server
welcomeScreenShown	boolean	true if first information screen on QuestionnaireChooser has been shown
pref_auto_upload	boolean	true if "Upload automatically" check box in the settings menu is checked
pref_network_mode_title	boolean	true if "Allow mobile network" check box in the settings menu is checked

4.4. Steps to Run a Study with the Questionnaire App

So far, the Questionnaire app, the corresponding backend server, QuestionnaireAdmin, DatabaseAdmin and the Evaluation framework were introduced and their main functionalities were illustrated in brief. These tools allow to prepare and conduct a study and analyze the results. The required steps to do so (see Figure 4.14) are summarized in the following.

First, the conductor has to be sure about the goal of the study and its relevant aspects to be examined. The Questionnaire app fits for a field study that asks the participant to answer questionnaires or write a diary. The questions or tasks, which have to be accomplished, have to be planned by considering the goal of the study. Then, all questions have to be structured in a reasonable and thoughtful way as a questionnaire [4, 6]. Afterwards, the conductor has to decide whether the participants should answer the questions voluntarily, whether there should be an additional reminder or if the questionnaire should open automatically after the usage of a specific app. Moreover, the number of participants and groups has to be chosen depending on how many are needed to prove the hypotheses. Finally, after acquiring the participants, they have to be

distributed to the groups in a reasonable way.

With this preliminary considerations being done, the conductor of the study is able to start preparing the technical side. Figure 4.14 shows the steps for the setup highlighted in blue. First, he has to install the server. In case there is no Python version 2.6 on the server, it needs to be downloaded and installed[3]. The next step is to install Pyramid with the installation guide[4]. Thus, the framework of the server is ready and the QServer file can be installed. For installation, the zip-File *qserve-1.0* has to be copied into the *virtualenv*-folder and has to be unpacked there. Then, switch into the virtualenv-folder by entering *"cd virtualenv"*. Finally, start the server with the command *"PATH/pserve qserve-1.0/production.ini &"* to run it in the background.

After that the conductor is able to use the first HTML page, QuestionnaireAdmin. QuestionnaireAdmin enables to create all prepared questionnaires and diary study forms together with the chosen questionnaire mode and the corresponding parameters. After creating all questionnaires, QuestionnaireAdmin uploads the data to the server. Next, the AppCreator tool has to be opened with a browser under: *http://IP address of server:port number/static/AppCreator/AppCreator.html*. There, the parameters *murTime*, *pdTime*, *slTime*, *groupId* and *mode* need to be defined as described in section 4.1.1. The chosen values for the parameters as well as the correct *serverIp* are automatically inserted into the Questionnaire app code. Afterwards, the created Questionnaire.zip needs to be downloaded from the server and imported into Eclipse. In Eclipse the project has to be opened and the APKs have to be created by compiling the project without any modifications. In order to create more APK versions for one or more groups, the parameters can be changed with the AppCreator or in the parameter section of the QuestionnaireService.java file.

All APKs then have to be sent via email to the participants. The participants are able to install the APK by opening the mail and pushing the "Install" button. To guarantee the immediate logging of the smartphone usage, it is recommended to ask the participants to briefly open the Questionnaire app after installing. Thereby, all technical preparations are done and the study can start. It is recommended to send the APKs at least a day before the study starts. This should guarantee that all participants have enough time to install the app and the conductor is able to fix potential problems with the installation that could occur due to different smartphone models and apps.

During the study, all given answers and the logged data will be uploaded to the backend server automatically. The backend server saves the data and enables the conductor to view it with DatabaseAdmin. At the end of the study, the participants are allowed to uninstall the Questionnaire app from their phone in case the backup copy of their data is not needed any more. For an easier analysis of the data, the SQL database can be viewed with an SQL editor. Another possibility is to use the Evaluation framework. The Evaluation framework is designed for the study conducted within this diploma thesis. However, the already existing files can be used as an example for how

---

[3]http://python.org/download/
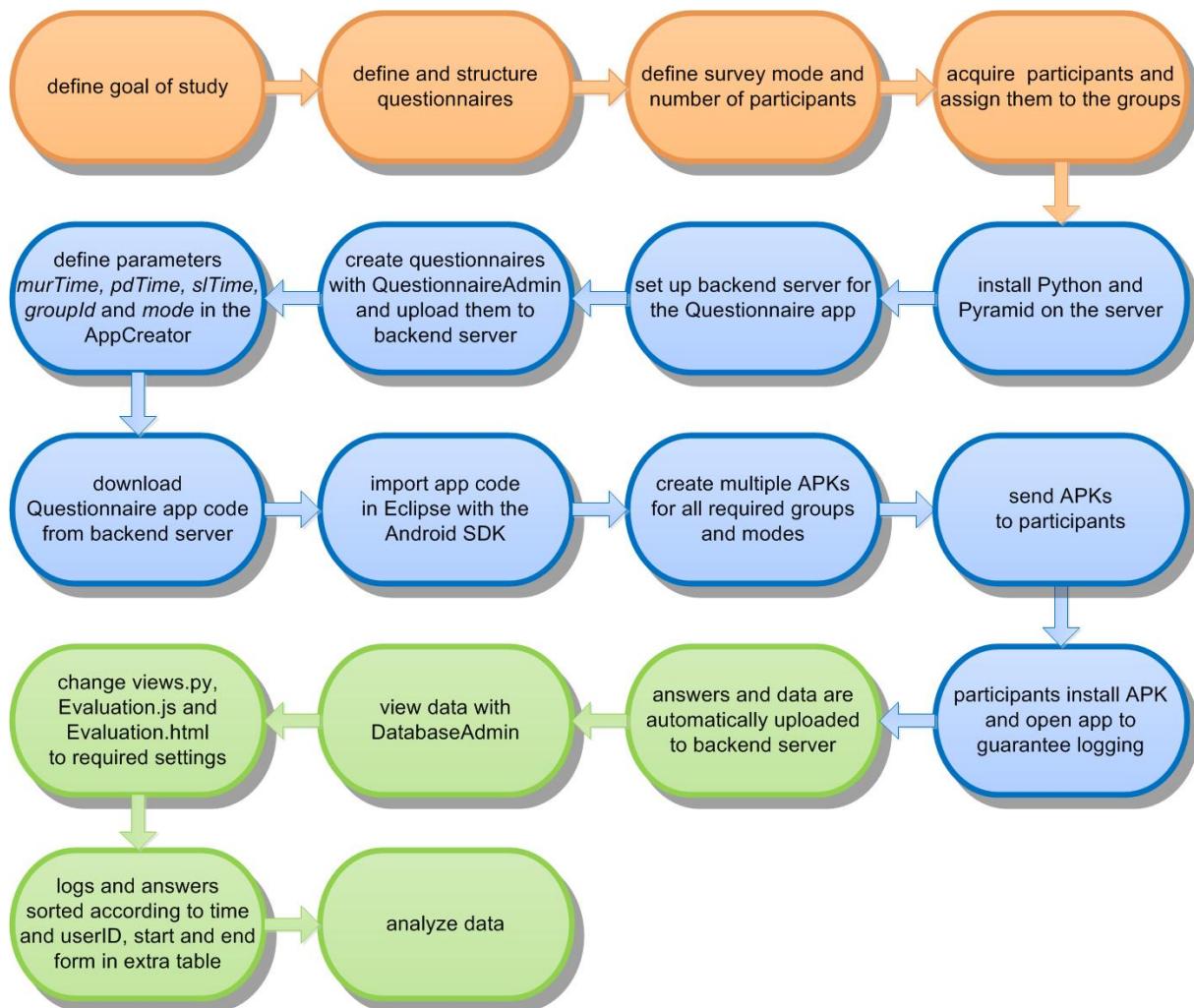[4]http://docs.pylonsproject.org/projects/pyramid/en/1.3-branch/narr/install.html

Figure 4.14.: Schema to create, conduct and analyze a study. The preparations for the study are colored orange, the technical setup blue and the collection and analysis of the data green.

the files "views.py", "Evaluation.js" and "Evaluation.html" have to be changed in order to fulfill the requirements of a new study. After everything is set up, the Evaluation tool presorts the data for an easy analysis. Finally, the data can be analyzed in order to test the hypotheses.

Chapter 5.

Pilot Study

In the previous chapter, the focus lay on the implemented Questionnaire app. This chapter describes the method and the analysis of the pilot study. The pilot study was run after the main part of the implementation was completed. It was conducted in order to test the usability of the app and to guarantee that the interaction of app and backend server works properly.

5.1. Method

Before the pilot study started, some preparations had to be done. First of all, questionnaires were needed. Therefore, the three questionnaires

1. start questionnaire: questions on the demographic data of the participant,

2. Facebook questionnaire: questions on the Facebook app usage and the

3. end questionnaire: questions on the study and the user experience with the app,

with a structure similar to the planned main study, were created. The forms were made to test them for the main study and not to gather answers. Second, the parameters for the app were chosen. At that point in time, the app had four configurable parameters. The parameters with the chosen values were

- the IP address of the server (serverIp),

- the interval for reminding the participant to upload answers (murTime): every 24 hours,

- the interval for trying to download new available questionnaires from the server (pdTime): every 12 hours,

- the interval for saving the logged data (slTime): every hour

(see Section 4.1.1 for further explanation of the variables). These values were chosen in order to minimize the impact on battery life. With the creation of the questionnaires and the choice of the

parameters, the server and the app were ready for the study.

Next, participants were needed. On the one hand, the pilot study needed to have enough users to test the functionality. On the other hand, the amount of data to be analyzed should not be too large. Therefore, the study was conducted with five participants, including myself. These participants got the APK for the installation of the app via email. The email also contained the instructions on how to participate in the study.

The pilot study ran for 72 hours (June 25, 2012, 12 p.m. until June 28, 2012, 12 p.m.). At the beginning of the study, the participants had to fill out the start questionnaire. During the study they should answer the Facebook questionnaire after every usage of the Facebook app. Additionally, every evening at 6 p.m. a notification appeared on the smartphone in order to remind the users to complete a questionnaire. This questionnaire mode is called interval-triggered in the main study. After the end of the study, the participants had six hours to fill out the end questionnaire.

5.2. Results

5.2.1. Results from the Start Questionnaire

The start questionnaire was used to obtain demographic data from the participants. The questions concerned the gender, the age and the occupation. Additionally, the participants were asked how long they already had a smartphone and used the Facebook app. The answers revealed that the four male and the one female participant were on average 25.8 years old. On average, they had a smartphone for 17.4 months and have used the Facebook app for 13 months.

5.2.2. Results from the Facebook Questionnaire

To investigate the usage of the Facebook app and the ability of the participants to self-assess, the Facebook questionnaire asked the following requests or questions.

1. Please estimate how long you last used the Facebook app.

2. Why did you start the Facebook app?

3. What did you do after the first Facebook activity?

4. How often did you use the Facebook app since the last time you filled out a questionnaire without answering a questionnaire?

The answered questionnaires were compared with the logged data. Three of the five log files were complete. One log file had an incomplete data set of the first day. For reasons of data

comparability, the answered questionnaires for the first day were ignored. The log file of the last participant was missing logs for three of seven questionnaire usages and four of seven Facebook usages. Consequently, the last participant was excluded from the analysis of the Facebook questionnaire.

Table 5.1 presents the results of the comparison between the answered questionnaires and the logged data. As can be seen in the first two columns, the first two test subjects (TS 1 and TS 2) answered a questionnaire for every Facebook app usage. The other two participants left out a few. TS 4 declared that he counted six Facebook app uses. However, the logs indicated 16 uses. On average, all four users filled out 5.75 Facebook questionnaires, but used the app 7.25 times. The next two columns focus on the first question, the duration of the app use. As can be seen from all entries and the general average, the participants overestimated their usage. Column E presents the average duration of app use of the not indicated usages. Since TS 1 and TS 2 always filled out the questionnaire, only TS 3 and TS 4 have entries for this column. Their general mean of duration of use is about 0.63 minutes = 38 seconds. Another interesting fact is the interval between the notification and the answering of the questionnaire. With a minimum of 77.00 minutes and a maximum of 232.67 minutes the general average was 136.25 minutes. Thus, the participants filled out the questionnaire about two hours after the notification appeared. The last column shows the average duration for answering the Facebook questionnaire. With a mean of 19.91 seconds the duration for three of the four users was under 20 seconds.

Table 5.1.: Statistics on the Facebook questionnaire; A: number of answered questionnaires, B: number of Facebook app usages, C: self-assessed Facebook app usage on average [min], D: logged Facebook app usage on average [min], E: mean of not specified Facebook app usage duration [min], F: average time until notificated questionnaire was filled out [min], G: average time to answer Facebook questionnaire [s].

	A	B	C [min]	D [min]	E [min]	F [min]	G [s]
TS 1	12	12	2.33	0.88	-	232,67	13,05
TS 2	7	7	2	1.24	-	146.67	19.46
TS 3	2	10	3	1.25	0.73	88.67	27.64
TS 4	2	self-reported:6, logged: 16	2	1.49	1.8	77	19.5
general average	5.75	7.25	2.33	1.21	0.63	136.25	19.91

The answers of the second and third question of the Facebook form, shown in Figure 5.1, reveal a significant trend. When asking the participants why they opened the Facebook app, 94% told that they wanted to read news. From the remaining 6%, one half read an event and the other half posted some news. The next question asked the participants what else they did with the Facebook app. The majority of 79% chose the same answer option as in the first question and thereby indicated that they did nothing else. From the remaining test subjects, 6% read news,

6% wrote a message and 9% did something else. In conclusion, the Facebook app is quite often opened for just one activity and this activity is mainly reading news.
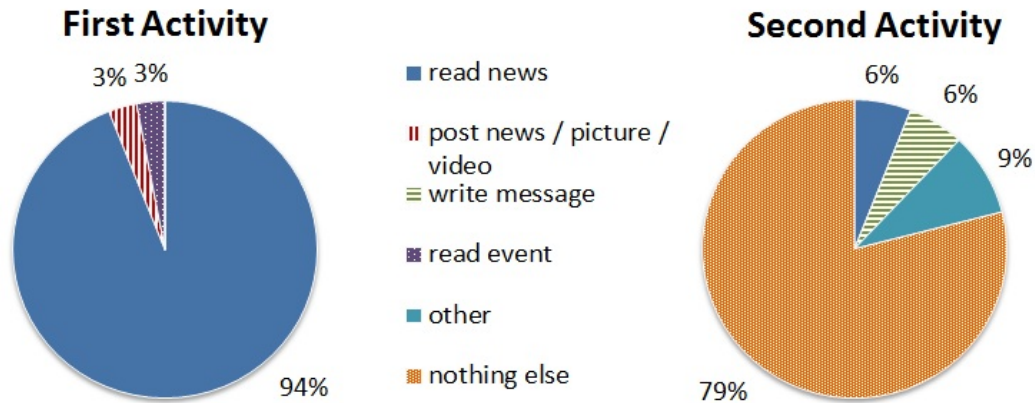

Figure 5.1.: Results of the second and third question of the Facebook questionnaire. The left pie chart represents the reason why the Facebook app was opened. The right pie chart specifies further activities in the Facebook app.

5.2.3. Results from the End Questionnaire

At the end of the pilot study, the participants filled out a questionnaire on the usage of the Questionnaire app. The questions along with the answers scale used in Figure 5.2 were

1. Answering the questionnaire after the Facebook app usage is: very much effort(1) - very little effort(5)

2. I have answered the questionnaire on average (immediately after the Facebook app use): very rarely(1) - always(5)

3. The graphical interface of the Questionnaire app is: very poor(1) - very good(5)

4. The operation of the Questionnaire app is: very poor(1) - very good(5)

5. My battery life with using the Questionnaire app was compared to usually: much worse(1) - much better(5)

6. What did you like or not like about the Questionnaire app? Please give feedback.

Figure 5.2 shows the given answers of four participants as box-and-whisker plots[1]. Due to my participation as the fifth subject, these data were removed to get an objective evaluation. The first column shows that the users perceived answering the Facebook questionnaire on average as little effort. Thus, the questionnaire is not too long and is not annoying to fill out. The next

---

[1]http://en.wikipedia.org/wiki/Box_and_whisker_plot

question asked the participants to self-assess how often they answered the questionnaire directly after the app usage. Two participants declared that they always filled it out directly after the use. The third and the fourth participants answered the questionnaire *most of the time* and *often* after the usage. The appearance and the handling of the app were considered as *okay* and *okay* to *good* respectively. This indicates that the subjects were comfortable with the app, but would appreciate some improvements. The fifth question focused on battery life. The battery life is an important point to consider because a draining battery would be a huge annoyance to users. With regard to the given answers, the use of the app does not have a noticeable influence on the battery life of the smartphone. The last question asked the participant to give feedback on the app. Some interesting feedbacks are discussed in the next section.
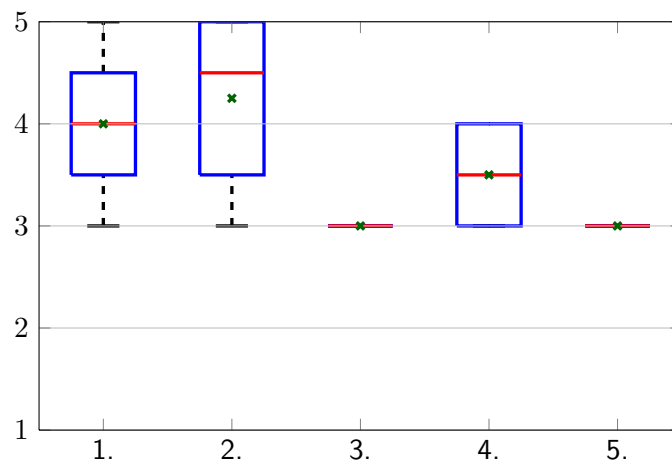
Figure 5.2.: Results of the end-questionnaire of the pilot study. On average, answering was little effort (1.) and the forms have been answered almost directly after app usage (2.). The app's interface was okay (3.), its operation was rather intuitive (4.) and the battery life stayed the same (5.). Definition of symbols:
——: median, **x**: average, **o**: outlier, —— upper: third quartile, —— lower: first quartile, —— upper: maximum value, —— lower: minimum value.

5.2.4. Usability Issues

The feedback that the participants gave in the last question provided hints to improve the appearance and the handling of the app. A question type that could be improved was the edittext type. A soft keyboard is automatically displayed on the smartphone whenever the user clicks on an editable element. However, users would have appreciated it if the soft keyboard would have disappeared automatically as soon as the next question was displayed. Additionally, two subjects would have liked the enter button on the soft keyboard to have a function because they tried to go to the next question by pushing it. The enter button now starts a new line of text. In order to hide the soft keyboard, a "Done" button right next to the editable text field was implemented.

In case the subject forgets to push the Done button, the soft keyboard automatically hides itself whenever the user skips to the next question. Another suggestion regarding the edittext question type was that the area to enter text could be larger in order to see the whole text that has been entered at once. Thereto, the "Number of Lines" field in the QuestionnaireAdmin (see Section 4.2.2) has been added. The supervisor of the study can now define the size of the text field. If she wants or expects a longer answer, she can create a larger text field.

Participants also wished for some improvements on the scaleedit question type. The scale bar should not have the width of the screen, but leave some space on both sides. A little space on both sides could improve the ability to pull the slider near the ends. For this reason, the scale bar has now more space to both screen borders. Another weakness was the appearance of the drop-down list. The old list had a radio button next to every answer. This confused one subject because the radio button did not get marked by pressing this answer. He remarked that it would be better if there are only the answers and no radio buttons. Hence, the new appearance of the dropdown question type is a simple drop-down list.

Finally, there were some general remarks on the Questionnaire app. One user noticed that during answering a questionnaire the app skipped back to the first question whenever he was rotating the smartphone. This bug can be annoying in case the user fills out a long questionnaire and has already answered a lot of questions. The bug was fixed and the app can now be operated even though it is rotated. Another weakness was the notification to fill out a questionnaire. One tester would have appreciated it, if the notification only appeared when he forgot to answer the questionnaire. Therefore, the notification trigger was changed and now tracks whether a form has not been answered after an app usage. The last note was with regard to the Facebook questionnaire. The third question concerning other activities on the Facebook app usage missed the option "nothing else".

5.2.5. Technical Issues

After removing the usability issues, there was one technical issue to be taken care of. As described in Section 5.2.2, some logged data of the participants was missing. An explanation for this bug is that the implementation removed the logged data from the SD card of the smartphone after sending the data to the server. This functionality was implemented in order to not overload the SD card with logger data. In case of a communication failure with the server, the data would not be resent properly in all cases. The app behavior was changed to catch all possible transmission errors and to additionally keep a complete backup copy of the log on the SD card. With these changes, the app was ready for the main study.

Chapter 6.

Main Study

This chapter describes the preparations for the main study, its execution and the data analysis. The questions to be answered are: Which questionnaire mode provides the highest response rates for which application? How high is the response rate after each week? When are the response rates too low to obtain statistically relevant data for a study?

6.1. Study Setup

6.1.1. Methodology

The study was run with 30 participants over the course of six weeks. The length of six weeks was chosen due to several reasons: first, in order to see how the response rates behave over a longer period of time. Second, it was assumed that the response rates decrease strongly after four weeks and a duration of six weeks should reveal this possible decrease. Third, due to the time limitations of a diploma thesis this was the maximum number of weeks the study could run.

At the beginning of the study, the users had to answer a start questionnaire with questions about demographic data and their smartphone experience. Then, a between-subject design was used with each participant experiencing only one condition. A within-subject design, on the contrary, lets the participants test all conditions of a system. Due to the six-week period, this method would have been inappropriate. A bias, which can be introduced in between-subject designs due to different group members, has to be prevented by choosing similar groups. Therefore, the 30 participants were distributed in three similar groups during the study, each of them having a different questionnaire mode. The survey modes of the groups were:

1. voluntary feedback diary and data logging,

2. interval-triggered, experience-sampled diary and data logging,

3. event-triggered, experience-sampled diary and data logging.

As explained in chapter 3, the first group of participants had to self-report. The second also had to self-report, but got a daily notification in case they missed to make an entry. The last group had the forms pop up automatically after an event occurred. All 30 subjects had to answer the same questions after every usage of the Facebook app or one of their mail apps. These questions asked the test subjects about their intention, other actions and the duration of the usage. Additionally, it asked for the number of forgotten diary entries since the last entry. During the study, reminder emails have been sent to the participants at the beginning of week three and week five. These emails included the information, how long the participant already took part in the study and when the study would end. It did not ask the user to improve his response rate, but indirectly showed him that he should still take part.

At the end of the study, a last questionnaire evaluated the app as well as the questionnaire mode. After the study, the given answers and the logged data were analyzed with regard to the aforementioned aspects.

6.1.2. Researched Hypotheses

The aim of this diploma thesis was to set up common guidelines for conducting a study with smartphones. Therefore, the Facebook and the mail app usage are researched in order to gather information for an app that is mainly used in leisure time and one that is used for work purpose and during the whole day. The collected information from the two study techniques - the self-reports and the logs of the app usages - will be analyzed and compared to each other. The analysis will test the hypotheses:

- **H1**: The average duration of an app usage is underestimated by the participants.

- **H2**: The average duration of the mail app is better estimated than of the Facebook app, because it is often shortly used to look up a new incoming mail.

- **H3**: The response rate of the interval-triggered group is higher than of the voluntary group.

- **H4**: The response rate of the event-triggered group is higher than of the interval-triggered group.

- **H5**: The response rates for the Facebook questionnaire are higher than those for the mail questionnaire, because the participants rather fill out questionnaires during their spare time.

- **H6**: The weekly response rates decrease over the six weeks.

- **H7**: The number of app usages decreases over the period of six weeks due to the additional effort for answering questionnaires.

- **H8**: The participants underestimate the number of app usages for both apps.

- **H9**: The event-triggered mode is the most burdensome for the participants.

6.1.3. Technical Setup

Before the study could start, the backend server was set up and the questionnaires were created. The start questionnaire was shown only until the participant answered it and was not shown again afterwards. The Facebook and the mail questionnaire could be answered as often as desired. The interval to trigger a notification of the second group was set to one day. Thereby, the mail questionnaire notification was triggered at 9:30 a.m., the notification for the Facebook questionnaire at 7:00 p.m.. The end form appeared at the last day of the study and could be filled out only once. All questionnaires were saved on the server and added to the assets-folder of the Questionnaire app code. All three groups had common values for the following parameters:

- the IP address of the server (serverIp),

- the interval for reminding the participant to upload answers (murTime): every 24 hours,

- the interval for trying to download new forms from the server (pdTime): every 12 hours,

- the interval for saving the logged data (slTime): every hour.

After the parameters were set, three different APKs were created for the three groups and their different questionnaire modes.

6.1.4. Participant Acquisition

The number of 30 participants was chosen in order to have ten participants per group. This number should provide enough data samples to collect a valid and significant data set. The requirements for the subjects were to own an Android smartphone and to use the Facebook and any mail app.

Due to the requirement to have the Facebook app, a call to participate was made via Facebook. Additionally, Facebook friends, who were known to have a smartphone, were contacted with a direct message. The approach via Facebook was made in the hope to trigger a snowball effect. Contacted friends were asked to inform their friends about the study and these friends their friends and so on. Another post was made in the online forum for electrical engineering students to reach a bigger audience that probably has a smartphone. Also, notes with a call to participate were hung up on the pinboards at the cafeteria of the Technical University of Munich.

The information given to the participants included that 30 subjects would have to answer questions with their smartphone about their Facebook and their mail app usage during a six week period. Additionally, small rewards were listed in order to motivate persons to participate.

The respondents, who wanted to participate, had to confirm if they fulfill all requirements. Some of them only used the Facebook Messenger app or used Facebook with their smartphone internet browser. Others missed the fact that only Android phone users could take part in the study. Of

those who fulfilled the requirements, data on their smartphone model and the name of the mail app they used were collected. The information about the model was needed to make sure that the phone has the same or a higher API level required by the app. A running app was not guaranteed if the API level was lower. Moreover, the names of the used mail apps had to be known for the mail questionnaire in order to be logged. A regular expression was used to log all packages that include the string "mail". However, this would have not logged all mail apps as, for example, the K-9 mail app does not include the word "mail" in its package name. For this reason, the apps had to be known before the Questionnaire app could be given to the participants.

6.1.5. Time Frame

The study started with 24 participants on July 24, 2012. These participants were distributed into three eight person groups. The end date of their participation was September 4, 2012. The study was started with 24 participants because it was not sure whether the goal of 30 would be reached. Additionally, there was a deadline that required it to be started. One by one the additional six participants joined in on July 26, 27, 28, 29, 31. The last installed the app on August 2, 2012. The six participants were distributed equally in the three groups and had to answer questionnaires accordingly longer.

6.1.6. Distribution of the Questionnaire App

Five days before the study started, an info mail was sent to those participants, who started on July 24. It asked the participants to choose between installing the Questionnaire app by themselves or taking part in a personal meeting. In the meeting, the aim of the thesis, the Questionnaire app and the steps to install the app were presented.

The three different APKs were sent to the subjects one day prior to the study. The mail for all those who did not take part at a meeting included an app installation guide and the specification of their task during the study. The study officially started at 12 p.m. on July 24 and the first 24 participants were asked to answer questionnaires after every Facebook and mail app usage.

The other six subjects joined in the following and a half weeks and got an email with the APK and an installation guide. They had to install the app by themselves to instantly take part in the study.

6.1.7. Data Preparation

The logs and answers of the Facebook and mail app usages and questionnaires were viewed with the Evaluation tool and copied into Excel. First, all entries that lay before the start and after the end date were deleted. The study lasted for exactly six weeks and other data should not be

considered. Then, all successive answers indicating that a log entry could be missing were analyzed. Due to third party task killer apps, QuestionnaireService was sometimes stopped on some phones for better CPU efficiency. Also, QuestionnaireService can be stopped by the Android operating system itself to free up memory. The KeepAliveReceiver and the BootCompletedReceiver are designed to minimize the time when QuestionnaireService is not running. Another reason may simply be the fact that the user by mistake filled out the wrong questionnaire. All affected questionnaires were therefore looked up in the overall timeline to check for the reason and were deleted after that. Then, the data was analyzed.

6.2. Results

The results section presents all data gathered with the four questionnaires and the logging of the app usage. In order to indicate values for each of the three groups, the abbreviations

- **VOG**: **VO**luntary feedback diary and data logging **G**roup,

- **ITG**: **I**nterval-**T**riggered, experience-sampled diary and data logging **G**roup and

- **ETG**: **E**vent-**T**riggered, experience-sampled diary and data logging **G**roup

are introduced. The participants are numbered from test subject (TS) 1 to 10 for the VOG, TS 11 to 20 for the ITG and TS 21 to 30 for the ETG to ease the allocation of a participant to a group. The answers to the four questionnaires are listed in the order

- start questionnaire,

- Facebook and mail questionnaire and

- end questionnaire.

The questions of a form are enumerated with capital letters. These capital letters are used in the tables and figures as representation of the particular questions.

6.2.1. Start Questionnaire

In order to gain more information about the participants, the start questionnaire asked the following questions about the participants and their level of experience with smartphones:

A. What is your gender?

B. How old are you?

C. What is your profession?

D. What is your highest degree?

E. How long (approximately) do you already have a smartphone?

F. How long (approximately) do you already use the Facebook app on your smartphone?

G. How often do you use the Facebook app?

H. I believe that I am good in self-assessing how long I use the Facebook app per day.

I. How often do you use the mail app?

J. I believe that I am good in self-assessing how long I use the mail app per day.

K. What apps do you normally use in addition to the Facebook and the Mail app?

L. How many different apps are you using in a week?

M. What type of internet connection do you use?

N. Please give yourself a code name. If any problems should arise, I will write a general mail addressed to the code names.
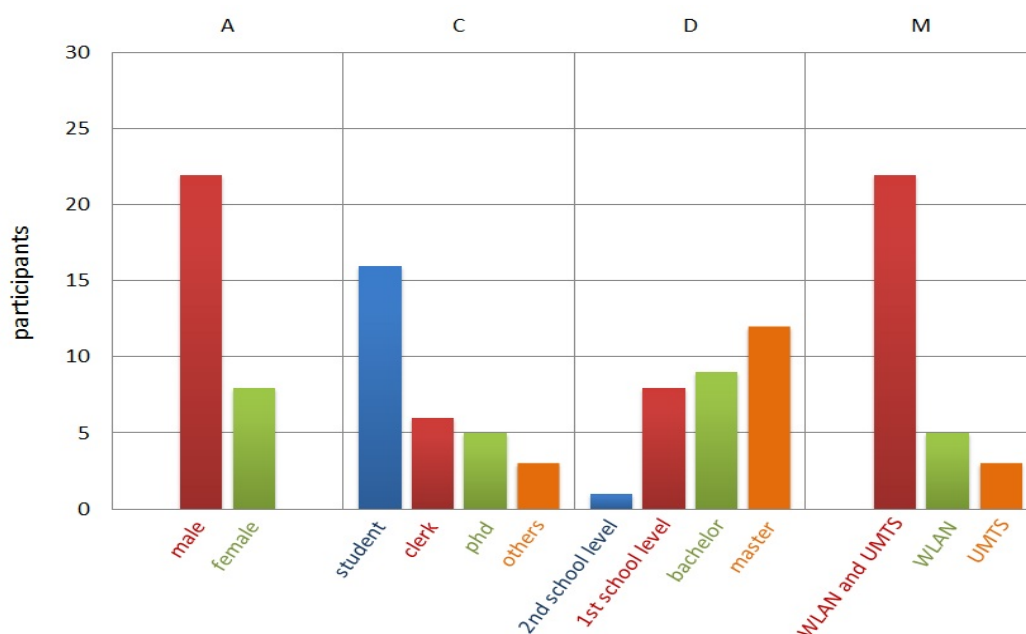
Figure 6.1.: Results of the start questionnaire of the main study for the questions A, C, D and M. Each bar represents the number of participants who chose this answer for the questions on the gender (A), the profession (C), the degree (D) and the used internet connection (M).

Figure 6.1 summarizes the results of the start questionnaire. The 30 participants consisted of 22 men and 8 women. They were almost equally distributed in the three groups. The event-triggered and the interval group had seven male and three female participants, the voluntary group eight male and two female. Therefore, no gender-based differences between the three groups should

occur. The profession of the subjects was queried in question C. 16 participants were students (VOG: 7, ITG: 4, ETG: 5), six were clerks and five were research assistants. Three subjects indicated that they were not students, clerks, freelancer, research assistants, professors or pupils, but something else. Question D called for the highest degree the participants achieved. There was one in the ETG with the second school level, eight with the first school level, nine with a bachelor and twelve participants with a master degree. The last question illustrated in Figure 6.1 reveals the kind of internet access with the smartphone. The majority of 22 subjects made use of both WLAN and the mobile network (VOG: 7, ITG: 7, ETG: 8). Five used only WLAN and three the mobile network.

Table 6.1 presents the answers for the questions B, E, F and L. The first query, query B, was on the age of the test subject. On average, the age of all 30 participants was 25.4 years. The distribution on the three groups was alike with the values 24.7 (VOG), 26.6 (ITG) and 24.8 (ETG). Thereby, the youngest participant was 18 years old and the oldest 32. Question E concentrated on the approximate number of months the participant had a smartphone. The general average was 19.9 months. Most of the participants from the voluntary group did not have a smartphone for that long, which led to an average of 11.8 months. The event-triggered group had a smartphone on average for 20.8 months. The majority of members in the interval-triggered group had a longer experience with an average of 27.2 months. Overall, there was one participant who had just recently bought it and two participants who chose the maximum value of 60 months. The next question, F, asked for the number of months since the participant has downloaded the Facebook app. The usage of the Facebook app ranged from minimum zero months to maximum 54 months. The average for the voluntary group was 10.0 months, for the interval-triggered 15.5 months and for the event-triggered 18.6 months, which led to a general mean of 14.7 months. Question L illustrates the average number of apps a participant uses in a week aside from the Facebook and the mail app. With a minimum average of 7.0 in the voluntary group and a maximum average value of 10.8 in the interval-triggered group, the general average of all 30 participants were 9.3 apps in a week. Thereby, the minimum number indicated by an user was three and the maximum number was 20.

Table 6.1.: Results of the start questionnaire for the questions on the age (B), the number of months since the participant has a smartphone (E) and has downloaded the Facebook app (F) and the number of apps used per week (L). The numbers are the averages of the three groups and the general averages.

	B [years]	E [months]	F [months]	L [apps]
VOG	24.7	11.8	10.0	7.0
ITG	26.6	27.2	15.5	10.8
ETG	24.8	20.8	18.6	10.0
general average	**25.4**	**19.9**	**14.7**	**9.3**

In order to gain more information about the used apps, question K offers a list of the most common ones. Figure 6.2 presents the answers to question K. As indicated, the maps or navigation app is the most common one. Only one subject does not make use of this app. With 26 of the 30 participants, the calendar app is the second most common one, third is the weather app with a number of 25 users. 20 subjects used other, not mentioned apps, 19 subjects a music app and 17 the YouTube application. The last option of apps named in the list was games. From the numerous game apps, only 14 subjects indicate to play them on the smartphone. In total 150 apps were reported. The aforementioned question L asked for the number of different apps used in a week. The total amount specified in this question was 278 per week. This value included the Facebook and the mail apps used by the participant. The value of 278 per week is almost the double of 150, which means that the participants either overestimated the used apps per weeks or that they have multiple apps for the listed categories in question K. From the listed apps, it can be assumed that a majority of persons uses only one calendar, one music, one weather and one YouTube application. Therefore, it can be assumed that only from the categories maps/navigation, games and others, multiple apps are used on one smartphone. The answers given by the participants indicate that some of them underestimated their number of used apps. Overall, the majority of subjects estimated a rather realistic number.
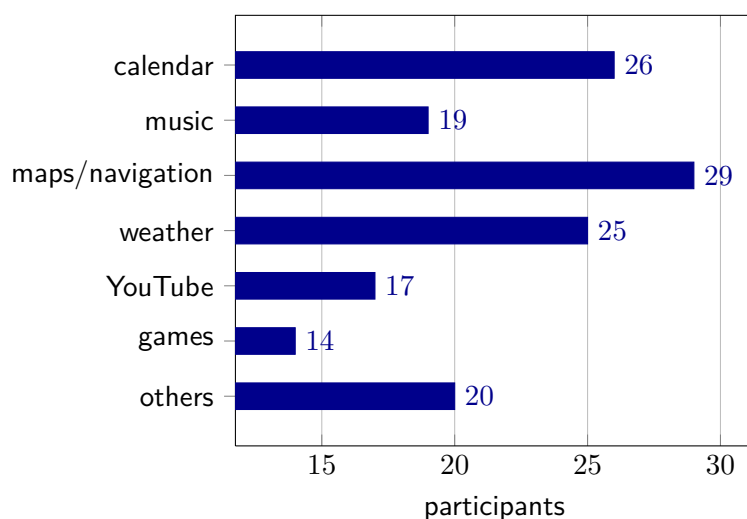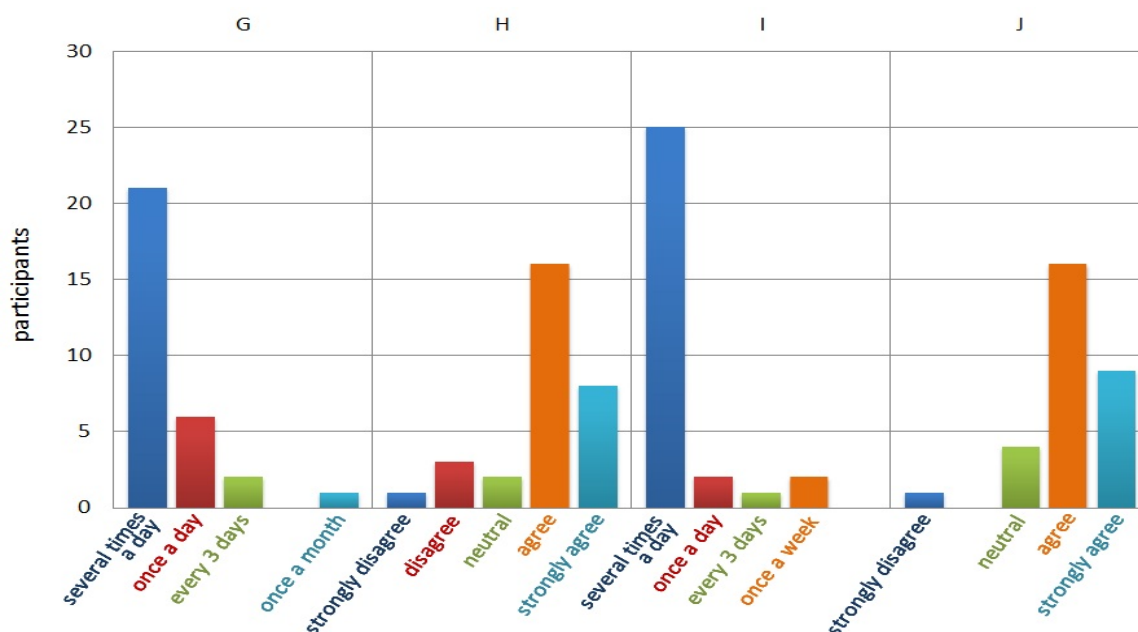

Figure 6.2.: Results of the start questionnaire of the main study. The bars represent the number of participants who indicated to use the app in question K. The most commonly used apps by the 30 participants are maps/navigation, calendar and weather.

Figure 6.3 presents the answers to questions G, H, I and J. Those questions were posed in order to gain information on the number of times the Facebook and the mail app are used and whether the participants are good in self-assessing the usage. Question G reveals that the Facebook app is started several times a day from 21 participants. They are equally distributed in the three groups. The same holds for the 6 participants that use the app once a day. From the remaining three

participants, two are using the app once every three days and one is using the app once a month. Asked about the ability to self-assess the usage of the app, one person chose the minimum value and disagreed strongly, three disagreed and two were neutral. The majority of 16 participants agreed and the remaining eight agreed strongly. Relying on this data, the comparison to the answers of the Facebook questionnaire and the logged data of the Facebook app usage should correlate for the majority of 24 participants.

The same questions were asked for the mail apps. 25 participants indicated in question I that they open the mail apps several times a day. Figure 6.3 shows that similar to the Facebook app, the number was almost equally distributed in the three groups. From the remaining five participants, two use the mail apps once a day, one person every three days and two start it once a week. The remaining question J demanded to estimate how good the participant is in self-assessing the mail apps usages. Similar to question H, 16 subjects agreed and nine even agreed strongly that they are good at self-assessing. One person disagreed strongly to this statement and the remaining four participants were neutral. Therefore, the comparison of mail questionnaire answers and the logged data should correlate for the majority of 25 subjects.

Figure 6.3.: Results of the start questionnaire of the main study for the questions G, H, I and J. Each bar represents the number of participants who chose this answer for the questions on the frequency of their Facebook (G) and mail (I) app usage and whether they are good in estimating the duration of their Facebook (H) and mail (J) app usage.

In summary, the three groups are similar in regards to gender, age, highest degree, Facebook and mail app usage. Therefore, the given answers should not be biased due to dissimilar groups of participants. Additionally, the majority of subjects use the two researched apps several times a

day and estimate their ability to self-assess their usages as good to very good. The logged data should, hence, correlate to the answers given in the questionnaires.

6.2.2. Facebook and Mail Questionnaire

During the six-week study, the participants had to answer a Facebook and a mail questionnaire. In the following, the questions of these forms will be presented. A complete version of both questionnaires including all answer options and prompts can be found on the enclosed CD.

Facebook Questionnaire

1. Why have you started the Facebook app?

2. What have you done after that?

3. Please estimate how long you have used the Facebook app at the last usage.

4. How often have you used the Facebook app since the last time you completed a questionnaire after its usage?

Mail Questionnaire

1. Why have you started the mail app?

2. What have you done after that?

3. Please estimate how long you have used the mail app at the last usage.

4. How often have you used the mail app since the last time you completed a questionnaire after its usage?

Activities in the Facebook and mail app

The first two questions of the Facebook as well as the mail questionnaire asked the participant to specify why he has started the app and what he has done after that. The second question was a multiple choice question. Therefore, all reported answers are summed up to 100% so that the combinations given by the participants are not listed separately. All answers of the first two questions were therefore analyzed equally for all participants for the Facebook and the mail app.

Facebook Questionnaire   Figure 6.4 presents all reported Facebook usages. The left pie chart shows the chosen answers to the first question on why the participants have started the app. The Facebook app was with a percentage of 73% most frequently started in order to read news. The

second most common cause with 16% was to read or write a message and the third most frequent one were other activities that were not mentioned in the answers option list. With a percentage of 2%, posting news, a picture or a video and adding or deleting a friend were rarely the reason for opening the app. The most infrequent reason to start the Facebook app was, with a percentage of 1%, reading or creating an event.
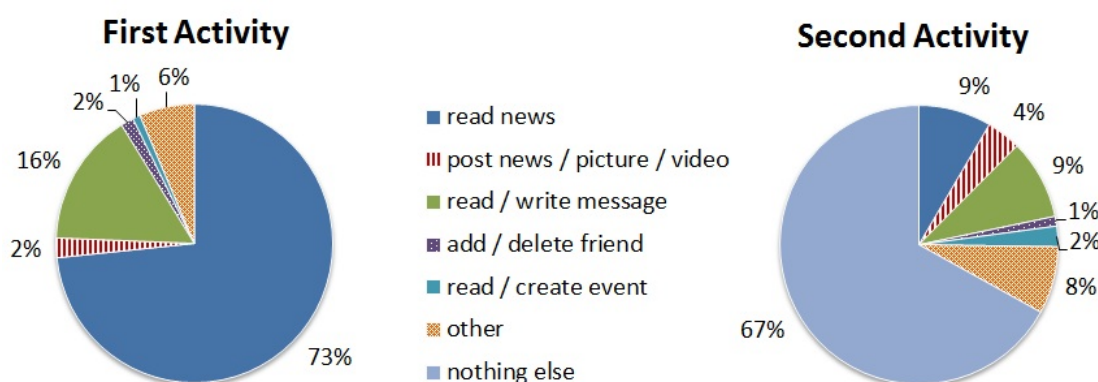
Figure 6.4.: Overall results for the Facebook app usage. The left pie chart represents the reported reason why the Facebook app was started. The right pie chart specifies further activities in the Facebook app.

The pie chart on the right, which displays further actions in the app, reveals that 67% of the participants did nothing else. 9% of the subjects read news and another 9% read or wrote a message. The third most common action were other activities than those listed in the options. 4% of the subjects posted news, a picture or a video and only 2% read or created an event. The least common action was to add or delete a friend.

In summary, the 30 participants opened the Facebook app most commonly to read news. After they had performed their first activity, a majority of 67% did nothing else and closed the app. Therefore, one can assume that the Facebook app is most often used for a quick news look-up. The majority of the remaining 33% read news, read or wrote a message and did other things. If we assume that the most common reason to choose "other" is the "like/dislike" functionality, the sequence of reading a new post and liking or disliking it could lead to the percentage of 8% for "other". The same holds for the answer option "read/write message", which was chosen with a percentage of 9%. The participants could have read news and, for example, wanted to write something personal to a friend who posted something. The last often chosen answering option with 9%, "read news", makes sense by looking at the 37% of users who did not open the app in order to read the news, but did something else. Some of these participants may wanted to read new posts after their first activity before they closed the app again.

Mail Questionnaire Figure 6.5 reveals that, similar to the Facebook activities, there is a common usage behavior when looking at the first and second activity. As can be seen in the left pie chart, the most common reason (75%) to start the mail app is to read mails. The second most chosen answer with 14% is that the participant deleted a mail, whereas 9% opened the mail app for other reasons than the given answer options. With a percentage of 1% "answering a mail" and "writing a new mail" were chosen and almost 0% chose the last option "sort mail".
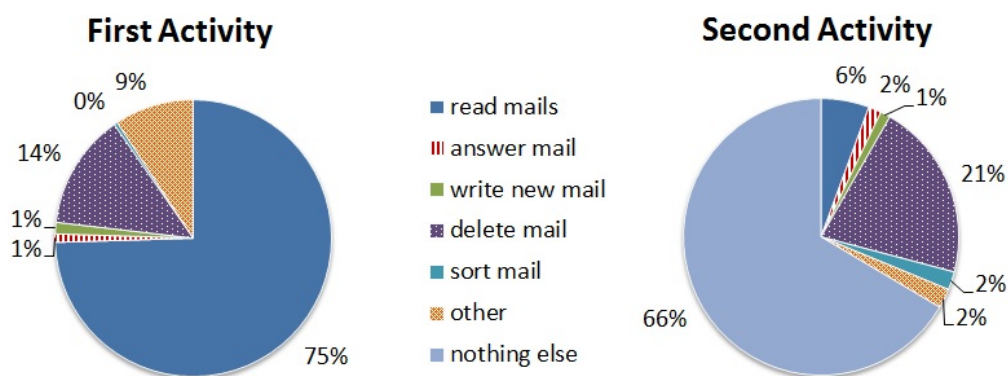
Figure 6.5.: Overall results for the mail app usage. The left pie chart represents the reported reason why the mail app was started. The right pie chart specifies further activities in the mail app.

As can be seen in the right pie chart, 66% of all app usages were only for one action and the participants did nothing else after the first action. With a percentage of 21%, mails were deleted as the second action. The third most common activity with 6% was sorting mails. Further 2% of all actions were answering a mail, sorting mail and others. Finally, 1% of the participants wrote a new mail as second activity.

Similar to the Facebook questionnaire, the most common reason to open the mail app is to read a mail. Usually, a notification appears at the top of the screen as soon as a new mail arrives. Then, the user clicks on the notification and the mail app starts. In case the subject deactivated automatic mail syncing or mail notifications, he has to check the mail app for new messages manually. The percentage of 9% for the answering option "other" could partly be explained as "checking for new mails". About 14% of the participants started the mail app in order to delete mails, for example spam mails, to free space in their mailbox.

As already seen in the Facebook app activities, most participants opened the mail app just for their original intention and did nothing else. Only 21% of all second actions were deleting a mail. Therefore, one can say, that most usages of the mail app were started because the participant wanted to read a mail and afterwards either deleted the mail or just closed the mail app. For this reason, the mail app could most often be not a replacement for the mail program on the computer, but an easy usable way for reading mails wherever the users are.

Reported Durations of App Usage

The third question of both forms asked the participants to estimate the last duration of the app usage (DAU). The histogram of all answered questionnaires can be seen in Figure 6.6 for the Facebook app and in Figure 6.7 for the mail app. It should be noted that this question was asked with the question type scaleedit of the Questionnaire app and a default value of two minutes.

Facebook Questionnaire   As can be seen in Figure 6.6, the most frequently chosen duration of app usage is two minutes, which corresponds to the default value. The value has been chosen 584 times, which is a percentage of 51.4% of all 1137 answered Facebook questionnaires. The maximum at two minutes can be explained in two ways. First, it could simply be the real estimation of the user for the duration of the app usage. Second, a majority of participants could just have left the default answer for this question unchanged. In order to determine what reason is more probable, the logged durations are evaluated on page 67. Figure 6.6 also reveals, that the other most commonly estimated durations are one minute (157 times), five minutes (91 times), three minutes (69 times), four minutes (62 times) and zero minutes (40 times). The amount of six to 16 minutes has been chosen between zero and 33 times. A duration of 20, 26, 30 and 58 minutes was chosen once and the maximal value of 60 minutes was estimated twice. In comparison to the logged data, which has a longest app usage of 32 minutes, especially the DAUs of 60 minutes seem overestimated.
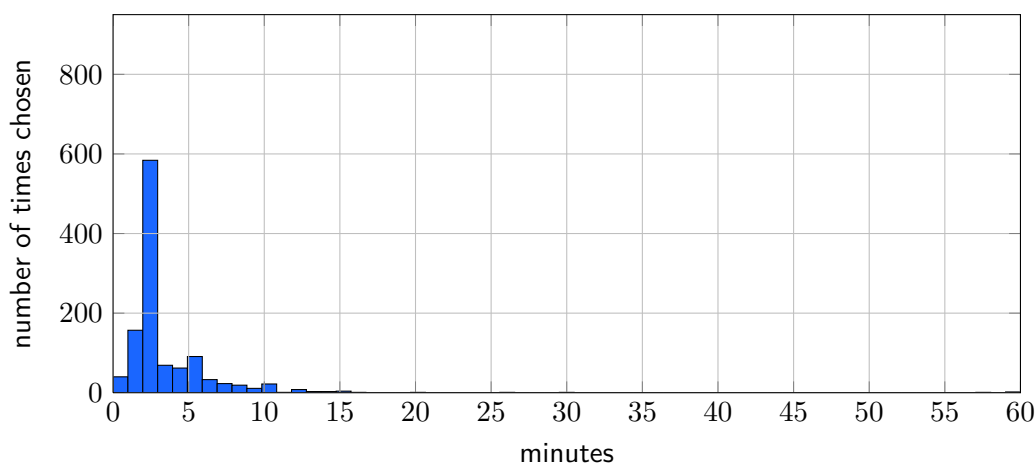
Figure 6.6.: Frequency distribution of the reported duration of Facebook app usage. The most frequently chosen duration was two and the maximum indicated value 60 minutes.

Mail Questionnaire   The histogram of the reported mail DAUs can be seen in Figure 6.7. It strikes at first sight that only the mail DAUs exclusively range from zero to eleven minutes. In contrast to the Facebook histogram of DAUs, the values are distributed differently. Again, the

most commonly reported DAU is two minutes. Overall, two minutes were chosen 908 times for the mail questionnaire, which represents 46.7% of the total 1943 questionnaires. In case the reported duration is correct, about the half of all mail app usages should last for about 2 minutes. The second most often reported duration is one minute with 489 times, followed by zero minutes with 373 times. The values three minutes (42 times), four minutes (39 times) and five minutes (34 times) are chosen similarly often. The same holds for six minutes (19 times), seven minutes (14 times) and eight minutes (11 times). The remaining values of nine, ten and eleven minutes are reported four, five and five times, respectively.
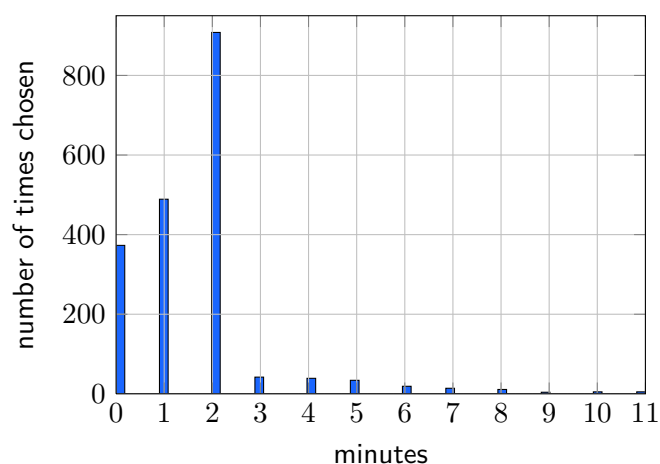
Figure 6.7.: Frequency distribution of the reported durations of mail app usage. The most frequently chosen duration was two and the maximum indicated value 11 minutes.

Unlike the answers for the Facebook questionnaire, the range of reported durations is a lot smaller for the mail questionnaire. Additionally, the reported durations, except for zero, one and two minutes, are slightly more evenly distributed for the mail app than for the Facebook questionnaire. Overall, both apps seem to be used between zero and two minutes in general according to Figure 6.6 and Figure 6.7.

Difference of Logged and Reported Duration of App Usage

The frequency of reported durations of app usages revealed a wide range of values for the Facebook and a smaller range for the mail app. In order to see how much the reported duration differs from the logged duration, box-and-whisker plots were created for the three groups. The difference was calculated by subtracting the logged duration from the reported duration. A positive value in the box plot is an overestimation of the DAU, while a negative value indicates an underestimation. The following figures are cropped to focus on the important regions and leave out extreme outliers. The box plots including all values can be looked up on the enclosed CD.

Facebook   Figure 6.8 presents the time differences between the reported and the corresponding logged durations of the Facebook app usage. The median for the voluntary group is zero minutes, while for the interval- and the event-triggered group it is an overestimation of one minute. The minimum value of VOG is one minute underestimation and the maximum value two minutes overestimation. Outliers can be found from minus five minutes up to fifteen minutes. The average of all the time differences for the VOG is 55 seconds. The interval-triggered group has a greater range with a minimum value of minus three and a maximum value of five minutes. Outliers for the ITG range from -11 up to 26 minutes. The average for ITG is 59 seconds. The event-triggered group holds the same values for the median, minimum and maximum value like the ITG. However, the outliers for the ETG are minus five up to 60 minutes. The average for the ETG is 111 seconds time difference.
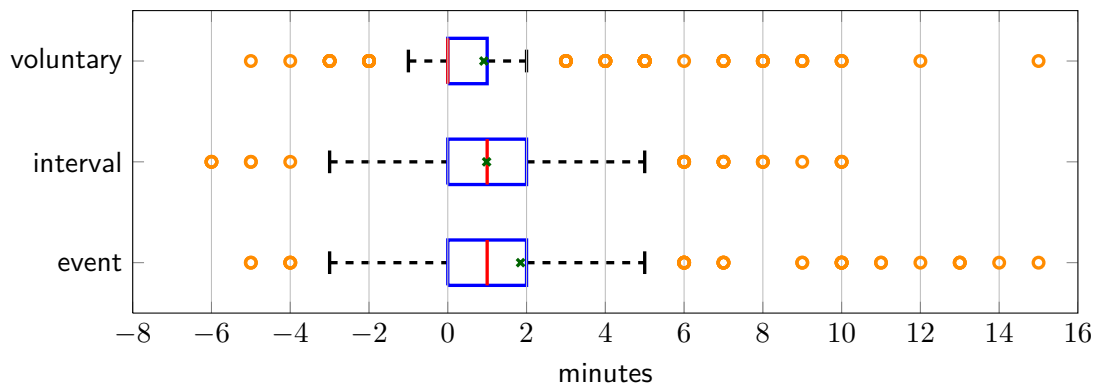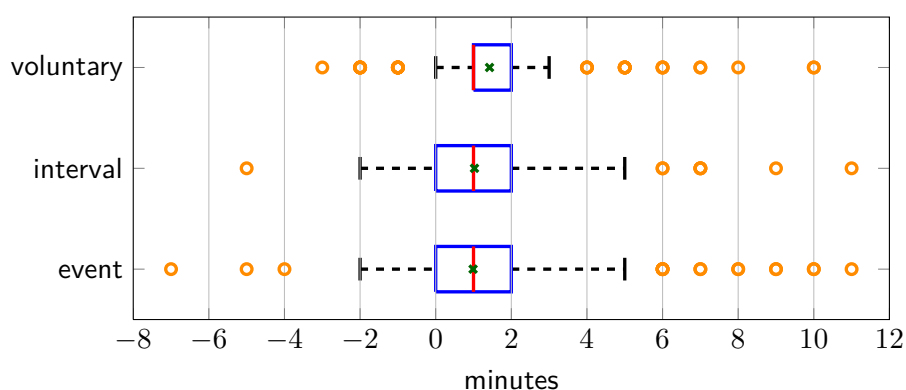

Figure 6.8.: Measured difference of reported and logged duration in minutes for the Facebook app usage (cropped). The ETG is the worst in self-assessing the DAU with an overestimation of 111 seconds, while the VOG is the best with 55 seconds.
Definition of symbols:
———: median, **x**: average, **o**: outlier, ——— upper: third quartile, ——— lower: first quartile, ——— upper: maximum value, ——— lower: minimum value

According to the averages of the three groups, all groups overestimated the application usage with one or two minutes, respectively. The upper and lower quartiles indicate a most common overestimation of zero to two minutes. According to the aforementioned numbers and the wide range of outliers, the event-triggered group seems to be the worst in self-assessing the Facebook DAU. The voluntary group, on the contrary, has the closest maximum and minimum values and the lowest average and is therefore considered to be best in evaluating the DAU.

Mail   The time differences for the mail durations of app usages have at first glance (see Figure 6.9) a smaller range than those of the Facebook app. Compared to the Facebook questionnaire, the box plot of the voluntary group is shifted by one minute. The minimum value is zero minutes, the median one minute and the maximum value three minutes. The outliers range from minus

three to ten minutes. This leads to an average overestimation of 85 seconds. The median of the interval-triggered group is also one minute. The minimum value is minus two minutes and the maximum five minutes. The outliers for the ITG range from five minutes underestimation up to 11 minutes overestimation. The average for the ITG is 62 seconds. The event-triggered group holds the same values as the interval-triggered group for the minimum and maximum, the upper and lower quartile and the median. The average of 59 seconds is similar to the one of the ITG. However, the outliers for the event-triggered group have the widest range with -24 up to 11 minutes.

Figure 6.9.: Measured difference of reported and logged duration in minutes for the mail app usage (cropped). The ITG and the ETG are on average equally good in self-assessing the DAU with an overestimation of ≈ 60 seconds, while the VOG performs worse with 85 seconds. Definition of symbols:
——: median, **x**: average, **o**: outlier, —— upper: third quartile, —— lower: first quartile, —— upper: maximum value, —— lower: minimum value

In comparison, the ITG and the ETG have a lower average as the VOG, while the median of one minute time difference is the same in all three groups. However, the outliers of the event-triggered group vary more than for the interval-triggered group. Therefore, the ITG seems to have had the best self-assessment for the mail DAU.

Overall The overall difference in the DAUs for the Facebook and the mail app is investigated by summing up the data of all three groups. Figure 6.10 presents the box plots and reveals that the median, the box and the whiskers are the same for both apps. Thereby, the minimum value is an underestimation of three minutes and the lower quartile is zero minutes. The median is one minute overestimation, the upper quartile two and the maximum value even five minutes. Outliers for the Facebook app range from 60 minutes overestimation to eleven minutes underestimation. The outliers of the mail app are distributed differently from 24 minutes underestimation to eleven minutes overestimation.

In average, the Facebook DAU is overestimated by 78 seconds and the mail DAU by 66 seconds.
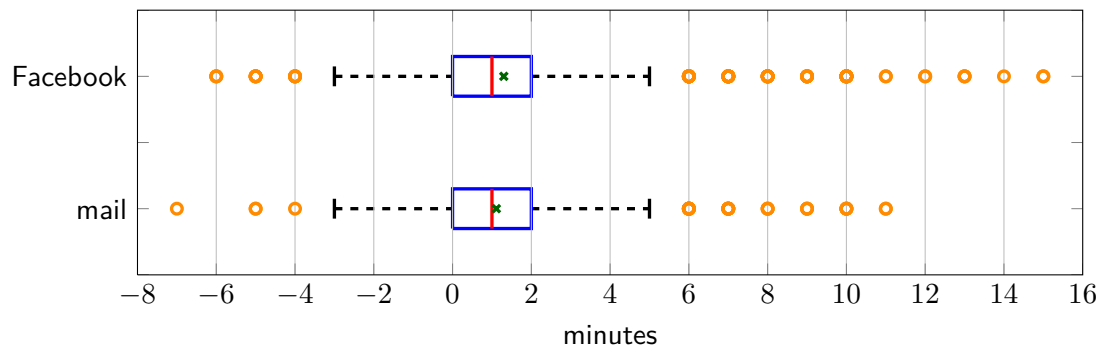
Figure 6.10.: Summary of measured difference of reported and logged duration in minutes (cropped). The self-assessment of the mail app usage was slightly better than the one of the Facebook app. Definition of symbols:
——: median, **x**: average, **o**: outlier, —— upper: third quartile, —— lower: first quartile, —— upper: maximum value, —— lower: minimum value

These numbers indicate that the participants are slightly better in self-assessing the duration of app usage for the mail than for the Facebook app. Especially the smaller range of outliers reveals a more realistic estimation of the app's usage duration. Also, the Facebook app usage seems to be counted more generously by the subjects. A time difference of 60 minutes suggests that these participants might have summed up a number of short usages. However, the overall box plots of the difference for the Facebook and the mail DAU reveal no great differences according to the distribution of the box plot values.

Comparison of Logged and Reported Average Duration of App Usage

After the analysis of the reported duration and the time differences between the reported and logged durations of app usages, the weightings of participants according to their produced amount of data will be deleted. The analyses made so far were calculated over the whole dataset of every participant. Now, the averages of the single participants will be used in order to equalize the weightings of all participants no matter how much data they produced. These newly calculated DAUs are required for further analyses in combination with the following results. The response rates, for example, will also be calculated with the averages of each participant. Figure 6.11 presents the average durations of every group. The left bar indicates the mean logged duration for the app usage of all ten participants of one group while the right bar shows the mean reported duration.

Facebook The average logged durations (ALD) for the Facebook app can be seen in Figure 6.11(a) and range from 1.49 minutes for the event-triggered group over 1.60 minutes for the voluntary to 1.69 minutes for the interval-triggered group. The average of the reported durations
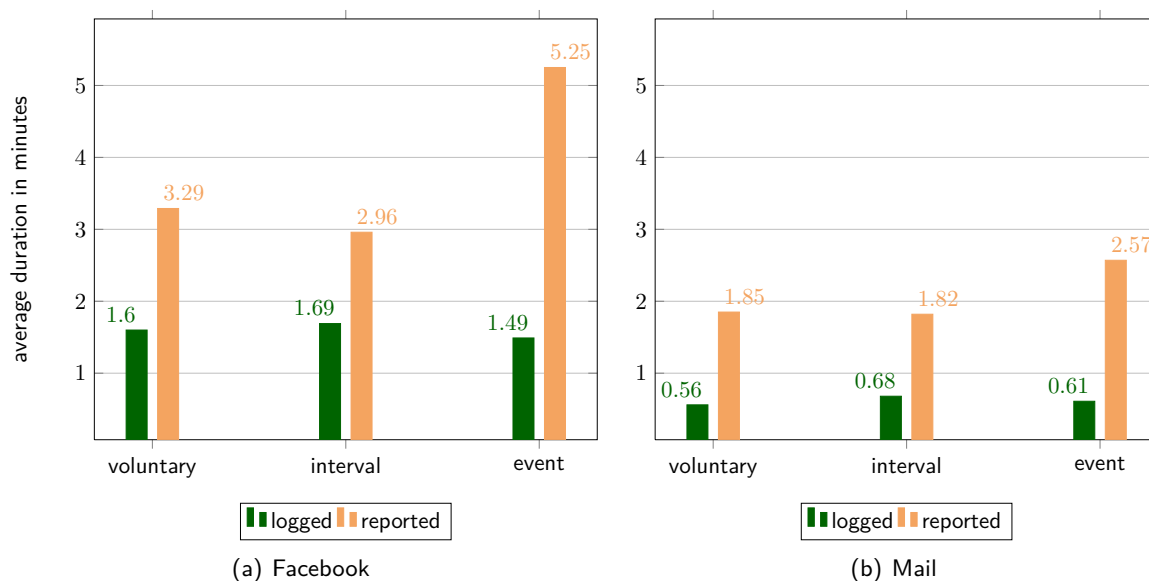
Figure 6.11.: Average duration of app usages for every group. The left bar (green) represents the average logged duration. The right bar (orange) shows the average reported duration.

(ARD) ranges from 2.96 minutes for the ITG over 3.29 minutes for the VOG to 5.25 minutes for the ETG. The difference between the two means is 1.69 minutes for the voluntary group. The minimum difference of 1.27 minutes is reached by the interval-triggered group. The worst group in self-estimating is the event-triggered with a difference of 3.76 minutes. The general average difference of duration is 2.24 minutes.

Mail The average logged durations for the mail apps are all lower than the ones for the Facebook app and range from 0.56 (VOG) over 0.61 minutes (ETG) up to 0.68 (ITG) (see Figure 6.11(b)). The average reported durations are also lower and range from 1.82 minutes (ITG) over 1.85 (VOG) to 2.57 minutes maximum for the event-triggered group. These values lead to a difference of 1.29 minutes for the VOG. The event-triggered group has the biggest difference with 1.96 minutes. Again, the minimum value is reached by the interval-triggered group with 1.14 minutes and the ITG's self-assessment can therefore seen as the best one for the mail app usage duration. Taking the average of all three groups, the time difference between ALD and ARD is 1.46 minutes.

Overall In comparison, the Facebook app was used about one minute longer than the mail app. Regarding the differences between reported and logged durations, the three groups together have an average difference of 2.24 minutes for the Facebook and an average difference of 1.46 minutes for the mail app. Thus, the average DAU is overestimated by the participants for Facebook as well as mail.

The first possible explanation for this is that the Facebook app seems to be used more unconsciously than the mail app and the real duration is overestimated. Another explanation could be the way of self-estimating the duration of app usage. The slider for this question in the Questionnaire app enables to enter only integers, while the real DAU is logged in milliseconds. The logged duration is rounded towards the closest duration in minutes. This leads to a value of zero minutes for 0-29 seconds, a value of one minute for 30-89 seconds, and so on. The differences between the logged and reported average durations for Facebook and mail vary between one and two minutes except for the Facebook app in the ETG with 3.76 minutes. A difference of one minute can be explained by a frequent app usage of under 30 seconds and a reported app usage of one minute. The subjects could have chosen one minute, because the app has been used and a value of zero could have seemed inaccurate to them. The mail app has a logged average duration of 37 seconds. These 37 seconds include a lot of logged mail app usages under 30 seconds which could lead to the aforementioned one minute difference. The third possible explanation for this time discrepancy is a different way of counting the number of app usages. The applied rule says that two logs can be merged into one usage in case they are interrupted for less than one minute. If this rule is applied with two minutes instead of one minute, more logs are merged, resulting in a smaller number of logs. This smaller number of logs includes logs with a longer duration of app usage. Greater values for the logged DAUs would have decreased the time difference between the average logged and the average reported durations.

Referring only to Figure 6.11 and not taking other factors into account, the interval-triggered group had the best self-assessment. Their difference between logged and reported average duration for both app usages was the smallest.

Response Rates of the Groups

The most interesting question this study can answer is: Which group has the highest response rate? Therefore, the response rates of all participants were averaged and resulted in the group response rates in Figure 6.12. The figure reveals that the response rates for the three groups are similar. The voluntary group has the lowest response rate for the Facebook questionnaire with 45.08% and the mail questionnaire with 52.09%. The interval- and the event-triggered group have an only slightly higher response rate of 52.59% and 52.86% for the mail form. Regarding to the Facebook questionnaire, the response rates between the groups range from 45.08% in the VOG over 48.88% in the ETG up to 58.53% in the ITG.

The last column in Figure 6.12 shows the "auto open" response rates of the ETG calculated by the ratio of the number of answered forms to the number of automatically triggered ones. For the Facebook questionnaire, the response rate is with 96.86% higher than for mail with 91.51%. Looking at the response rate of the automatically triggered questionnaires, the event-triggered group looks the most favorable. In comparison, the auto open response rates are almost doubled
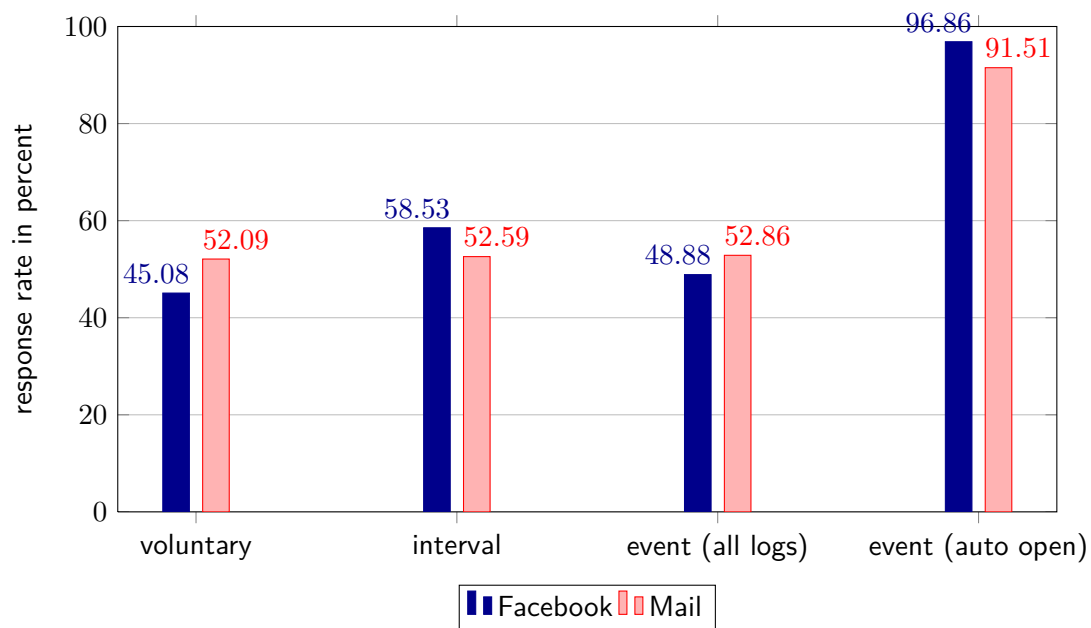
Figure 6.12.: Response rates of the three groups. The left (blue) bars represent the response rate for the Facebook questionnaire. The right (red) bars show the response rate for the mail questionnaire. The response rates for the Facebook and the mail questionnaire are similar for the VOG, the ITG and the all logs rate of the ETG. The response rates of the auto opened ETG are almost 100%.

to the response rates for all logs. This means, that the automatically triggered form only popped up half of the times the Facebook or mail app was closed. This effect is due to the fact that the questionnaire only appears in case the user switched from the app to the home screen. This limitation was implemented in order to not disrupt the user experience. Additionally, the phone model influences the time interval from app ending to the opening of the questionnaire. Samsung smartphones seemed to open the form instantly within a second, whereas some HTC phones took up to two seconds. Therefore, some users were able to open another app before the questionnaire popped up. A combination of these two reasons caused that questionnaires were triggered in 43.5% of all times the Facebook app was closed.

The response rates for the mail questionnaire between the three different groups are almost the same. For the Facebook questionnaire, the differences are bigger. This could be explained by seeing the Facebook app usage as time killer in periods where the participant has nothing else to do. He opens the app in some sort of automatism because he is used to, for example, read new posts in case he is bored or some time has passed. After such a start of the Facebook app it is likely to forget to answer a questionnaire. Therefore, the response rates of the Facebook questionnaires are lower than the ones from the mail questionnaire. The mail response rate for the interval-triggered group is the exception. An explanation could be that the daily notification for a missed form appeared for the mail questionnaire at 9:30 a.m. and for the Facebook questionnaire

at 7:00 p.m.. The participant could have ignored or deleted the notification for the mail form because the motivation to work or the work efficiency is higher in the morning and he does not want to fill out a form. Another explanation could be that the mail app is used more frequently during the day than the Facebook app. A daily notification only corrects one missed form. Therefore, the response rate of the Facebook form could be higher than the one of the mail.

In summary, one can say that the automatically opened forms are answered very reliably. However, it has to be considered that the participants were asked only about 50% of the time. The end questionnaire has to show, how convenient the users found that questionnaire mode and if a higher rate of automatically opened forms would still be okay for them. The interval-triggered group can be preferred in case all app usages are taken into account.

Response Rates of the Groups Including the Reported Number of Missed Questionnaires

In the Facebook as well as in the mail form, the last question asks the participant to estimate the number of times he has used the app without filling out the questionnaire since the last time he has answered it. By adding this number of reported forgotten forms to the number of answered questionnaires, a second type of response rate can be calculated. This response rate should be in the ideal case 100%, since the number of filled out questionnaires plus the number of missed ones should be equal to the number of logs. The question has been formulated in a way that all not reported app usages should be counted and, therefore, the auto open response rate for the ETG can be left out. Figure 6.13 illustrates the ratio of the self-estimated number of usages and the real app usages. The response rates for the mail form are similar and range from 54.83% in the event-triggered group over 58.01% in the voluntary group up to 59.14% in the interval-triggered group. The Facebook form response rates differ more, similarly to those in Figure 6.12. The VOG has with 51.19% the lowest ones, the ETG is with 58.15% in the middle and the best response rate has again the ITG with 69.18%.

As already mentioned, the response rate in this diagram would be 100% in the ideal case. This ideal rate of 100% was not reached by any group. One reason for this could be that the participants count an app usage differently from the applied norm. For example, the participants could rate a Facebook app usage which is interrupted by the sleeping mode and started again after five minutes as one use. However, following the norm it would be counted as two usages. Another reason could be that the participants simply underestimate the number of usages without answering a questionnaire. For this reason, the self-reports should not be used as indicator for the number of app usages according to this data. The combination of self-reports and logs presents a more reliable overview. However, if logging cannot be conducted the interval-triggered mode proved to have the highest reliability in this context.
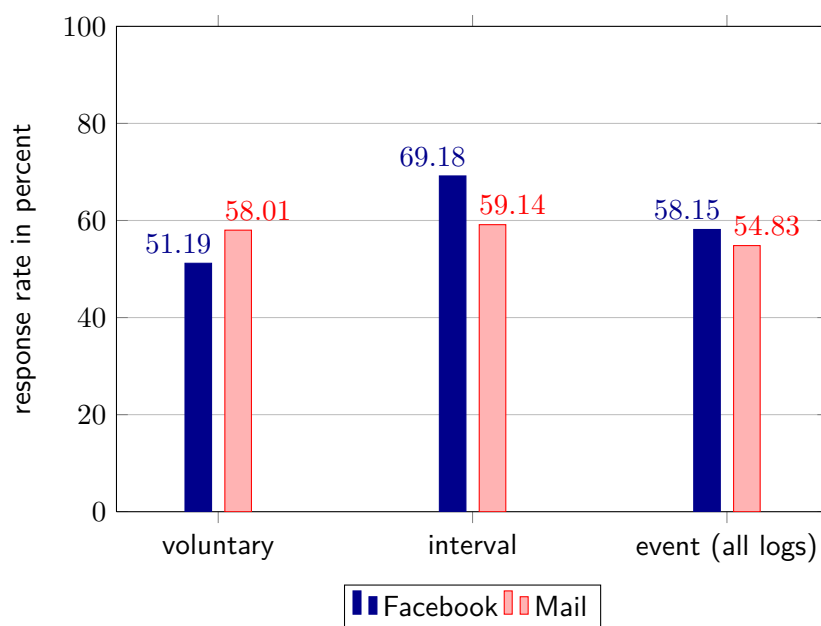
Figure 6.13.: Response rates of the three groups including the reported number of missed questionnaires. The left (blue) bar presents the response rate for the Facebook questionnaire. The right (red) bar shows the response rate for the mail questionnaire. The interval-triggered group achieves the best rates for the Facebook as well as the mail form.

Weekly Response Rates of the Groups

Another interesting aspect of the response rate is the trend over the course of six weeks. Figure 6.14 presents the weekly response rates for all three groups of the Facebook and the mail questionnaire, respectively. It should be noted that reminder mails were sent to the participants at the start of week three and week five.

Facebook   The response rates for the Facebook questionnaire ranged from 41.61% to 68.03%. The curve of the voluntary group starts in the first week with a response rate of 55.31%. The rate declines to the overall minimum of 44.47% in week two and another minimum of 46.22% in week four. In week three, the response rate is slightly higher probably due to the reminder mail. In the last two weeks, the rate increases to a final value of 56.14% in week six. This results in an almost 1% higher response rate in the last week than in the first.

The weekly response rates of the interval-triggered group are in general higher than the ones of the voluntary group and start in the first week with a value of 62.32%. Then, they alternately decrease and increase between 59.40% and the maximum 68.03% in week three. Week three and week five show higher values than the weeks before. In the last week, the response rate has a value of 55.59%.
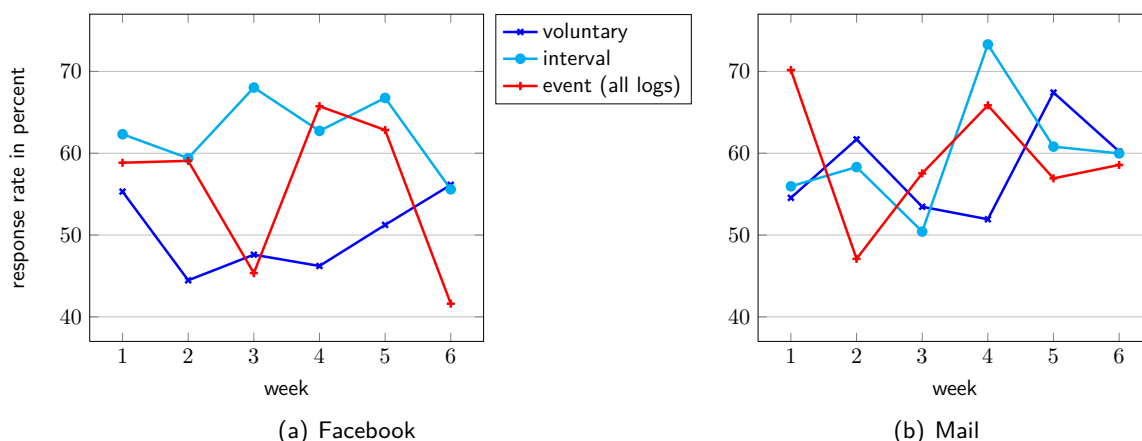
Figure 6.14.: Weekly response rates of the three groups for the Facebook and the mail questionnaire. For the Facebook and the mail questionnaire the response rates lie in a range between 40% and 74%.

The event-triggered group shows the greatest fluctuations throughout the whole study. It starts with 58.84%, stays almost the same in week two (59.07%) and decreases strongly to a value of 45.35% in week three. Then, the response rate increases to the maximum of 65.75%. After week four, the value decreases to 62.84% in week five and to the overall minimum of 41.61% in the last week.

In comparison, the voluntary group is the only one, whose end value is higher than the start value. The response rate of the interval-triggered group decreases from week one to week six by 6.73% and the one for the event-triggered group even by 17.23%. Thus, the response rates of the latter two groups reveal a decrease of motivation in the last week. For the event-triggered group, the response rate decreases by 21.23% in the last week. Therefore, it should be considered, whether the interval of six weeks is too long for a study that investigates an app that is most commonly used in the leisure time.

Mail   The differences between the weekly response rates of the mail questionnaire are bigger than those of the Facebook questionnaire and the values range from 47.09% to 73.30%. The voluntary group starts with a response rate of 54.55% and increases to 61.70% in week two. The response rate then decreases to the minimum of 51.93% in week four. It increases to the maximum of 67.42% in week five and decreases again to an end value of 60.22%. This results in an 5.67% higher end response rate than the start response rate.

The interval-triggered group starts with a response rate of 55.96% and increases to 58.31% in week two. In week three the minimum response rate is reached with 50.43%, while week four holds the maximum value of 73.30%. The response rate then decreases in the last two weeks to 59.98%. This end value of the ITG is 4.02% higher than the start value.

The last group, the event-triggered group, starts with the highest response rate of all three groups with a percentage of 70.17%. In week two, this response rate falls sharply to 47.09%. The value increases to 65.86% in week four. After this second maximum is reached, the response rate drops again in week five to a percentage of 56.93%. Finally, the participants answer more frequently in the last week with a response rate of 58.57%. The overall trend of the response rate from week one to week six is a decrease of 11.6%.

In comparison, the voluntary and the interval-triggered group reach a higher response rate in the last week of the study than in the first week. The minimum response rate occurred in week three and four, respectively. Therefore, the length of six weeks seemed to be too long for these two groups. The event-triggered group began with a high response rate of 70.17% and dropped by 23.08% in week two. The response rate increased again in the following weeks, but did not reach the maximum of week one. For this reason, the event-triggered mode seems to be convenient for only one week and then starts to be annoying. It also seems to depend on the users motivation.

Weekly Auto Open Response Rates for the ETG Calculated by the Ratio of the Number of Answered Questionnaires to the Number of Automatically Triggered Questionnaires   The aforementioned response rate for the ETG was calculated by the ratio of the number of answered questionnaires to the number of all logs. However, this *all logs* response rate is not completely convincing. Although the participants were asked to answer a questionnaire after every usage, it cannot be assumed that they remembered to fill it out when it did not appear automatically. Therefore, the weekly *auto open* response rates of the automatic triggered questionnaires were calculated by the ratio of the number of answered questionnaires to the number of automatically triggered questionnaires and can be seen in Figure 6.15. A value greater than 100% is possible if the majority of participants answered questionnaires additionally to those which were triggered automatically.

The response rates for the Facebook questionnaire start with 97.95% in week one and decrease to 90.28% in week two. In week three, the response rate increases again to 98.59% and decreases slightly in week four to 96.94%. In week five, the overall maximum of 103.29% is reached. The last week ends with a response rate of 95.95%. This value shows only a slight decrease of 2% between the start and the end value. The response rates for the mail questionnaire vary in a much greater range. The start value of 99.44% is even increased to 106.66% in week two. Then, the response rate drops strongly to 87.87% in week three, rises to the overall maximum of 111.63% in week four and declines to the minimum of 85.17% in week five. The last week has a higher response rate of 93.70%. From week one to six, the response rate decreases by 5.74%.

The mail questionnaire seems to depend more on the participant's motivation, which changes from week to week. The reminder mails probably increased the response rates of week three and five for the Facebook questionnaire, while the response rates for the mail questionnaire where higher in week four and six.
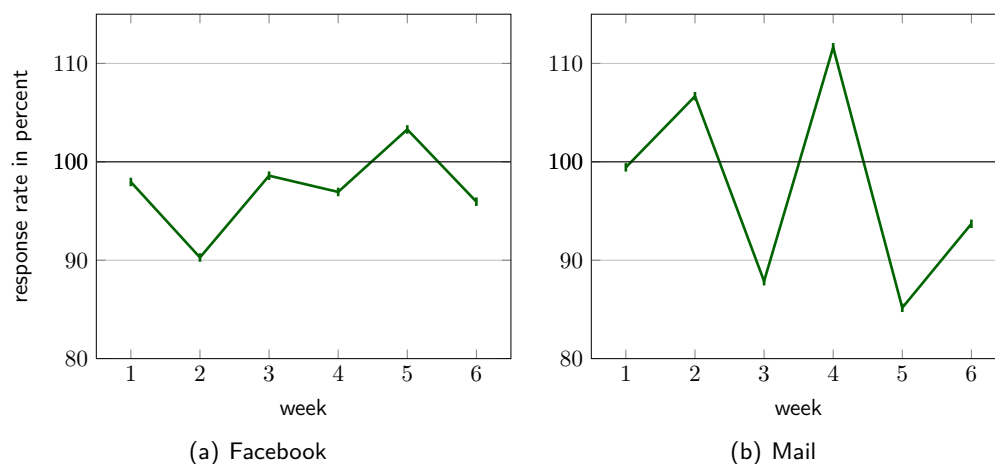
Figure 6.15.: Weekly auto open response rates of the ETG for the Facebook and the mail questionnaire. These rates are the ratio of the number of answered questionnaires to the number of automatically triggered questionnaires. A value greater than 100% is possible if the majority of participants answered questionnaires additionally to those which were triggered automatically.

Summary of the Weekly Response Rates For the **Facebook questionnaire**, the weekly response rates are, except for all logs of the event-triggered group, more constant during the period of six weeks than for the mail questionnaire. In comparison, the trends of the two ETG response rates do not match to each other. The ETG all logs curve in Figure 6.14 has its minimum in week three and its maximum in week four, whereas the auto open curve in Figure 6.15 has its minimum in week two and its maximum in week five. Furthermore, even though the overall response rate of the ETG in week three is only 45.35%, the auto open questionnaire response rate is almost 100%. Additionally, week four holds the overall maximum with 65.75%, while the auto open response rate decreases slightly by 1.65%. Therefore, it cannot be assumed that a more frequent appearance of automatically forms decreases the response rates. For further analysis, the number of logs per week have to be taken into account.

Considering the reminder mails sent at the first day of week three and five, an increase of the response rates in these weeks for all three groups (ETG: auto open curve) can be found. Hence, it can be assumed that for the Facebook questionnaire the mails of the conductor influenced the motivation of the participants. It remains unclear if it is due to feeling guilty not having answered all forms. Another reason could be the information how long the study will still last and the new motivation they got from it.

As aforementioned, the duration of the study of six weeks was too long and led in some weeks to low response rates. For the ITG, the recommendable study length with a consistent high response rate seems to be five weeks. For the voluntary group, the response rate decreases in the middle and increases at the end. This can be caused by new thrust of motivation. Therefore, a possible ideal study length could be three to four weeks. The reduced number of weeks could prevent the

low response rates in the middle of the study and the test subjects could stay motivated for the whole period of time. For the event-triggered group, an ideal length seems to be two weeks in case the frequency of events is as high as in this study. Week four and five have a higher response rate after the strong decrease in week three. The response rate decreases in week six and leads to the conclusion that this period of time is too long. Similar to the voluntary group, an ideal length seems to be three to four weeks.

A comparison of the two ETG curves for the **mail questionnaire** reveals their different trends. The all logs response rate in Figure 6.14 has its minimum of 47.09% in week two and never falls below 55% afterwards. The auto open curve in Figure 6.15, on the contrary, changes its trend every week. From week four to week five, it drops by a value of 26.46%. Thereby, it reaches its minimums in week three and five. The reminder mails do not seem to have positively influenced increasing the response rate as the values of these weeks are even the lowest ones. A similar trend can be seen for the curve of the interval-triggered group with week three and five showing the lowest response rates. Taken together, the reminder mails for ETG and ITG seem to have helped for the Facebook questionnaire, but not for the mail questionnaire. As the ETG and ITG curves drop steeply in week five (see Figure 6.14), a shorter duration of four weeks might have been appropriate. Furthermore, if the frequency of events is very high, the ideal length for the event-triggered group might be one week.

The voluntary group does not have an increase of the response rate from week two to week three, but from week four to week five. It seems that both reminder mails improved the mail response rates. It can be assumed that answering the mail form was easily forgotten by the VOG due to the number of mails a participant gets in one day. If the frequency of mail app starts is high, the participant would have to answer a lot of forms afterwards. The motivation might have risen with the second reminder mail when the subjects realized that the study would soon be over. The highest response rate in week five could also lead to the assumption of a possible ideal study length of five weeks for the VOG and the mail app. If a consistent high response rate during the study is needed, the number of weeks has to be further decreased. A length of three to four weeks could have been appropriate.

6.2.3. End Questionnaire

After the study ended, the participants filled out the end questionnaire. The statements of the end questionnaire had to be rated by the participants on a five-point Likert scale between "strongly disagree" to "strongly agree". The averages of the answers will additionally be given as numbers between 1 ("strongly disagree") and 5 ("strongly agree"). The end questionnaire consisted of the following statements and questions:

A. Answering the Facebook questionnaire was little effort.

B. I have always filled out the Facebook questionnaire directly after the Facebook app usage.

C. Answering the mail questionnaire was little effort.

D. I have always filled out the mail questionnaire directly after the mail app usage.

E. In which week did you have the feeling that you haven't filled out every questionnaire any more?

F. The graphical interface of the Questionnaire app is attractive.

G. The operation of the Questionnaire app is intuitive.

H. The performance of my battery has remained equally good despite the Questionnaire app.

I. Would you prefer the same questionnaire method in another study?

J. Answering the questionnaires has changed my usage behavior of the Facebook app.

K. Answering the questionnaires has changed my usage behavior of the mail app.

L. If your usages have changed, please explain briefly in what kind of way.

M. General comments on the Questionnaire app and the study.

Answers on the Questionnaire app

The Questionnaire app was created in order to conduct this study. It had already been tested in a pilot study and had been further improved. However, the three-day pilot study could not simulate the experience of participants who use it for six weeks. Therefore, the subjects were asked to rate the statements that the graphical interface of the app is attractive (question F), that the operation of it is intuitive (question G) and that the app did not change the performance of the battery (question H). The answers of all 30 test users are visualized as box-and-whisker plots in Figure 6.16.

The first plot in Figure 6.16 shows that the participants were on average neutral to the statement "The graphical interface of the Questionnaire app is attractive.". A slight positive trend of the mean 3.2 indicates that overall more participants found the app attractive. Three participants disagreed strongly to the comment. Two of them were from the event-triggered group and one was from the voluntary group. Test subject (TS) 6 from the VOG as well as TS 29 from the ETG gave no explanation in the free text. Only TS 25 stated that he would have preferred it if all questions were on one scrollable screen. Pushing the next button in order to get to the next question was too much effort for him. Additionally, he suggested to improve the slider for the scaleedit question type and to leave out the last screen, which thanks the participants for
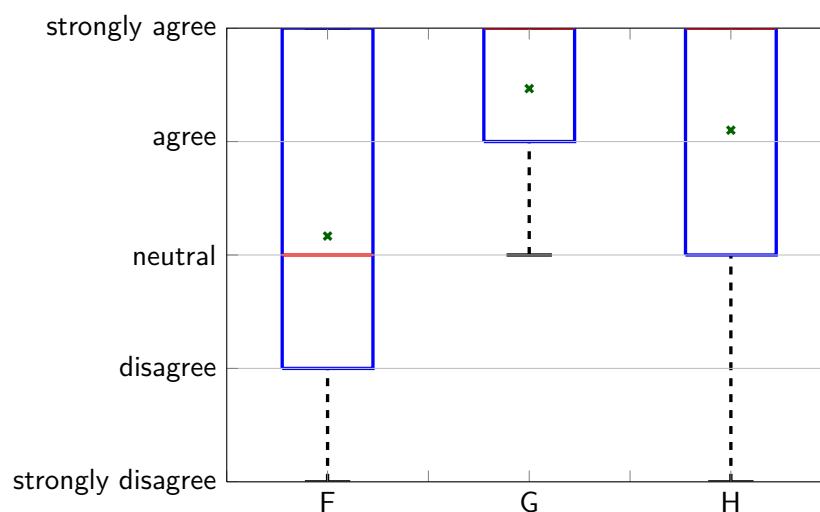
Figure 6.16.: Results for the statements F, G and H of the end questionnaire. Participants were on average neutral to the graphical interface of the app (F), agreed that the operation of it is intuitive (G) and that the app did not change the battery life (H). Definition of symbols:
——: median, **x**: average, **o**: outlier, —— upper: third quartile, —— lower: first quartile, —— upper: maximum value, —— lower: minimum value

answering the form. The lower quartile for the question is at the value "disagree". Two subjects (TS 27 and 30) from the ETG explained that they would have preferred the slider to have a smaller maximum value in order to ease the setting of the desired value. TS 27 also said, that the delay between closing the Facebook or mail app and the appearance of a form was too long. TS 1 rated the statement neutral and made the comment that a changeable default value for the scaleedit question type would be attractive to him. He stated that in most cases he used the Facebook app for just one minute and would have preferred this value as the default value. The slider was commented as improvable from TS 14 and TS 12 who said that they often did not change the value. Additionally, TS 12 would have preferred the app to allow returning to the last question in case something was answered incorrectly. From the participants who agreed strongly that the app is attractive, TS 8 would have wanted the free text answers in the end questionnaire to allow skipping to the next question even though nothing has been entered. Overall, he found the app "well designed". TS 10 stated that the Questionnaire app had an "attractive design".

In summary, the most frequently stated improvement to be made was the slider from the question type scaleedit. Therefore, the maximum value for another study of app usages should not be 60 minutes. Although some participants chose this value, a lower maximum, for example 20 minutes, would ease the setting for most of the participants. The suggestion to show all questions on one screen may seem preferable for such short questionnaires. However, it would be complicated for longer forms such as the end questionnaire. Additionally, questions could be skipped by mistake without answering.

The ratings of second statement presented in the Figure 6.16 shows the given ratings for "The operation of the Questionnaire app is intuitive.". The median of subjects agreed strongly to this comment. The average of persons agreed (strongly) with a value of 4.5. The minimum value was chosen by three participants who were neutral to it. They were all from the ETG and two of them indicated the slider not to be sufficiently. Participants who agreed strongly to it, made the comments that it is "easy to use" and that a form could be completed "fast and playfully"' (TS 3) and "well and fast" (TS 4). The operation of the app was rated as "very intuitive" (TS 7, TS 10).

The last ratings presented in Figure 6.16 are for the statement "The performance of my battery has remained equally good despite the Questionnaire app.". One subject disagreed strongly and marks the minimum of the box plot. Unfortunately, no comments were made how strong this change in battery life seemed to the subject. Being neutral was the lower quartile and indicates that the app changed the battery performance slightly. Apparently, the effect on battery life was moderate. The average of the participants agreed with a value of 4.1 and more than the half of them agreed strongly that the app did not change the battery life. Thus, the Questionnaire app did not seem to increase the perceptible power consumption for most of the subjects.

Answers on the Performance with the Facebook and the Mail Questionnaire

After analyzing questions on the app itself, the answers given on the performance of the Facebook and the mail questionnaire will be presented. These included the effort (A and C), the answering behavior (B and D) and the app usage changes (J and K). The given answers will be shown in box-and-whisker plots for each group.

Voluntary Group   As shown in Figure 6.17, the median of participants agreed with the statement "Answering the **Facebook Questionnaire** was little effort." (A). The average of 4.1 shows a slight trend towards "strongly agree" that marks the upper quartile. The minimum rating was "neutral" and only one subject disagreed and indicated that it was a great effort for him. However, for the majority of the voluntary group the Facebook questionnaire was easy to be answered. Being asked to rate "I have always filled out the Facebook Questionnaire directly after the Facebook app usage." (B), the median and the average (3.6) of answers was "agree". Two subjects agreed strongly. The minimum and the lower quartile for this answer was "disagree". TS 5 commented that "whether [because I'm] in a hurry or because I often forget it, I have not always filled out the questionnaire". The last statement that "Answering the questionnaires has changed my usage behavior of the Facebook app." (J) was answered with all possible ratings. Four subjects disagreed strongly and one of them commented that "nothing has changed with me...". TS 10 explained that for him there was "No more endless surfing and [I] reduced [the] usage". There was one test subject, TS 2, who agreed strongly that the form changed his Facebook usage and he also explained that he
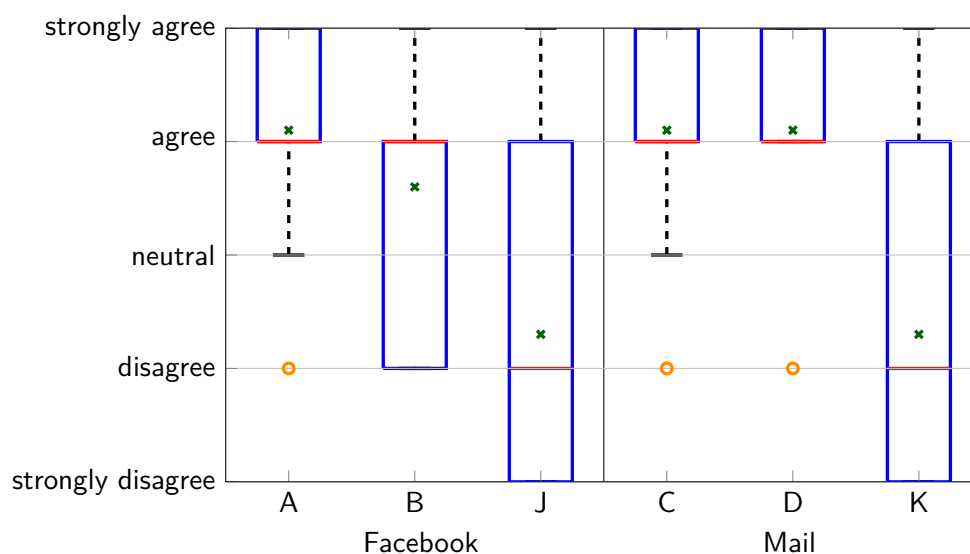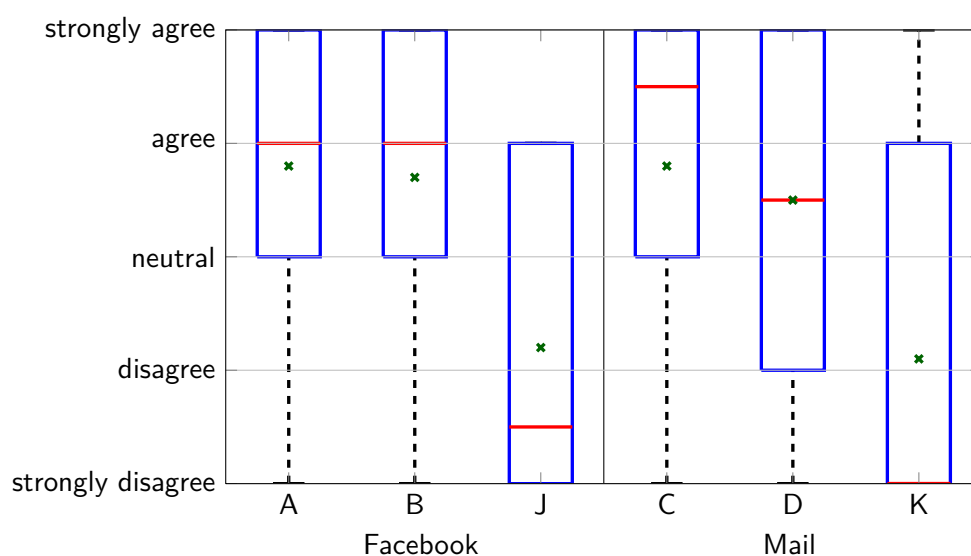
Figure 6.17.: Results for the statements A, B, J, C, D and K of the end questionnaire for the voluntary group. In average, both forms caused little effort (A, C), were often answered directly after the usage of the app (B, D) and changed the app usage slightly (J, K). Definition of symbols:
——: median, **x**: average, **o**: outlier, —— upper: third quartile, —— lower: first quartile, —— upper: maximum value, —— lower: minimum value

"preferred to look [it] up at the PC, if nearby. [It] is faster than app and questionnaire together ... and looking up is more comfortable...". Although the average of participants indicated with a value of 2.3 that their Facebook app usage did not change strongly, the overall number of logs did decrease over the six weeks due to the subjects who agreed (strongly) to this statement.

The given answers for the statement "Answering the **Mail Questionnaire** was little effort." (C) result in the same box-and-whisker plot as for the Facebook questionnaire. It can be assumed that for the majority of participants, filling out was due to the mean of 4.1 only little effort and only a few assessed it as a higher effort. A comment made by TS 2 revealed that it is "too much effort to fill out a questionnaire for 30 seconds after reading mails for 10 seconds". Contrary to the Facebook questionnaire, the box plot for the mail questionnaire indicates with a value of 4.1 that the majority of subjects always or most often filled out the corresponding form immediately (D). Only TS 4 disagreed to the statement and explained thereby, that he often forgot to answer the form. For the last statement shown in Figure 6.17 (K), the box plot is nearly identical to the one of the Facebook questionnaire. With an average of 2.3, the obligation to fill out forms changed the mail app usage slightly. It did not change the usage at all for some subjects, while TS 2 indicated again that he changed his app usage strongly like for the Facebook app. TS 3 commented that he "ha[s] no longer used both apps as frequently as before the study.". This change of behavior might explain the decrease of the number of logs for both apps.

Interval Triggered Group Figure 6.18 presents in the first box plot (A) that on average (3.8) the participants of the ITG estimated filling out the **Facebook questionnaire** as no big effort. However, there was also one subject, TS 16, who answered with "strongly disagree", but made no comment. The statement that they always filled out the Facebook questionnaire immediately (B) had the same distribution of ratings as statement A with an average of 3.7. TS 17, who chose "strongly disagree", explained that "[he] got more careless over time, but it wasn't disturbing!". According to the majority of subjects who stated that they always or most often filled out the form, the response rate for the ITG for the Facebook questionnaire should be relatively high. When considering the answers made to "Answering the questionnaires has changed my usage behavior of the Facebook app." (J), the average of 2.2 tends to the answer "disagree". However, the upper quartile indicates that some of them agree and, hence, changed their app usage. TS 14, who chose this rating, explained that he "used [it] less, in order to not have to fill out questionnaires". TS 12 wrote that his usage was "less [often and he] switched to the PC". TS 20 was neutral to the statement, but stated "Given that answering the form took longer than shortly reading a message or a mail, I [did not read the message].". Therefore, the number of Facebook logs did decrease over the course of the six weeks.

Figure 6.18.: Results for the statements A, B, J, C, D and K of the end questionnaire for the interval-triggered group. On average, both forms caused no effort (A, C) and were often answered directly after the usage of the app (B, D). They disagreed on average to the fact that they have changed their app usage behavior (J, K).
Definition of symbols:
——: median, **x**: average, **o**: outlier, —— upper: third quartile, —— lower: first quartile, —— upper: maximum value, —— lower: minimum value

The average, the upper and lower quartile as well as the minimum of "strongly disagree" for the statement "Answering the **Mail Questionnaire** was little effort." (C) was rated similarly to the

same statement for the Facebook form. The median for the mail form, however, is a little higher because more participants chose "strongly agree". Nevertheless, the effort for answering the form seems to be low for the majority of subjects. Being asked if they filled out the questionnaire directly after each usage (D), an average of 3.5 of the subjects agreed or were neutral to it. Only one test subject, TS 17, rated similarly to the statement for Facebook that he disagrees strongly and indicated that he rarely to never answered the form after the app usage. His response rate was, hence, very low because he probably used the apps very often during the day and reacted only to the notifications. According to the box plot, the response rate for the mail form should be lower than for the Facebook form. However, the ITG got a notification once a day in case they missed the answering. So, participants who did not use the apps multiple times a day should have a good response rate. The usage of the mail app has not greatly changed for most of the subjects (K). Seven participants disagreed or disagreed strongly to the fact that they changed their behavior, while the average has a value of 2.1. TS 12 comments, that he "[did] not look up every single mail" and TS 13 stated that he "[did] not check for mails as often as before (if not visible that a new mail [was] available)". Therefore, similar to the Facebook questionnaire, the number of logs for the mail app for the ITG did decrease during the six weeks.

Event Triggered Group   As can be seen in Figure 6.19, an average of 3.7 of the subjects agreed that answering the **Facebook questionnaire** is little effort (A). Three participants agreed strongly, whereas one test subject, TS 24, disagreed strongly and made no further explanation. However, he commented that the slider is improvable and this could have influenced the rating. The next statement shows no box and no whiskers because eight subjects agreed with an average of 4.2 that they have always filled out the questionnaire directly after the Facebook app usage (B). The remaining two participants indicated a strong agreement. This high rate of agreement shows that the event-triggered mode of the form leads to a high response rate. Being asked whether the form changed the app usage (J), the subjects agreed or were neutral with an average of 3.5. Three participants admitted a strong agreement. However, there were three participants who disagreed or disagreed strongly. TS 23 commented that "[he] used the apps considerably less" and TS 25 said that "you wonder (primarily within the last weeks) whether you really should start the app". Another comment from TS 25 was that he switched to other apps like the Facebook messenger. Therefore, the number of logs for the Facebook app did decrease strongly over six weeks study.

The answers on the effort of the **mail questionnaire** (C) reveal that the median of chosen ratings was agreement and the average has a value of 3.4. Two subjects agreed strongly that it was little effort. However, two participants disagreed and one disagreed strongly to the statement. TS 24 stated that "the questionnaire should start directly after closing the app". This comment can be explained with the short delay that is caused by the HTC phone of the participant. Eight subjects agreed with the statement that they directly answered the questionnaire (D). The remaining two subjects agreed strongly. Therefore, the average has a value of 4.2. This can be explained by the
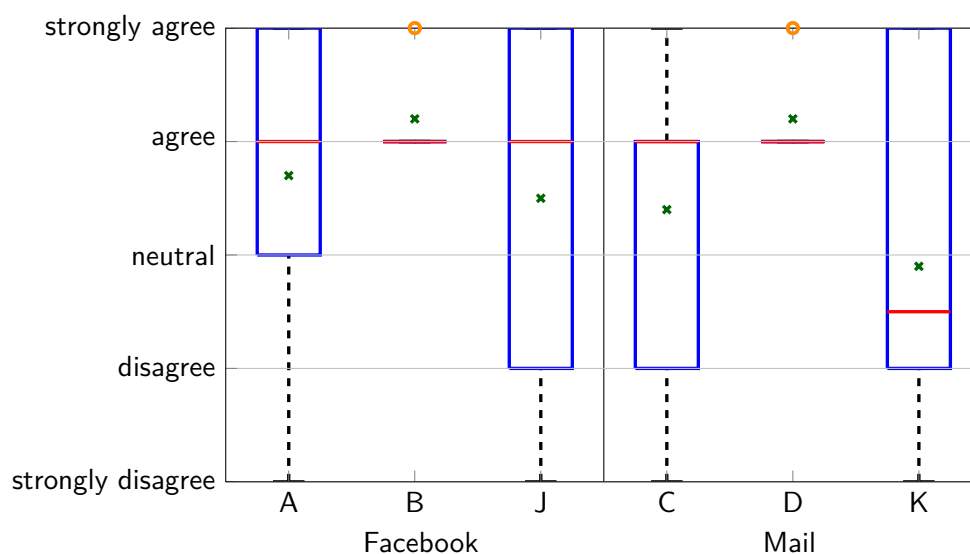
Figure 6.19.: Results for the statements A, B, J, C, D and K of the end questionnaire for the event-triggered group. The ETG rated the forms on average as little effort (A, C) and indicated to have answered the forms usually directly after the usage (B, D). The members of the ETG changed the mail app and the Facebook app usage during the study (J, K). Definition of symbols:
——: median, **x**: average, **o**: outlier, —— upper: third quartile, —— lower: first quartile, —— upper: maximum value, —— lower: minimum value

event-triggered mode that automatically opens the form after an app usage. However, the all logs response rate for the Facebook as well as the mail app was only about 50%. The participants have therefore either interpreted the statement to refer only on the automatically triggered forms or forgot to also answer questionnaires when they did not appear. The task to fill out questionnaires did change the app usage of three participants strongly (K). Two subjects were neutral to the statement. The remaining five subjects stated no or a slight change. One of them, TS 21, commented that he "partially looked up mails directly at the PC, when [he] was to lazy to fill out the questionnaire on the smartphone afterwards". TS 22, who changed his app usage strongly, wrote that "[he] used the mail app in the end only when [he] had no laptop around". Another explanation from TS 24 was that "[he] often just read the notification, but did not start the app with it, in case something new came in". Therefore, the usage of the mail app has been changed, similar to the Facebook app, with the automatically opened questionnaires.

Given that the form appeared on average only in 50% of the app uses, this change of has to be considered for other studies. In case the conductor of the study wants a self-reporting for every usage, he has to make himself aware that this is going to change the usage of the studied object. In comparison, the average of participants of the other two groups did not change the usage that strongly, according to the given ratings. The advantage of a event-triggered form is

that the subjects answered on average more questionnaires after the usage than participants from the other groups.

Answers on the Weekly Performance

Statement E from the end questionnaire asked the participants to choose in which week they did not fill out every questionnaire any more. Figure 6.20 shows the frequency distribution of answers and shows an increase over the six weeks. No test subject stated that he missed forms in the first week, but TS 26 from the event-triggered group admitted his lower response rate in week two. Nine participants indicated not having answered all questionnaires in week three (VOG: 4, ITG: 1, ETG: 4). The lower response rate for the ETG can be explained from the effort the forms seemed to be for some of them. In week four, 14 out of 30 participants indicated that they have not answered every questionnaire (VOG: 6, ITG: 4, ETG. 4). Week five reaches the maximum of 16 participants with five members of the VOG, five of the ITG and six of the ETG. Thus, 50% of each group either forgot or simply did not want to answer a form. For the voluntary and interval-triggered group, the participants frequently seemed to forget the task. However, the same cannot be assumed for the event-triggered group with the questionnaires being automatically opened. Although a reminder mail was sent at the beginning of this week that highlighted the near end of the study, the event-triggered group seems to have lost their motivation. This assumption is also transferable to some subjects of the other two groups. In week six, fewer participants indicated to have missed forms. The reason could probably be a thrust of motivation that the study would soon end.
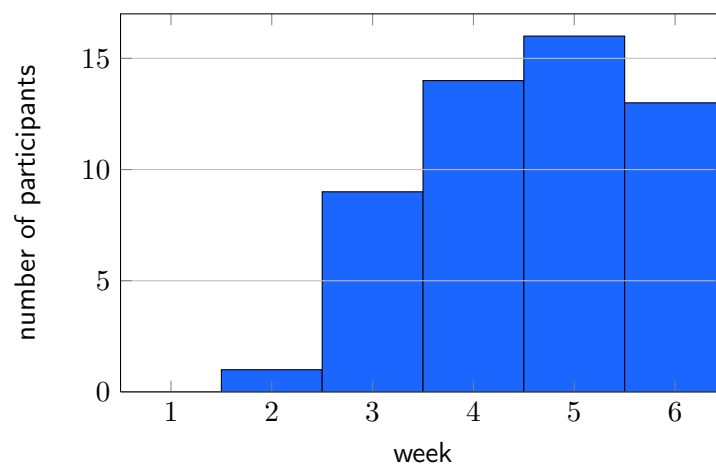
Figure 6.20.: Frequency distribution to statement E of the end questionnaire. The number of participants who reported a lower response rate increases from week to week and reaches its maximum of 16 in week five.

Answers on the Questionnaire Mode

Being asked about the questionnaire mode and whether the participant would want the same method for another study, six subjects of the voluntary group would have preferred the same method for another study. TS 4 wrote that this is due to "the low effort and [the fact] that it can be filled out any time". TS 10 additionally stated that "the question mode fitted perfectly for this study and that [he] could picture this method also for another study". From the four remaining members of the VOG, two indicated that they would have preferred the form to open itself automatically after the usage "so that the forgetfulness is prevented" (TS 8). TS 5 "would have liked to be asked automatically, once or several times a day for filling out a form.[..] A daily notification would have been enough". Another questionnaire mode was demanded by TS 2 by saying that it was "too much effort to answer a questionnaire for 30 seconds after reading mails for 10 seconds... rarer - every fifth time or so would perhaps be better...". Depending on the aim of a study, the conductor has to decide how much of the usage he wants to be self-reported.

In the interval-triggered group, two participants liked the questionnaire mode. One participant simply wrote "no" and one would have preferred only logging and no self-reports. The event-triggered group had one participant who was irresolute concerning this question. Two subjects did not like the questionnaire mode because it was "too time intensive and circumstantial. The 'periodical reply' of the same questions again and again was disturbing. You partly even avoid the use of the apps." (TS 23). TS 26 wrote that "it was too annoying. Rather log in the background instead of asking for a self-assessment." The remaining subjects liked the automatically triggered questionnaire even though "the frequency is partly very exhausting ... over time" (TS 22).

Overall, eight participants liked the interval-triggered mode, seven the event-triggered mode and six the voluntary mode.

Answers on the Changed App Usage

Besides the aforementioned comments on the changed behavior concerning the apps, there were some statements about its positive side effects. TS 1 wrote that he had a "more conscious use of the apps, only when [he] really wanted to use it and [did] not simply start [it] because there was nothing to do". TS 10 added that "[there was] no endless surfing any more and a reduced usage". A "more conscious estimating of time" was made by TS 19 and TS 29 "paid more attention on updates". This comment can be interpreted that he did not check for mails or Facebook news when there was no notification. Finally, TS 28 stated that "now he knows how much time [he] waste[s] with it! [He] should use [his] time with other matters.".

Although the change of behavior is not helpful for the conductor of the study who wants to investigate the usage of the apps, the participants are able to become more conscious on how they spend their time.

General Comments on the Study and the App

Some comments which were made at the last, free text question were already mentioned before when discussing the answers on the performance and the usage of the apps. Additionally to them, TS 2, a member of the voluntary group, declared the "kind of survey as annoying". He did not like the effort of answering a questionnaire after a short app usage. TS 9 stated that it is "very annoying to fill out a questionnaire after each usage [of the app] since nowadays there are permanently notifications for Facebook or for new mails". From the interval-triggered group, TS 20 wrote that "since [he] partly very frequently read[s his] mails, the effort was considerably higher than [he] would have imagined it to be in advance". TS 13 explained that it has been "interesting for [himself] how often one starts the app" and that it has been an "acceptable expenditure of time". TS 7 wrote simply "super" while TS 6 stated that it is a "very intuitive and actually ingenious survey method".

Chapter 7.

Discussion

In the last section, the individual results of the main study were presented and analyzed. In this chapter, the combination of all collected results will be investigated. This analysis is made with regard to the hypotheses of section 6.1.2.

7.1. Ability to Estimate the Average Duration of App Usage

Facebook   In the start questionnaire, 24 of the 30 participants agreed or even agreed strongly to the statement that they are good in self-assessing the duration of the Facebook app usage. Therefore, most of the reported app durations should be close to the logged app duration. The weighted time differences of the three groups, which can be seen in Figure 6.8, led to the following results. In the "weighted" time differences, participants who produced a large data set have more impact on the group time difference than those who produced a small data set.

The time differences for the Facebook app show that the voluntary group has the best self-estimation with an average of 55 seconds overestimation, second is the interval-triggered group with 59 seconds and the event-triggered group appraise themselves worst with 111 seconds over-estimation. When looking at the unweighted but averaged time differences in Figure 6.11(a), the event-triggered group also has the highest value of 226 seconds. However, the estimations from the interval-triggered group are with a mean of 76 seconds better than those from the voluntary group with 101 seconds. The overall mean difference is 134 seconds overestimation. Considering the logged mean app usage duration of 96 seconds for the Facebook and 37 seconds for the mail app, hypothesis **H1**, which states that the participants underestimate their app usage, can be rejected for the Facebook app.

Mail   Similar to the Facebook app, 25 of the 30 participants indicated that they are good at estimating the duration of the mail app usage. Thereby, only five of the 30 participants should be responsible for a lower rate of correctly estimated mail app durations. Looking at the weighted time differences, the averages for the event-triggered and the interval-triggered group are both

about 60 seconds overestimation, while the voluntary group overestimated the logged app usage by 85 seconds. If the mean of all participants is investigated, the interval-triggered group has the best self-assessment, followed by the voluntary group. The event-triggered group estimated the mail app usage worst by 118 seconds overestimation. Overall, the 30 participants overestimated the mail app usage by 88 seconds. This leads to a rejection of hypothesis **H1**.

Comparison of Mail and Facebook   The analysis of the time difference between the reported and the logged durations of app usage proved hypothesis **H1** to be wrong because both DAUs have been overestimated by the participants. In comparison, the Facebook app usage has been overestimated by 134 seconds, while the mail app usage has been overestimated by 88 seconds. The hypothesis **H2**, which says that the mail DAU is better estimated than the Facebook DAU, cannot be accepted because the difference is not significant (T-test: $t(29)=2.05$, $p=0.10>0.05$). Although the difference is not significant, the Facebook DAU is estimated worse than the mail DAU. This leads to the conclusion that the mail app is used more consciously than the Facebook app. Another explanation could be the default value of two minutes for the slider of the reported duration. As can be seen in Figure 6.6 and 6.7, this default value has been chosen for both apps in about half of the forms. However, the logged duration of the Facebook app is with an average of about 95 seconds closer to the default two minutes than the mean of 37 seconds of the mail app. The estimation of the mail app should be worse due to the greater difference of the logged duration to the most frequently reported value of two minutes. Therefore, the influence of the default value of the slider can be eliminated as an explanation. A third explanation could be that the shorter DAU of the mail app is easier to self-assess than the longer DAU for the Facebook app.

7.2. Response Rates

One aim of this diploma thesis is to investigate, which kind of questionnaire mode provides the highest response rate in which context. The formulated hypothesis **H3** states that the response rate of the interval-triggered group is higher than the response rate of the voluntary group. As can be seen in Figure 6.12, the response rate for the mail questionnaire is 0.50% higher for the ITG than for the VOG. For the Facebook questionnaire, the difference is even 13.45%. However, neither the difference for the Facebook response rates (T-test: $t(18)=2.10$, $p=0.35>0.05$) nor for the mail response rates (T-test: $t(18)=2.10$, $p=0.97>0.05$) are significant. Therefore, hypothesis **H3** is not proven, although the interval-triggered mode with a daily notification increased the response rates in comparison to a simple voluntary diary slightly more. The higher difference between the groups for the Facebook questionnaire can be caused due to the number of times the Facebook app is opened in order to kill some spare time. The action to open the Facebook app is

unconscious and the task to fill out the form can be forgotten. The ITG gets a daily notification in case the participants missed to answer a form so that they are able to fill out the questionnaire at a later point of time. This notification also reminds them to answer forms in general. However, this did not impose a great difference on the response rates for the mail questionnaire. The ratings of the VOG made in the end questionnaire can be seen in the columns B and D of Figure 6.17. They show that the participants estimated their response rates correctly. The average of the chosen degrees of agreement to a high rate of reported app usages is higher for the mail questionnaire just like the response rates in Figure 6.12. The same correct self-estimation can be seen for the ITG response rates and the given ratings in the end questionnaire in Figure 6.18.

Another hypothesis, **H4**, was that the event-triggered group has a higher response rate than the interval-triggered group. Figure 6.12 illustrates that if all logs are taken into account, the hypothesis is not true for the Facebook questionnaire but for the mail questionnaire. However, if only the response rates for the automatically triggered forms are considered, hypothesis **H4** can be accepted due to an almost doubled response rate of the ITG (Facebook: 58.53%, mail: 52.59%) in comparison to the response rate of the ETG (Facebook: 96.86%, mail: 91.51%). The event-triggered mode, which triggers the appropriate form to appear automatically on the screen, thereby improves the response rates for the Facebook questionnaire (T-test: t(9)=2.26, p=0.00002<0.05) as well as the mail questionnaire (T-test: t(9)=2.26, p=0.00026<0.05) significantly. As aforementioned, the self-estimation of the response rate was correct for the ITG. The ETG ratings in the end questionnaire in Figure 6.19 are identical for the Facebook and the mail form, although there is a slight difference in the actual response rates.

Overall, the response rates for the Facebook questionnaire were 5.94% and 5.55% higher compared to the mail questionnaire for the interval-triggered group and the automatically event-triggered group, respectively. Only the voluntary group answered the mail questionnaire in relation more often than the Facebook questionnaire. The averages of the VOG, the ITG and the automatically triggered response rates of the ETG are 66.82% for the Facebook and 65.40% for the mail questionnaire. However, this difference is not significant (T-test: t(29)=2.05, p=0.59<0.05) and hypothesis **H5** can therefore not be accepted. The slightly higher response rate of the Facebook questionnaire could be caused by the inappropriateness that answering the mail questionnaire sometimes imposes during work time.

7.3. Weekly Response Rates

Another interesting aspect are the specific weekly responses of every group and their development during the course of six weeks. **H6** states, that the weekly response rates decrease over the six-week study length. In Figures 6.14 and 6.15, no linear decrease of the response rates neither for the Facebook questionnaire nor for the mail questionnaire can be seen.

The response rates for the Facebook questionnaire increase for the VOG by 1.00% from week one

to week six, while they decrease for the ITG (6.73%) and for the ETG (all logs: 17.23%, auto open: 200%). The total mean of the VOG, the ITG and the auto open response rate of the ETG is in week one 71.86% and 69.23% in week six. Therefore, **H6** has to be rejected for the Facebook questionnaire in case the single groups are examined, but can be accepted for the mean of all 30 participants. However, it has to be noticed that the decrease is not significant (T-test: $t(29){=}2.05$, $p{=}0.51{>}0.05$).

The mail questionnaire response rates, on the contrary, increase for the VOG (5.67%) and the ITG (4.02%), while the ETG response rates decrease by 11.60% for all logs and by 5.74% for the auto opened forms. The mean of all 30 participants shows an increase of the response rate from week one (69.98%) to week six (71.30%). For this reason, hypothesis **H6** has to be rejected for the mail questionnaire. However, the t value of the T-test $t(29){=}2.05$, $p{=}0.80{>}0.05$ proves that this increase is not significant.

In summary, the response rates decrease and increase only slightly for the Facebook and the mail questionnaire, respectively, although the reported numbers of missed forms present another trend, as can be seen in Figure 6.20. There, the number of participants who did not fill out every form any more increases from week one to week five and decreases only slightly to week six. This should lead to a linear decrease of the response rate in the first five weeks, which is not visible in the weekly response rates. Although the reported overall response rate for the Facebook and the mail form are confirmed by the logged data, the weekly response rates do not match. Therefore, the self-assessment of the weekly response rates is less credible than the logged data. This can be caused by a wrong retrospection of the single weeks. In case information about the weekly performance are required, logging is preferable to the self-assessment at the end of the study.

7.4. Number of App Usages

The end questionnaire asked the participants whether they changed their app usage during the six-week period of the study. The given ratings can be seen in the columns J and K in Figure 6.17 for the VOG, in Figure 6.18 for the ITG and in Figure 6.19 for the ETG. The voluntary and the interval-triggered groups indicated a change by the majority of the participants. For the ETG, the participants indicated on average that their app usage has changed even more. From the comments given in the end questionnaire, subjects tended to look up new posts and mails on the computer or reduced the usage of the app to situations where they really wanted to use it.

Table 7.1 presents the logged number of app usages (NAU) for both apps and each of the three groups. Neither the NAU of the Facebook nor the mail questionnaire decreases linearly. However, the number of Facebook app usages in week six is 24% lower for the VOG, 35% lower for the ITG and 15% lower for the ETG compared to week one. For the mail app, the NAU in week six is 12% lower for the VOG, 18% lower for the ITG and 31% lower for the ETG. A t-test with the number of app usages of all 30 participants in week one and in week six shows that the NAU

decreases for the Facebook app (T-test: t(29)=2.05, p=0.003<0.05) as well as for the mail app (T-test: t(29)=2.05, p=0.004<0.05) significantly. For this reason, hypothesis **H7** is accepted. The decreases for the Facebook app usage are higher than the ones for the mail app for the voluntary and the interval-triggered group. The event-triggered group has a higher decrease for the number of mail app usages. Overall, one cause for the decrease of the app usages could be the task to answer a questionnaire afterwards. Some subjects could have reduced the NAU in order to avoid having to fill out the forms. Another cause could be that the study was conducted during the summer holidays. Subjects who went on vacation probably used the apps less often or not at all. The last reason is simply a reduced usage of the app due to a lower Facebook or mail activity. Therefore, the hypothesis **H7** is accepted, but the specific reason for the decrease cannot be identified.

Table 7.1.: Number of Facebook and mail app usages during the six-week period. The number of app usages in week one is for both apps and for all groups higher than in week six.

Facebook	week 1	week 2	week 3	week 4	week 5	week 6	SUM
VOG	241	238	180	228	121	182	1190
ITG	234	151	126	125	147	151	934
ETG	237	198	140	137	144	201	1057
Mail	week 1	week 2	week 3	week 4	week 5	week 6	SUM
VOG	227	123	155	152	161	199	1017
ITG	236	219	199	152	186	194	1186
ETG	324	216	259	196	213	223	1431

A fact about the number of app usages that can be stated is that the participants underestimated this number. If the subjects had counted their number of app usages correctly, the bars in Figure 6.13 would all have had a value of 100%. The group response rates, which include the reported number of missed questionnaires, have values between 50 and 70%. These percentages are only 2% to 10% higher than the usual response rates without the reported missed forms. Therefore, the participants underestimate the number of times they have used the Facebook and the mail app and hypothesis **H8** is proven.

7.5. Best Overall Questionnaire Mode

There are three questionnaire modes that are supported by the Questionnaire app: voluntary, interval-triggered and event-triggered. For the conducted study, the interval-triggered group had the best response rate for the Facebook questionnaire, while the event-triggered group was with a percentage of 0.27% slightly better for the mail questionnaire.

The high response rate of the ITG can be explained by the daily notification of a missed form that should have been filled out. In case the daily number of mail app usages is higher than for the Facebook questionnaire, the single notification per day imposes a higher increase of the Facebook response rate than on the more frequently used mail app. According to the number of app usages in Table 7.1, this assumption is correct because the ITG used the mail app 21% more often than the Facebook app.

The VOG, on the contrary, used the Facebook app about 14% more often than the mail app, but the response rates for the mail questionnaire were 7% higher than for the Facebook questionnaire. Therefore, it can be assumed that a higher number of usages leads to a lower response rate because the task to fill out a questionnaire can be forgotten more often and the VOG got no notifications. The VOG may have used the Facebook app more unconsciously than the mail app and forgot the form after the app usage. The same holds for the ETG in case all logs are considered because the mail app has been used 26% more often and the Facebook questionnaire has a higher response rate.

In case only the number of automatically triggered questionnaires instead of all logged app usages is considered, there is a different distribution. Questionnaires for the event-triggered group have only been triggered after every second app usage. As aforementioned, only a switch to the home screen triggers the appropriate form. However, every time an automatic questionnaire appeared, it has almost always been answered. With a response rate of 96.86% for the Facebook and 91.51% for the mail form (see Figure 6.12), this questionnaire mode is the preferable one. If more than every second app usage should have been logged, it is not clear how high the response rate will stay. Looking at the answers given in the end questionnaire, the event-triggered mode had the most negative comments. Therefore, hypothesis **H9**, which states that the event-triggered group is the most burdensome, can be accepted. Although three voluntary group members would have liked the app to automatically start the questionnaires, four members of the ETG criticized the automatically triggered forms to be annoying. If the frequency of forms was higher, the response rate might be lower. Therefore, the conductor of the study has to decide whether he wants to have forms answered for every app usage or if he wants to have a high response rate for the sample number of times the form has been triggered.

7.6. Recommendable Study Length

In addition to the best overall questionnaire mode, the three questionnaire modes can be analyzed for the best study lengths in the different contexts. Figure 6.14 and Figure 6.15 can be considered for this question.

For the Facebook questionnaire, an ideal study length for the voluntary and the event-triggered group proved to be three to four weeks. The voluntary group had lower response rates in the middle, which could be eliminated by reducing the number of weeks. The response rate for the

event-triggered group fell strongly in the third and sixth week. A length of three to four weeks could reduce the drops and could lead to a higher overall response rate. In case the frequency of automatically triggered questionnaires should be higher than the one for this study, a length of two weeks could be ideal. The interval-triggered group had the most consistent response rates for the Facebook questionnaire. The drop of the response rate in week six leads to a recommendable study length of five weeks.

The response rates for the mail questionnaire also result in a recommendable length of three to four weeks for the VOG in order to avoid the lower response rates in the middle of the study. For the interval-triggered and the event-triggered group, a length of four weeks is recommendable. The response rate of the mail questionnaire for the interval-triggered drops, similar to the event-triggered group, after week four. If a higher frequency of automatically triggered questionnaires is required, a recommendable study length for the ETG is one week.

In comparison, the recommendable lengths for a study of the Facebook app usage, which represents an activity of the spare time, are one week longer than for the mail app usage, which is an example of an activity during work time.

Chapter 8.

Long-Term Effects Study

The analysis of the main study results showed that a lot of participants changed their Facebook and mail app usage behavior during the study. This change of behavior is especially interesting with regard to the long-term effects after the six-week study. Therefore, the participants were asked to complete an online questionnaire on October 4, 2012, one month after the end of the main study. The questionnaire was not announced at the end of the study in order to not bias the long-term effects. The six questions on the long-term effects were the following:

A. I use the Facebook app compared to - a) before the study, b) during the study
   (1) much less, (2) less, (3) just as often, (4) more, (5) much more.

B. I use the mail app compared to - a) before the study, b) during the study
   (1) much less, (2) less, (3) just as often, (4) more, (5) much more.

C. Would you participate again in a similar study?

D. For how long would you install the app again? I would install the app again for ... weeks.

E. What have you learned about yourself in the course of the study? Enter new insights about yourself, about the participation in the study, on the use of the Questionnaire app, on the use of the Facebook and the mail app, etc.

F. Would you recommend your questionnaire mode also for other types of studies? Can you think of an example? For how long would you recommend the proposed questionnaire method?

The participants were able to answer the questionnaire until October 7, 2012. Within this period of time, 23 from the 30 participants filled out the form. From the 23 participants, nine were members of the voluntary group, seven were from the interval-triggered group and seven were from the event-triggered group.

8.1. Changed App Usage

The first two questions, questions A and B, asked the participants to estimate their current Facebook and mail app usage in comparison to before and during the study. The given answers are presented in Figure 8.1 as box-and-whisker plots. It should be noted that these questions were only answered by 22 of the 23 participants.

(a) Facebook   (b) Mail
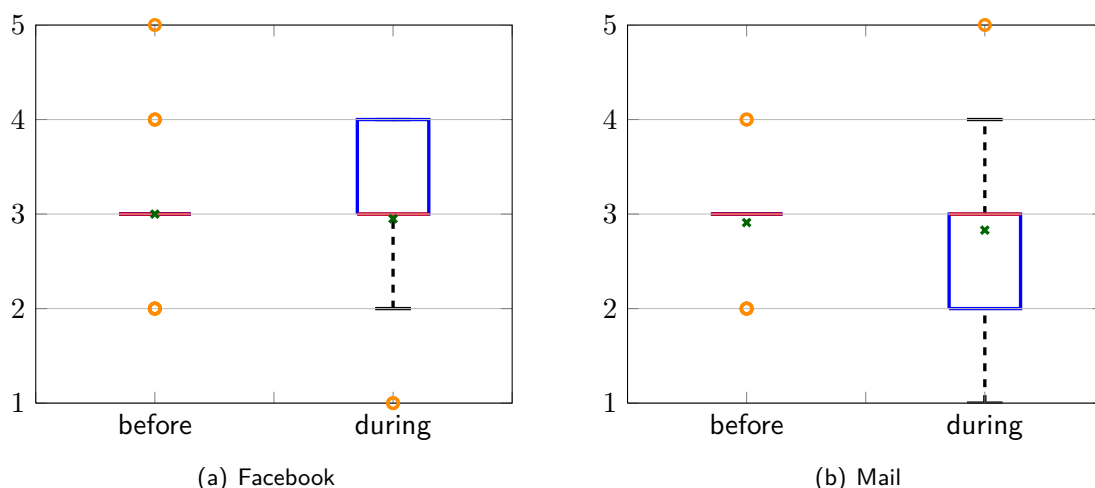
Figure 8.1.: Results for the statements A and B of the long-term effects questionnaire. The usage of both apps after the study was on average the same as before the study. For the Facebook app, the current usage is the same as during the study, while for the mail app it is surprisingly slightly less often. Definition of symbols:
——: median, **x**: average, **o**: outlier, —— upper: third quartile, —— lower: first quartile, —— upper: maximum value, —— lower: minimum value

As can be seen in the left bars of both plots in Figure 8.1, the current Facebook and mail app usage is on average the same as it was before the study. For the Facebook app, 13 participants reported that they started the app with the same frequency as before the study, while five participants used it less often. The remaining four participants used the Facebook app currently more often and much more often than before. The mail app was used by seventeen subjects with the same frequency. Four participants used it less and two used it more often. On average, the current Facebook and mail app usage was the same as before the study.

Another trend can be seen for the comparison current usage to usage during the study. Figure 8.1(a) reveals a upper quartile for "more usage". Six participants stated that they used the Facebook app more often after the study had ended. One of them, TS 26, wrote that he "open[s] Facebook often, even though [he] do[es] not want to ([he opens it] unconsciously because [he] is bored)". With an average of 2.95, the subjects indicated the same usage, while the minimum was at the rating "less usage".

In comparison to the Facebook app, the mail app was rather used less often after the study than

during the study (see Figure 8.1(b)). The minimum rating chosen by one subject was "much less" and the rating "less" was with seven agreeing subjects the lower quartile. The median and the average of 2.83 indicated that the participants used the app currently as often as during the study. Three participants used it more often and one even much more often. TS 27, who indicated a less frequently usage, noticed: "I learned that I checked my mails surprisingly often. I have begun to check my mails considerably less often since the [additional] appearance of the [Questionnare] app [after the mail app] has lasted considerably longer as only looking up [the mails] shortly [with no form appearing afterwards].".

General comments about the usage of the apps were made in question E. TS 3 stated that he "used both apps during the study more consciously. Before, [he] often used the Facebook app shortly. During the study, [he was] more conscious about the usage as well as the frequency.". TS 1, TS 4, TS 7, TS 22, TS 23 and TS 29 made similar comments about the more conscious usage of the Facebook or the mail app. TS 1 also added that he "used the Facebook and the mail app more consciously. [...] However, that was only because [he] knew that [he] should fill out a questionnaire afterwards, now [he] click[s] on it again just because [he] ha[s] nothing better to do.". TS 18 was more conscious about the duration of the app usage and TS 9 assumed that the DAU is often higher than the self-reported usage duration. Considering the logged DAU of TS 9, his self-assessment is rather false because his self-reported DAU for the Facebook as well as the mail app were on average higher than the logged ones.

TS 2, TS 21 and TS 24 indicated that they either switched to the PC in order to look up new posts and mails or that they reduced the usage. TS 24 wrote: "I have often read only the mail notifications during the study. [After the study], I have maintained it for a few days. But then I fell back on the instant reading. The [number of] usage[s] of the Facebook app has felt less in total. In general, I feel that I am 'wasting' less time with my smartphone since the study. But it could also be due to something else, such as the change of my mobile phone during the study.". TS 13 added, that during the study "[...] often [he] realize[d] that [starting the app was] not necessary and you [start the app] 'because you just want to look [it up] again' and you are used to [do] it". A comment from TS 6 shows not only an increase of awareness for app durations and frequencies, but that he also realized "how dependent [he is] on [his] smartphone".

In summary, one can say that the study changed some participants' behavior with the apps in a more positive way because they are more conscious about the app usage and how much time they spend on it.

8.2. Renewed Participation in a Study and Reinstallation of the Questionnaire App

Question C asked the participants, whether they would take part in a similar study. From the 22 participants, which answered the questionnaire, 19 stated to be open for another participation. The remaining four subjects did not want to participate again. This concludes in a percentage of 82.6% of test subjects, which would participate again in a similar study.

The next question asked the participants, for how long they would reinstall the Questionnaire app. Figure 8.2 presents the given values.
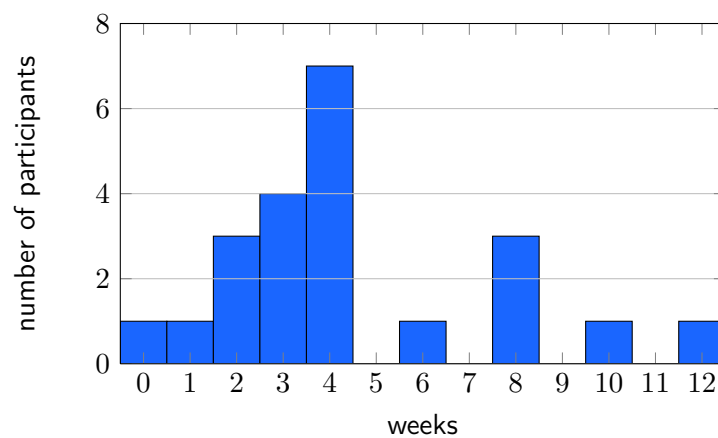
Figure 8.2.: Frequency distribution for the statement D of the long-term effects questionnaire. The preferred length for a study was four weeks and the maximum of weeks were 12.

With a maximum of seven from the 22 participants, about one third of all participants would reinstall the Questionnaire app for four weeks. Four participants would agree to three weeks and three to two weeks. Thereby, 14 of the 22 subjects indicated two to four weeks as an acceptable reinstallation period. A shorter period was declared by one person with zero weeks and another one with one week. A longer period of six, ten and even twelve weeks were reported by one person each. A period of eight weeks was indicated by three people.

In summary, the maximum acceptable length for having the Questionnaire app installed on the phone is 12 weeks. It has to be noted that the participants probably made these specifications with regard to the effort during the six-week main study. The seven members from the ETG, which answered this question, stated to reinstall the Questionnaire app again for zero, two or three weeks, respectively. The seven ITG members chose one week, four or six weeks and the nine participants of the VOG would agree from two up to twelve weeks. This distribution shows that the recommendations for the length of a study in Section 7.6 were made relatively correct.

8.3. General Comments on the Questionnaire Mode

The last question asked the participants whether they would recommend the questionnaire mode with which they were asked for other studies. Additionally, they were requested to give an example and specify an acceptable study length.

From the voluntary group, TS 7 would recommend the voluntary feedback diary for a length of two to three months. However, TS 2 would not recommend this mode "because it requires a lot of personal initiative". He indicated a maximum of two weeks as appropriate. A comment from VOG member TS 4 was that he wanted the event-triggered automatic appearing of the questionnaire. Furthermore, he liked the Facebook and the mail questionnaire due to their shortness. A recommendation from him would be to use the app for location-based services with a length of four to six weeks. Location-based services enable to present, for example, points of interest or advertisements to the user by identifying his location.

TS 11 would also prefer an automatic opening of the questionnaire instead of his interval-triggered notifications. Three other members of the ITG, TS 13, TS 16 and TS 18, liked the daily notifications due to being "not that time-consuming" (TS 13). An acceptable study length would be a maximum of four weeks for TS 16 and TS 18. While TS 16 stated that the interval-triggered questionnaire mode could be used for any media consumption, TS 18 explained that there already are a lot of studies, which use this mode.

Although some of the members from the other two groups would have preferred the event-triggered mode, TS 21 and TS 26 found the automatic appearance of questionnaires "disturbing" because it "disrupts [the] flow of use" (TS 26). For being interrupted, TS 21 suggested a maximum of two to three weeks as study length. Contrary to these two participants, TS 29 liked the event-triggered mode due to being "easy and fast". He recommended a period of four weeks for a study with the event-triggered mode.

There were also suggestions for further studies that could be conducted with the Questionnaire app or the specific questionnaire mode. TS 1 wrote that it could be useful for other studies in the smartphone or app usage context because there would be no break from using the phone to answering a questionnaire. TS 3 recommended a usage in the field of medicine. He says, that the electronic device makes it easier to "send the results from, for example, a diary about the symptoms of allergies or headaches, directly to the doctor, which probably accelerates the diagnosis and simplifies the analysis.". TS 22 and TS 23 recommended a usage in exercise or nutrition contexts. A "forced 'monitoring' of eating behavior" and a daily input for the duration of an applied sport seems appealing for TS 23 due to the "obvious 'advantages' a participant has" from making entries. The suggestion of these various research fields shows that the already conducted studies, which were mentioned in section 2.2, are interesting to the users.

Finally, a statement made by TS 24 offers a reasonable usage of the event-triggered mode. He was at some point in time annoyed by the automatic appearance of the questionnaire. Therefore, he

suggests to use this mode for "'mobile-phone-addicts' or 'internet-addicts'[, because] it reminds you of another 'automatic' opening of, for example, [the] Facebook [app]". Another advantage for TS 24 is that the appearing questionnaires become annoying and they probably reduce the demand for the usage of an app. For these reasons, the event-triggered mode with a high frequency of automatically opened questionnaires after an app usage could help addicts to be more aware of their usage and reduce it.

Chapter 9.

Conclusions

The goal of this diploma thesis was to set up some common guidelines for the research with smartphones. The research technique of using a smartphone is relatively new. Therefore, practical experiences have to be gained for an optimal execution. Already existing studies with smartphones have not stated general rules, but only use this research technique as a tool. In order to investigate common guidelines, a research app is needed. The research app should include different research modes and should be designed in Android. Android is the preferable operating system as it has the greatest market share and it is subsequently easier to get a high number of participants. Already existing research apps lack features like the multiple questionnaire modes or are commercial solutions that have to be paid for per use.

Therefore, a research app named Questionnaire app was implemented in Android. The Questionnaire app offers the possibility to either conduct a survey with questionnaires or a diary study. There are three different modes: First, a diary entry or questionnaire can be answered voluntarily by the participant if something relevant to the studied object had happened. The second mode triggers notifications in pre-defined intervals if an entry has been missed to be made. The notifications should also improve the response rates as they are a reminder that the study is still running. A third mode is the event-triggered mode. It automatically opens the questionnaire after a defined event, like the closing of an app. The app enables the conductor of the study to decide which group of test subjects should use which one of these three modes and which questionnaires. In order to easily create, edit and delete questionnaires, a tool named QuestionnaireAdmin was implemented. The questionnaires as well as the automatically uploaded answers given by the participants are stored on the backend server. The server additionally offers DatabaseAdmin and the Evaluation framework, which enable an easy way to view the answers of the participants.

A six-week study has been conducted with the Questionnaire app that requested the participants to fill out a questionnaire whenever they used the Facebook or one of their mail apps. The 30 participants were distributed equally into three groups, each of them having one of the three different questionnaire modes.

Results proved the event-triggered mode with its automatically triggered forms to have the highest response rates for both questionnaires. Almost 100% of the event-triggered forms have been

filled out. However, the questionnaires appeared only half of the time an app was used and even that rate caused some of the event-triggered group members to criticize the mode as annoying. For these reasons, researcher have two options: decrease the study length for a higher frequency or lower the frequency for a longer study length. In case the response rates for all logged app usages are analyzed, the interval-triggered group had the best response rates. When comparing the response rates for the two apps, it can be seen that participants answered more forms for the Facebook app.

The weekly response rates led to the conclusion that a study can last one week longer for the Facebook app than for the mail app. For the voluntary mode a study length of three to four weeks is recommendable for both app types. For the interval-triggered mode a five-week study for a spare time app and a four-week study for a work time app has to be preferred. The event-triggered mode is advisable for a four-week study in case the frequency is similar to the one used in this study. If a higher frequency is required in order to gain more information about the studied object, the usage of a spare time app can be questioned for two weeks and a work time app for one week. During a study, a combination of self-reports and logging is helpful for the interpretation of the data. Although the participants indicate that they are good in self-assessing, the average duration of app usages was overestimated and the number of apps used was underestimated.

Future studies could use the Questionnaire app to investigate the weekly response rates in case there is no pre-defined length of the study. For the study conducted within this work, the participants have been told that the study would end after six weeks. This information led to some increase of the response rates at the end of the six weeks. Test subjects may have felt some guilt for not participating as much as they could have done and wanted to compensate for this at the end. It would be interesting how the response rates would have been if no end date had been specified to the users. Another interesting aspect to be investigated could be the influence of the reminder emails. In the weekly response rates of this study, an influence can be seen, but not how big this effect is. Therefore, different groups could get a different number of reminder emails or the same number of emails at different points of time. A third future research field could be the frequency of automatically triggered questionnaires for the event-triggered mode. It can only be assumed that a higher frequency of events would decrease the response rate, but not to what extent. Additionally, the ideal lengths for studies with different event frequencies could be examined.

The Questionnaire app could also be useful for other research fields. Two possible research fields are, for example, the evaluation of MobiDics, a mobile didactic platform for university teaching staff [42, 43], or research in the large [44–46].

In conclusion, the implemented Questionnaire app offers the functionality to conduct studies with smartphones in order to set up common guidelines and to research objects and behaviors. The study conducted with the Questionnaire app tested hypotheses that are of fundamental importance for the research with smartphones.

List of Figures

List of Tables

List of Acronyms

app   application

ALD   average logged duration

APK   application package file

ARD   average reported duration

DAU   duration of app usage

ETG   event-triggered group

ITG   interval-triggered group

NAU   number of app usages

TS   test subject

TUM   Technische Universität München

VMI   Fachgebiet Verteilte Multimodale Informationsverarbeitung

VOG   voluntary group

Bibliography

[1] Jonathan Lazar, Jinjuan Heidi Feng, and Harry Hochheiser. *Research methods in human-computer interaction*. John Wiley & Sons Ltd, Chichester, 2010.

[2] Rainer Schnell, Paul B. Hill, and Elke Esser. *Methoden der empirischen Sozialforschung*. Oldenbourg Wissenschaftsverlag GmbH, 2008.

[3] John Krumm. *Ubiquitous computing fundamentals*. Chapman & Hall / CRC, Boca Raton, Florida, 2009.

[4] Jakob Nielsen. *Usability engineering*. Morgan Kaufmann, 2007.

[5] Jon Froehlich, Mike Y. Chen, Sunny Consolvo, Beverly Harrison, and James A. Landay. MyExperience: a system for in situ tracing and capturing of user feedback on mobile phones. In *Proceedings of the 5th international conference on Mobile systems, applications and services*, pages 57–70, San Juan, Puerto Rico, 2007. ACM.

[6] Ben Shneiderman and Catherine Plaisant. *Designing the user interface. Strategies for effective human-computer interaction*. Addison-Wesley, 2009.

[7] Scott Carter and Jennifer Mankoff. When participants do the capturing: the role of media in diary studies. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 899–908, Portland, Oregon, USA, 2005. ACM.

[8] Leysia Palen and Marylin Salzman. Voice-mail diary studies for naturalistic data capture under mobile conditions. In *Proceedings of the 2002 ACM conference on Computer supported cooperative work*, pages 87–95, New Orleans, Louisiana, USA, 2002. ACM.

[9] Sami Hulkko, Tuuli Mattelmäki, Katja Virtanen, and Turkka Keinonen. Mobile probes. In *Proceedings of the third Nordic conference on Human-computer interaction*, pages 43–51, Tampere, Finland, 2004. ACM.

[10] Sunny Consolvo, Beverly Harrison, Ian Smith, Mike Chen, Katherine Everitt, Jon Froehlich, and James A. Landay. Conducting in situ evaluations for and with ubiquitous computing technologies. *International Journal of Human-Computer Interaction*, 22(1-2):103–118, April 2007.

[11] Lisa Feldman Barrett and Daniel J. Barrett. An introduction to computerized experience sampling in psychology. *Social Science Computer Review*, 19(2):175–185, May 2001.

[12] Arthur A. Stone and Saul Shiffman. Capturing momentary, self-report data: A proposal for reporting guidelines. *Annals of Behavioral Medicine*, 24(3):236–243, August 2002.

[13] Niall Bolger, Angelina Davis, and Eshkol Rafaeli. Diary methods: Capturing life as it is lived. *Annual review of psychology*, 54:579–616, January 2003.

[14] Jennifer Mankoff and Scott Carter. Crossing qualitative and quantitative evaluation in the domain of ubiquitous computing. In *CHI2005 workshop "Usage analysis: Combining logging and qualitative methods*, 2005.

[15] Bernard J. Jansen, Isak Taksa, and Amanda Spink. *Handbook of Research on Web Log Analysis*. IGI Global, October 2009.

[16] Frank P. Deane, John Podd, and Ron D. Henderson. Relationship between self-report and log data estimates of information system usage. *Computers in Human Behavior*, 14(4):621–636, December 1998.

[17] Kasper Lovborg Jensen, Christian Fischer Pedersen, and Lars Bo Larsen. Diasnet mobile: A personalized mobile diabetes management and advisory service. In *Second Workshop on Personalization for E-Health*, number June, Corfu,Greece, 2007.

[18] Jakob E. Bardram, Mads Frost, Károly Szántó, and Gabriela Marcu. The MONARCA self-assessment system: a persuasive personal monitoring system for bipolar patients. In *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium*, pages 21–30, Miami, Florida, USA, 2012. ACM.

[19] Robert N. Jamison, Stephen A. Raymond, Jonathan G. Levine, Ellen A. Slawsby, Srdjan S. Nedeljkovic, and Nathaniel P. Katz. Electronic diaries for monitoring chronic pain: 1-year validation study. *Pain*, 91(3):277–85, April 2001.

[20] Katie A. Siek, Kay H. Connelly, Yvonne Rogers, Paul Rohwer, Desiree Lambert, and Janet L. Welch. When do we eat? An evaluation of food items input into an electronic food monitoring application. In *Pervasive Health Conference and Workshops, 2006*, pages 1–10. IEEE Computer Society, November 2006.

[21] Zahid Hussain, Martin Lechner, Harald Milchrahm, Sara Shahzad, Wolfgang Slany, Martin Umgeher, and Peter Wolkerstorfer. Concept and design of a contextual mobile multimedia content usability study. In *Proceedings of the 2009 Second International Conferences on Advances in Computer-Human Interactions*, pages 277–282. IEEE Computer Society, February 2009.

[22] Zahid Hussain and Wolfgang Slany. Analyzing real mobile web usage of a multimedia stream-

ing application through log files. In *Proceedings of the international conference on Multimedia information retrieval*, pages 401–404, Philadelphia, Pennsylvania, USA, 2010. ACM.

[23] Nathan Eagle, Alex (Sandy) Pentland, and David Lazer. Inferring friendship network structure by using mobile phone data. *Proceedings of the National Academy of Sciences*, 106(36):15274–15278, September 2009.

[24] Atsunori Minamikawa, Nobuhide Kotsuka, Masaru Honjo, Daisuke Morikawa, Satoshi Nishiyama, and Masayoshi Ohasi. RFID supplement for mobile-based life log system. In *International Symposium on Applications and the Internet Workshops, 2007*, pages 20–23. IEEE Computer Society, 2007.

[25] Tuula Kärkkäinen, Tuomas Vaittinen, and Kaisa Väänänen-Vainio-Mattila. I don't mind being logged, but want to remain in control: a field study of mobile activity and context logging. In *Proceedings of the 28th international conference on Human factors in computing systems*, pages 163–172, Atlanta, Georgia, USA, 2010. ACM.

[26] Thorsten Bohnenberger, Anthony Jameson, Antonio Krüger, and Andreas Butz. Location-aware shopping assistance: evaluation of a decision-theoretic approach. In *Human Computer Interaction with Mobile Devices*, pages 48–48. Springer Berlin / Heidelberg, 2002.

[27] Enrico Rukzio. *Physical mobile interactions: mobile devices as pervasive mediators for interactions with the real world*. PhD thesis, University of Munich, 2006.

[28] J. Maitland, S. Sherwood, L. Barkhuus, I. Anderson, M. Hall, B. Brown, M. Chalmers, and H. Muller. Increasing the awareness of daily activity levels with pervasive computing. In *Pervasive Health Conference and Workshops*, pages 1–9, 2006.

[29] Ning Liu, Ying Liu, and Xia Wang. Data logging plus e-diary: towards an online evaluation approach of mobile service field trial. In *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services*, pages 287–290, Lisbon, Portugal, 2010. ACM.

[30] Minna Isomursu, Kari Kuutti, and Soili Väinämö. Experience clip: method for user participation and evaluation of mobile concepts. In *Proceedings of the eighth conference on Participatory design: Artful integration: interweaving media, materials and practices - Volume 1*, pages 83–92, Toronto, Ontario, Canada, 2004. ACM.

[31] Louise Barkhuus and Valerie E. Polichar. Empowerment through seamfulness: smart phones in everyday life. *Personal and Ubiquitous Computing*, 15(6):629–639, December 2010.

[32] Myoung Hoon Jeon, Dae Yol Na, Jung Hee Ahn, and Ji Young Hong. User segmentation & UI optimization through mobile phone log analysis. In *Proceedings of the 10th international*

conference on Human computer interaction with mobile devices and services - MobileHCI '08, pages 495–496, Amsterdam, The Netherlands, 2008. ACM Press.

[33] Rachel Demumieux and Patrick Losquin. Gather customer's real usage on mobile phones. In *Proceedings of the 7th international conference on Human computer interaction with mobile devices & services*, pages 267–270, Salzburg, Austria, 2005. ACM.

[34] Hossein Falaki, Ratul Mahajan, Srikanth Kandula, Dimitrios Lymberopoulos, Ramesh Govindan, and Deborah Estrin. Diversity in smartphone usage. In *Proceedings of the 8th international conference on Mobile systems, applications, and services*, pages 179–194, New York, NY, USA, 2010. ACM.

[35] Matthias Böhmer, Brent Hecht, Johannes Schöning, Antonio Krüger, and Gernot Bauer. Falling asleep with Angry Birds, Facebook and Kindle: a large scale study on mobile application usage. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*, pages 47–56, Stockholm, Sweden, 2011. ACM.

[36] Timothy Sohn, Kevin A. Li, William G. Griswold, and James D. Hollan. A diary study of mobile information needs. In *Proceedings of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, pages 433–442, Florence, Italy, 2008. ACM.

[37] Stina Nylander and Teres Lundquist. At home and with computer access: why and where people use cell phones to access the internet. In *Proceedings of the 27th international conference on Human factors in computing systems*, pages 1639–1642, Boston, MA, USA, 2009. ACM.

[38] Kenton O'Hara, April Slayden Mitchell, and Alex Vorbau. Consuming video on mobile devices. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 857–866, San Jose, California, USA, 2007. ACM.

[39] Michael. R. Hufford and Alan L. Shields. Electronic diaries: Applications and what works in the field. *Applied Clinical Trials*, (April):46–59, 2002.

[40] Scott Carter, Jennifer Mankoff, and Jeffrey Heer. Momento: support for situated ubicomp experimentation. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 125–134, San Jose, California, USA, 2007. ACM.

[41] Mika Raento, Antti Oulasvirta, Renaud Petit, and Hannu Toivonen. ContextPhone: a prototyping platform for context-aware mobile applications. *Pervasive Computing, IEEE*, 4(2):51–59, April 2005.

[42] Andreas Möller, Angelika Thielsch, Beate Dallmeier, Andreas Hendrich, Barbara E Meyer, Luis Roalter, Stefan Diewald, and Matthias Kranz. MobiDics - Eine mobile Didaktik-Toolbox

für die universitäre Lehre. In *DeLFI 2011 - Die 9. e-Learning Fachtagung Informatik der Gesellschaft für Informatik e.V.*, pages 139–150, Dresden, Germany, 2011.

[43] Andreas Möller, Angelika Thielsch, Beate Dallmeier, Luis Roalter, Stefan Diewald, Andreas Hendrich, Barbara E. Meyer, and Matthias Kranz. MobiDics - Improving University Education With A Mobile Didactics Toolbox. In *Ninth International Conference on Pervasive Computing (Pervasive 2011), Video Proceedings*, pages 1–4, San Francisco, CA, USA, 2011.

[44] Matthias Kranz, Lukas Murmann, and Florian Michahelles. Research in the Large: Challenges for Large-Scale Mobile Application research - A Case Study about NFC Adoption using Gamification via an App Store. *IJMHCI*, 2013.

[45] Lukas Murmann, Florian Michahelles, and Matthias Kranz. NFC Heroes - Observing NFC Adoption through a Mobile Trading Card Game. In *Proceedings of the 3rd International Workshop on Research in the Large. Held in Conjunction with Mobile HCI*, pages 19–22, San Francisco, USA, 2012.

[46] Andreas Möller, Florian Michahelles, Stefan Diewald, Luis Roalter, and Matthias Kranz. Update Behavior in App Markets and Security Implications : A Case Study in Google Play. In *Proceedings of the 3rd International Workshop on Research in the Large. Held in Conjunction with Mobile HCI*, pages 3–6, San Francisco, USA, 2012.