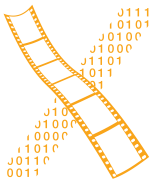

Chair for Computer Science 10 (Media Computing and Human-Computer Interaction)

RWTH AACHEN UNIVERSITY

Participants, Incentives, and User Studies: A Survey of CHI 2019

Bachelor's Thesis
submitted to the
Media Computing Group
Prof. Dr. Jan Borchers
Computer Science Department
RWTH Aachen University

by
Jenny Reinhard

Thesis advisor:
Prof. Dr. Jan Borchers

Second examiner:
Prof. Dr. Christian Remy

Registration date: 14.05.2020
Submission date: 14.09.2020

Zentrales Prüfungsamt/Central Examination Office

Eidesstattliche Versicherung
Statutory Declaration in Lieu of an Oath

Name, Vorname/Last Name, First Name

Matrikelnummer (freiwillige Angabe)
Matriculation No. (optional)

Ich versichere hiermit an Eides Statt, dass ich die vorliegende Arbeit/Bachelorarbeit/ Masterarbeit* mit dem Titel
I hereby declare in lieu of an oath that I have completed the present paper/Bachelor thesis/Master thesis* entitled

selbstständig und ohne unzulässige fremde Hilfe (insbes. akademisches Ghostwriting) erbracht habe. Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel benutzt. Für den Fall, dass die Arbeit zusätzlich auf einem Datenträger eingereicht wird, erkläre ich, dass die schriftliche und die elektronische Form vollständig übereinstimmen. Die Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.
independently and without illegitimate assistance from third parties (such as academic ghostwriters). I have used no other than the specified sources and aids. In case that the thesis is additionally submitted in an electronic format, I declare that the written and electronic versions are fully identical. The thesis has not been submitted to any examination body in this, or similar, form.

Ort, Datum/City, Date

Unterschrift/Signature

*Nichtzutreffendes bitte streichen

*Please delete as appropriate

Belehrung:
Official Notification:

§ 156 StGB: Falsche Versicherung an Eides Statt
Wer vor einer zur Abnahme einer Versicherung an Eides Statt zuständigen Behörde eine solche Versicherung falsch abgibt oder unter Berufung auf eine solche Versicherung falsch aussagt, wird mit Freiheitsstrafe bis zu drei Jahren oder mit Geldstrafe bestraft.
Para. 156 StGB (German Criminal Code): False Statutory Declarations
Whoever before a public authority competent to administer statutory declarations falsely makes such a declaration or falsely testifies while referring to such a declaration shall be liable to imprisonment not exceeding three years or a fine.
§ 161 StGB: Fahrlässiger Falscheid; fahrlässige falsche Versicherung an Eides Statt
(1) Wenn eine der in den §§ 154 bis 156 bezeichneten Handlungen aus Fahrlässigkeit begangen worden ist, so tritt Freiheitsstrafe bis zu einem Jahr oder Geldstrafe ein.
(2) Straflosigkeit tritt ein, wenn der Täter die falsche Angabe rechtzeitig berichtigt. Die Vorschriften des § 158 Abs. 2 und 3 gelten entsprechend.
Para. 161 StGB (German Criminal Code): False Statutory Declarations Due to Negligence
(1) If a person commits one of the offences listed in sections 154 through 156 negligently the penalty shall be imprisonment not exceeding one year or a fine.
(2) The offender shall be exempt from liability if he or she corrects their false testimony in time. The provisions of section 158 (2) and (3) shall apply accordingly.

Die vorstehende Belehrung habe ich zur Kenntnis genommen:
I have read and understood the above official notification:

Ort, Datum/City, Date

Unterschrift/Signature

Contents

List of Figures

List of Tables

Abstract

Research in Human-Computer-Interaction (HCI) mostly relies on user studies. Regarding CHI 2019, user studies are included in 88 % of all papers. One of the biggest challenges of this research approach is the successful recruitment of participants. While there are many factors influencing the efficiency of the recruitment process, the focus of this thesis is on the offer of incentives.

All papers of CHI 2019 were analyzed in order to give an overview about current practices in the HCI community regarding incentives in user studies. It was found that compensation differs depending on the analysis approach, the duration of the study, the number of participants and components as well as on the recruitment medium, the topic, the method, the study and recording type. Of these, the recruitment medium and the duration have proven to be two of the main influencing factors, whereby participants were most frequently compensated when recruited via Amazon Mechanical Turk and received higher amounts for studies with longer durations.

The results might be beneficial to researchers planning studies in the field of HCI. Findings can be used as an orientation and as a basis for a critical assessment of current practices. In addition, the source code of the analysis can be reused to investigate other conferences and compare the results to the findings of this thesis.

Überblick

Die Forschung im Bereich der Mensch-Computer-Interaktion (HCI) beruht hauptsächlich auf Benutzerstudien. In der CHI 2019 sind Benutzerstudien in 88 % der Publikationen enthalten. Eine der größten Herausforderungen dieses Forschungsansatzes ist die erfolgreiche Rekrutierung der Teilnehmer. Während es viele Faktoren gibt, welche die Effizienz des Rekrutierungsprozesses beeinflussen, liegt der Fokus dieser Arbeit auf dem Angebot von Vergütungen.

Alle Publikationen der CHI 2019 wurden analysiert, um einen Überblick über die aktuellen Praktiken in der HCI-Gemeinschaft hinsichtlich der Vergütungen in Benutzerstudien zu erhalten. Es wurde festgestellt, dass die Vergütung je nach Analyseansatz, Dauer der Studie, Teilnehmer- und Komponentenzahl sowie je nach Rekrutierungsmedium, Thema, Methode, Studien- und Aufzeichnungsart unterschiedlich ausfällt. Von diesen Faktoren haben sich das Rekrutierungsmedium und die Dauer der Studie als zwei der wichtigsten Einflussfaktoren herausgestellt, wobei Teilnehmer, die über Amazon Mechanical Turk rekrutiert wurden, am häufigsten eine Vergütung erhielten und solche, die an längeren Studien teilnahmen, mit einem höheren Betrag vergütet wurden.

Die gewonnenen Ergebnisse können für Forscher im Bereich HCI bei der Planung zukünftiger Studien von Nutzen sein. Sie können sowohl als Orientierung als auch als Grundlage für eine kritische Beurteilung aktueller Praktiken dienen. Des Weiteren kann der Quellcode der Analyse wiederverwendet werden, um andere Konferenzen zu untersuchen und die Ergebnisse mit den Erkenntnissen dieser Arbeit zu vergleichen.

Acknowledgements

First, I would like to thank Krishna Subramanian, my supervisor, for his guidance, feedback and advice throughout the course of my thesis.

I would also like to thank Prof. Dr. Jan Borchers, my thesis advisor and Prof. Dr. Christian Remy, my second examiner, for their time and support.

Finally, I would like to thank Paul-Joachim Niehoff, my fiancé, for always being supportive and patient throughout this time.

Conventions

Throughout this thesis we use the following conventions.

Text conventions

Definitions of technical terms or short excursus are set off in coloured boxes.

EXCURSUS:
Excursus are detailed discussions of a particular point in a book, usually in an appendix, or digressions in a written text.

Definition:
Excursus

The whole thesis is written in American English.

Download links are set off in coloured boxes.

File: myFile[a]

[a] http://hci.rwth-aachen.de/public/SurveyCHI2019/file_number.file

Chapter 1

Introduction

Research in HCI relies on user studies

Since Human-Computer-Interaction (HCI) deals with the behavior of humans interacting with technology [Lopes, 2016], research in this field mostly relies on user studies [Sohail, 2020]. Regarding the *Conference on Human Factors in Computing Systems* (CHI), the premier international conference of HCI, user studies are included in 88 % of all papers in 2019. In CHI 2014, the percentage is even higher. 91 % of all manuscripts contain at least one user study [Caine, 2016].

The efficiency of recruitment is determined by several factors

A major challenge of this research approach is the successful recruitment of participants. Researchers often have to deal with low response rates or the lack of participant diversity [Barkhuus and Rode, 2007]. In general, the efficiency of the recruitment process is determined by many factors. The relationship from the study director to the participants and the recruitment medium as well as the study topic, tasks and duration are key aspects that affect user participation [Christensen et al., 2017, Sohail, 2020]. For example, prior research has shown that it is possible to reach an increase of 25 % in the number of completed interviews by decreasing the announced interview time from 20 to 15 minutes. [Hansen, 2007].

Incentives in user studies

In addition to the factors mentioned, incentives also play an important role in recruiting [Sohail, 2020]. They can influence the rate of participation [Fiore et al., 2014] or be used to

target desired participants [Hsieh and Kocielnik, 2016]. But incentives are not offered in every user study. For example, in CHI 2019, offer of compensation is indicated in only 36 % of all studies. Furthermore, there is a high variance in the amount of compensation regarding studies where participants were incentivized [Latterman and Merz, 2001]. So how do researchers decide whether to offer incentives or not? And how do they determine the amount participants should receive? Are there any characteristics of the study or the participants that influence the offer of compensation?

Aim and contributions of the thesis

The aim of this thesis is to give an overview about current practices in the HCI community regarding incentives in user studies. For this purpose, all papers of CHI 2019 [1] were analyzed. Differences in the frequency and amount of compensation between the studies are highlighted and an attempt is made to find explanations for these differences. The findings might be useful for researchers in the field of HCI because they could benefit from these results in future studies. On the one hand, they could learn from the findings and use them as an orientation. On the other hand, the findings could be assessed critically. Researchers may wonder if they want to stick to current practices or if there are good reasons to consciously deviate from them. Another contribution of this work is the reusable source code of the analysis.

File: Analysis CHI 2019[a]

[a] http://hci.rwth-aachen.de/public/SurveyCHI2019/Analysis CHI 2019.ipynb

It can be used to investigate other conferences and compare the results with the findings of this thesis. This way, it might be possible to reveal trends over time or to identify differences to other fields of research.

[1] https://dl.acm.org/doi/proceedings/10.1145/3290605

Chapter 2

Related Work

Since this work is not the first to survey CHI publications, a brief overview of several prior findings is given in this chapter. Afterwards, the recruitment process in user studies is discussed, focusing on the recruitment medium and the role and offer of incentives.

2.1 Survey of CHI Conferences

How user studies in CHI changed over time

The analysis of CHI papers published between 1983 and 2006 reveals an increase in qualitative studies over a 24 years' time frame. In addition, it was found that more recent studies often last longer and use multiple types of methods [Barkhuus and Rode, 2007]. By extracting and clustering keywords of CHI papers published between 1994 and 2013, Liu et al. showed that topics of HCI research have shifted towards mobile interaction, while new issues such as crowdsourcing and privacy have involved. In contrast, topics related to visualizations or virtual reality (VR)/augmented reality (AR) were already addressed in early research and are still counted as two of the most popular topics [Liu et al., 2014]. Regarding sample size, it was found that the median number of participants in qualitative studies has increased over time, whereas the median number of participants in quantitative studies has decreased.

While there were usually between fifty and one hundred subjects in earlier studies, most commonly less than twenty participants take part in more recent quantitative research [Barkhuus and Rode, 2007].

Sample size differs according to the study setting and the methodology

The sample size within the CHI community was also investigated by Caine. She revealed local standards by analyzing all manuscripts published at CHI 2014. It was found that sample size differs according to the study setting and the type of methodology. The mean sample size is lower for qualitative studies than for quantitative ones and lower for studies conducted in-person than for remotely conducted studies. Regarding the methodology in remote studies, for example, the mean sample size is lower for interviews, diary studies and observations than for surveys and experiments [Caine, 2016].

2.2 Recruitment in User Studies

Factors to consider in the recruitment process

The sample size requirements are an essential aspect to consider in the recruitment process. Other factors that need to be taken into account are the ethical principles, the type of study design and the methodology used [Patel et al., 2003]. Since the course of the recruitment can determine the quality of the study results, it is very important to know where and how to recruit participants [James et al., 2014]. An adequate recruitment medium has to be chosen based on the aspects mentioned above.

2.2.1 The Recruitment Medium

Online methods outstand offline ones in terms of efficiency and costs

For the recruitment process, online methods seem to be the most appealing. They outstand offline methods not only in terms of efficiency (number of participants enrolled) but also, in terms of costs. Prior research has shown that the average cost per recruited participant was lower for online than for offline methods [Christensen et al., 2017]. For example, social media recruitment is characterized as inexpensive, simple, and efficient [Herbell and Zauszniewski,

2018]. On Amazon Mechanical Turk, hundreds of users are recruited in short time for low costs [Kittur et al., 2008].

AMAZON MECHANICAL TURK:
"Mechanical Turk (MTurk) is a website marketplace run by Amazon that connects requesters who have discrete, repetitive tasks (known as Human Intelligence Tasks or HITs) with workers from around the world" who get paid for completing these tasks [Hitlin, 2016].

Definition: *Amazon Mechanical Turk*

However, the recruitment via email, discussion boards, forums, and websites did not turn out to be effective. Koo and Skinner explain these results by highlighting the drawbacks of online recruitment. A major challenge for potential subjects is to distinguish between trustworthy and spam messages or postings on the Internet [Koo and Skinner, 2005]. Regarding drawbacks of online recruitment, the limitations of the single recruitment methods have to be considered. Since 89 % of all HITs on Amazon MTurk consists of surveys [Hitlin, 2016], this recruitment medium does not seem to be appropriate for every other type of data collection method. Furthermore, it is difficult to reach 'hidden' populations via the Internet. For this, snowball sampling is seen as a valuable tool [Waters, 2015].

Drawbacks of online methods

SNOWBALL SAMPLING:
"In this method, the existing study subjects recruit future subjects [with similar views or situations] among their acquaintances. Sampling continues until data saturation" [Naderifar et al., 2017].

Definition: *Snowball Sampling*

It is also seen as an effective method to study sensitive or private matters [Waters, 2015] and is mainly used in qualitative research [Biernacki and Waldorf, 1981]. Since participants in a chain are likely to have similar characteristics, one of the main drawback of this method is distortion [Waters, 2015].

Snowball sampling to study sensitive or private matters

2.2.2 The Role of Incentives

Incentives influence user participation

Apart from choosing an appropriate recruitment medium, researchers also have to decide whether or not to offer incentives. Incentives can influence the rate of participation positively [Fiore et al., 2014], whereat the effect sizes are larger for field than for lab studies, larger for qualitative than for quantitative approaches, and smaller for less complex tasks [Garbers and Konradt, 2014]. In general, pre-paid are more effective than promised incentives and monetary incentives are more effective than, for example, lotteries, charities, vouchers, or in-kind incentives [Hansen, 2007]. Prior research has shown that there are no differences in participation rates between nonmonetary incentives and no incentives [Kelly et al., 2017]. Regarding the amount of financial incentives, higher amounts can increase the quantity of participation [Mason and Watts, 2009]. For example, participants were more willing to participate when receiving 75 USD than when receiving 25 USD. Nevertheless, there was no difference for 50 USD and 75 USD [Kelly et al., 2017]. In conclusion, the impact of the amount of monetary incentives on participation rates is not monotonic, but equilibrium stages emerge as the monetary incentives increase [Liu and Feng, 2016]. Although the quantity of participation can be improved, there is no increase of the quality of work performed by participants when increasing the amount of financial incentives [Mason and Watts, 2009]. Indeed, extrinsic rewards could actually undermine the intrinsic motivation of participants [Ma et al., 2014], which could lead to a lower quality of work performed.

Pre-paid monetary incentives are most effective

Equilibrium stages emerge as monetary incentives increase

Incentives to target desired participants

Besides influencing the rate of participation, incentives can also be used to target desired participants because different types of incentives may attract different groups of people. For example, "lottery rewards attracted participants who held stronger openness-to-change values while a charity reward attracted those with stronger self-transcendence orientation" [Hsieh and Kocielnik, 2016]. Therefore, incentives can be helpful to recruit a sample that is diverse in race, ethnicity, and gender [Grady et al., 2005].

2.2.3 The Offer of Incentives

Factors that influence the amount of compensation

Prior research has shown that the amount of compensation is determined by several factors. Higher amounts are offered for studies that involved multiple interactions between participants and researchers, longer study durations, invasive procedures, and greater number of tasks [Ripley et al., 2010]. For example, the mean amount of compensation for studies including two tasks is 28.9 USD, while the mean amount for studies with three or more tasks is 189.7 USD. Another influencing factor is the methodology. For studies including a survey, the mean amount of compensation was 13 USD, while it was 24.20 USD for interview studies and 25 USD for focus groups [Latterman and Merz, 2001]. Furthermore, the recruitment medium affects the amount of compensation. For example, participants recruited via Amazon MTurk mostly conduct very short tasks with a payment of 0.10 USD or less [Hitlin, 2016].

Chapter 3

Method

In this chapter, the method is presented on which the thesis is based. Firstly, the collection of data is described, followed by the explanation of the data analysis. Finally, a brief description of the data set is given, including all factors that are considered in the further analysis in chapter 4 "Findings".

3.1 Data Collection

Papers of CHI 2019 were analyzed

The collection of data is based on contributions of the CHI conference to identify current practices of the HCI community. A systematic literature review of all papers published at CHI 2019 [1] was conducted. Data was extracted manually from each paper and saved in a spreadsheet.

File: Data Set CHI 19[a]

[a] http://hci.rwth-aachen.de/public/SurveyCHI2019/Data Set CHI 19.xlsx

List of investigated factors

The spreadsheet contains columns for the participants' details (number, gender, age, occupation, source, special characteristics, relationship to the study director), the details of

[1] https://dl.acm.org/doi/proceedings/10.1145/3290605

the study (topic, method, number of components, analysis approach, location, type of study, duration, time period, number of sessions, type of recording), the recruitment media, and details about the compensation (occurence, type, amount) as well as funding.

Participants' details

The number of participants is specified in two columns, including both the number before and after cleaning. The first column indicates the total number of participants that took part in a user study, while the second column shows the number of participants who were included in the analysis. The gender of the participants is tabulated in three columns. In the first column, the number of female participants is recorded, followed by the column in which the number of male participants is specified. The third column represents the number of participants who identified themselves neither as female nor as male. The age of the participants is saved in intervals whenever an interval is given. Otherwise, the average age is recorded.

Details of the study

The topic column consists of the keywords given at the beginning of each paper. If none are listed, the main keywords of the heading and the introduction were manually extracted, with a maximum of 10 keywords. In order to decide whether a qualitative, quantitative or mixed-method analysis approach is present, the result section of each paper was investigated. If only statistical findings are reported, the study is classified as quantitative, if only non-statistical results are presented, it is categorized as a qualitative study. Mixed-method studies include both, statistical and non-statistical results. Regarding the number of components, each survey, interview, etc., is counted as a single one. If a study contains two surveys, interviews, etc., on different topics, they are counted as two components whereas tutorials are not counted as a separate element. The duration of a study is specified as the average duration per session in minutes and the period is stored as the average period in weeks. Regarding the type of study, online studies are conducted via the Internet, phone studies via telephone. For all other studies, it was investigated whether the study was controlled or not. If a study was controlled, it is listed as a lab, otherwise as a field study.

The amount of compensation is given as the average amount in USD. Other currencies were converted to USD between February and June 2020 using Google's currency converter. In order to decide whether or not a study received funding, the statements in the acknowledgement section of each paper were considered.

Details about compensation and funding

Every row of the spreadsheet represents a single study. Studies that do not include human subjects are not excluded from the data set but are not further analyzed with regard to the factors listed above either.

Each user study is stored in a single row of the spreadsheet

3.2 Data Analysis

The data set was analyzed using Jupyter Notebook. The download link can be found in chapter 1 "Introduction".

To get a first overview about the data, categorical factors are described with frequency distributions. For each distribution, only those studies are considered that actually reveal information about the corresponding factor. Therefore, the number of included studies N is given for each investigation. In order to calculate the frequency distribution of the topics, the keywords that appeared most frequently in the data set were considered and clustered manually. The breakdown of keywords that were taken into account for each individual topic is shown in table 3.1. Regarding the frequency distribution of the recruitment medium, all social media channels that originally occurred in the data set (*Facebook*, *Instagram*, *LinkedIn*, *Reddit*, *Slack*, *Telegram*, *Tumblr*, *Twitter*) have been combined as *social media*. For metric factors, it is tested whether they are distributed normally. They are described by giving the mean and range of values as well as using boxplots to identify the median and outliers of the data points. Again, only those studies are considered that reveal information about the factor that is investigated.

The description of single factors

After focusing on the description of single factors, it was investigated whether there are differences in the frequency and amount of compensation according to the used analy-

Compensation and the analysis approach

Topic	Keywords
AR/VR/Mixed reality	AR, augmented reality, mixed reality, virtual reality, VR
Mobile Devices/Wearables	mobile device, smartphone, smartwatch, wearable
Health/Wellbeing	health, wellbeing, well-being
AI/Machine Learning	AI, artificial intelligence, chatbot, machine learning, robot
Visualizations/Analytics	analytics, visualization
Privacy/Security	authentication, privacy, secure, security
Haptics	haptic

Table 3.1: Keywords considered in the Analysis of the Topic

sis approach. In figure 3.1, this is visualized as the first step of the analysis. The differences in the frequency of compensation are investigated using frequency distributions, while the differences in the amount of compensation are analyzed with boxplots.

How do qualitative, quantitative and mixed-method studies differ?

To explain the differences in the frequency and amount of compensation for the various analysis approaches, it was analyzed whether qualitative, quantitative and mixed-method studies differ with regard to the number of participants, the duration of the study, the number of components, the recruitment medium, the topic of the study, the used method, the type of study, and the type of recording. In figure 3.1, this is visualized as the second step of the analysis. Categorical factors are analyzed using frequency distributions, whereas boxplots were generated for the investigation of metric factors.

Differences in compensation according to several factors

After identifying factors in which qualitative, quantitative and mixed-method studies differ, it was tested whether the frequency and amount of compensation also differ with regard to these factors. In figure 3.1, this is visualized as the third step of the analysis. In particular, for the frequency of compensation, it was investigated whether it differs based on the recruitment medium, the topic of the study, the used method and the type of study. Again, frequency distributions are used for the analysis. For the amount of compensation, it was tested whether it differs based on the number of participants, the duration of the study, the number of components, the recruitment medium, the topic of the study, the used method, the type of study, and the type

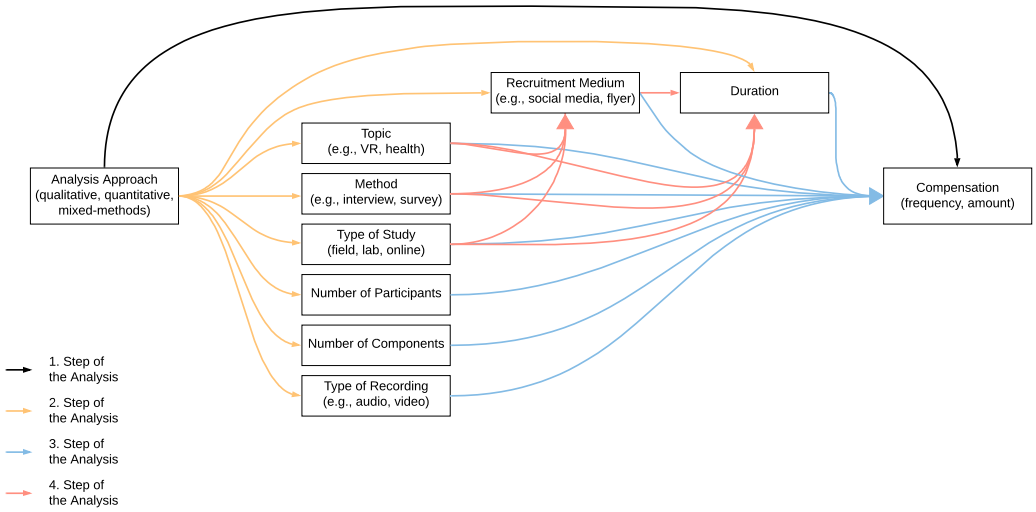

Figure 3.1: The Steps of the Data Analysis

of recording. Categorical factors were analyzed using boxplots. For the investigation of metric factors, scatter plots were generated and Pearson's correlation coefficients were calculated. For this, identified outliers were excluded. The correlation coefficient *r* is reported in chapter 4 "Findings". It is interpreted as follows [Pallant, 2011]:

- r = .10 to .29 or r = -.10 to -.29 weak correlation
- r = .30 to .49 or r = -.30 to -.49 moderate correlation
- r = .50 to 1.0 or r = -.50 to -1.0 strong correlation

Amazon MTurk and the duration of a study as explanations for the differences in compensation

In the fourth step of the analysis (see figure 3.1), it was investigated whether the recruitment medium and/or the duration of a study might be one of the main factors that could explain the differences in the frequency and amount of compensation. For this, the recruitment via Amazon MTurk and the duration are analyzed with regard to the topic of the study, the used method and the type of study. The duration is additionally analyzed according to the recruitment medium. Frequency distributions were generated to investigate the use of Amazon MTurk, boxplots were created to investigate the duration of the study.

Correlations and plots in Python

All correlations are performed using the module *Corrcoef* of

the library *Numpy* [2]. All plots are created using the module *Pyplot* of the library *Matplotlib* [3]. It should be noted that the scales of the boxplots were shortened manually in order to focus on the more common values instead of outliers and to emphasize differences more clearly.

3.3 Description of the Data Set

1031 user studies were analyzed with a median sample size of 16

The data set consists of 702 papers, including 1031 user studies. Thus, on average, 1.47 studies are reported in each paper. The number of participants ranges from 1 to 69,174 with a median sample size of 16 ($N = 992$). Sample sizes that include more than 75 participants are counted as outliers (see figure 3.2).

Details about the study

A qualitative analysis approach is used in 47.91 % of all studies ($N = 1031$). 31.81 % use a quantitative approach whereas 20.27 % are counted as mixed-method studies. The field of virtual reality, augmented reality and mixed reality is one of the most addressed fields in CHI 2019. 13.29 % of all studies ($N = 1031$) are dealing with VR, AR and mixed reality related topics. The absolute number of studies that are related to a specific topic is visualized in figure 3.3. The most used methods are experiments, interviews and surveys. Experiments are used in 44.71 %, interviews in 38.80 % and surveys in 37.63 % of all studies ($N = 1031$). The absolute number of studies that include a specific method is visualized in figure 3.4. The number of components ranges from one to ten with one being the most common number of components ($N = 1025$). Studies that include more than three components are counted as outliers. The duration ranges from three minutes to 66.5 hours with a median duration of 60 minutes ($N = 511$). Outliers are those studies that last longer than 153.3 minutes (see figure 3.5). Regarding the studies in which the type is indicated ($N = 905$), 45.41 % of all studies took place in the field, 34.70 % in the lab and 23.31 % took place online. Audio recording is used

[2] https://numpy.org/doc/stable/reference/generated/numpy.corrcoef.html

[3] https://matplotlib.org/api/pyplot_api.html

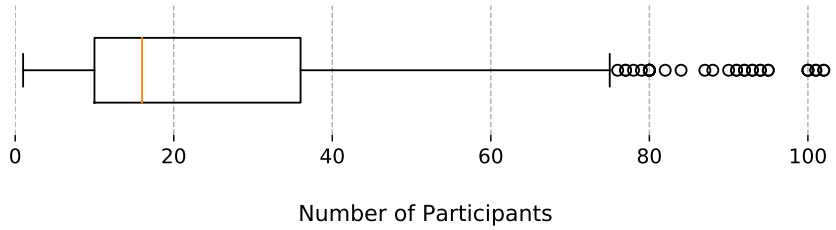

Figure 3.2: The Number of Participants

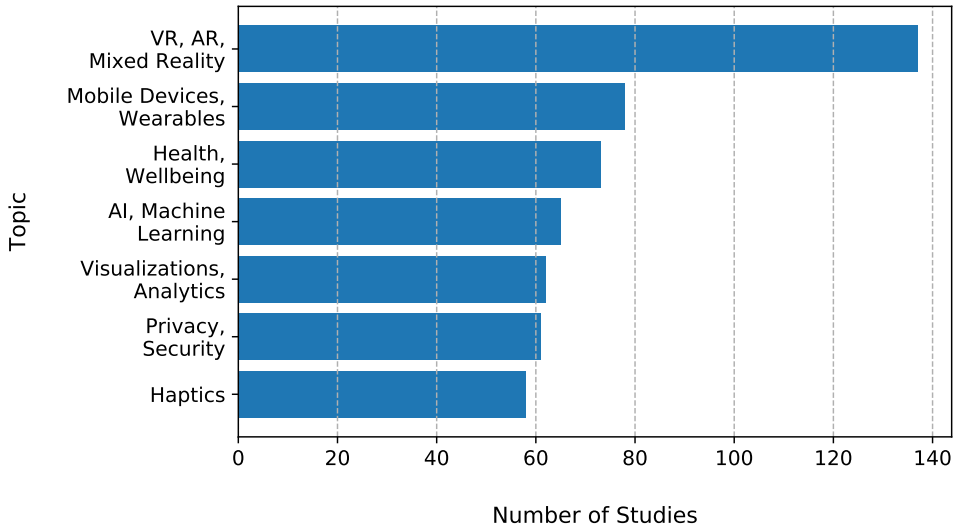

Figure 3.3: The Number of Studies according to the Topic

in 22.01 % and video recording is used in 8.96 % of all studies ($N = 1027$). In 6.13 % both, audio and video recording is used.

Details about the recruitment

The most used recruitment media are social media, mailing lists, snowball sampling and Amazon MTurk. Social media are used in 23.90 %, mailing lists in 18.78 %, snowball sampling in 15.61 % and Amazon MTurk in 15.12 % of all studies ($N = 410$). The absolute number of studies that use a specific recruitment medium is visualized in figure 3.6. In 36.37 % of all studies, compensation was offered to the participants, in 62.37 % compensation was not offered and in 1.26 %, compensation was not offered to everyone

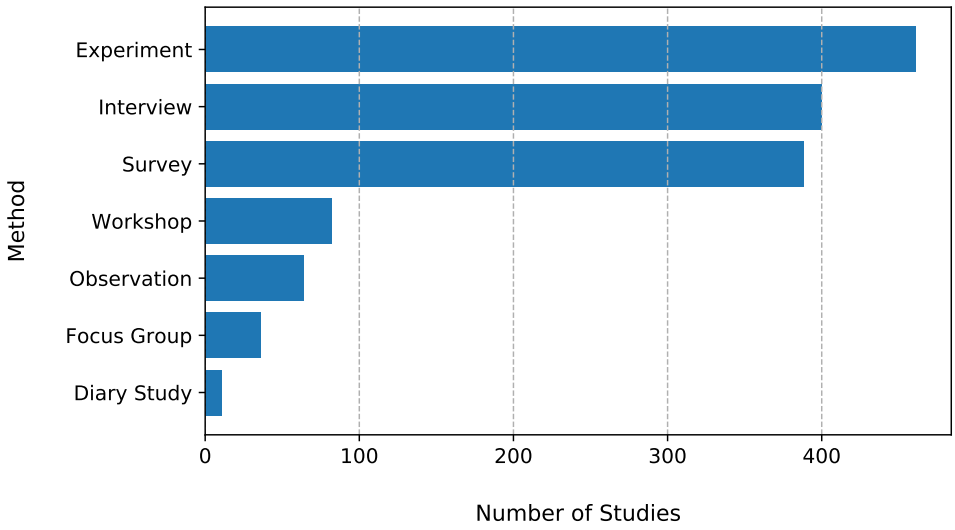

Figure 3.4: The Number of Studies according to the Method

but there was a lottery ($N = 1031$). The amount of compensation ranges from 0.05 USD to 1,330 USD with a median amount of compensation of 15 USD ($N = 318$). Compensation amounts higher than 60.93 USD are counted as outliers (see figure 3.7).

Explanation for outliers

Outliers for the number of participants and components as well as for the duration of the study and the amount of compensation can be explained by the large number of study features with high variety of characteristic values. Two extreme examples would be:

- A study with 69,174 students that lasted an entire school year [Hutt et al., 2019]
- A study with four designers that lasted between 60 to 73 hours with an hourly compensation rate of 20 USD per participant [Huang et al., 2019]

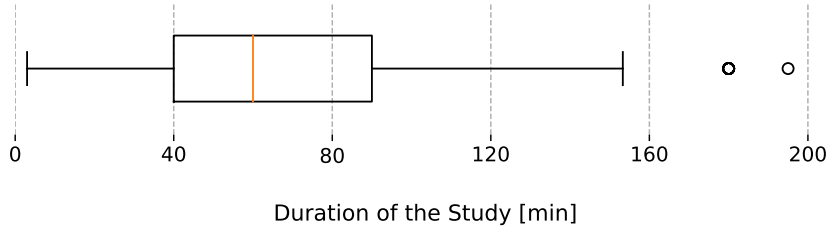

Figure 3.5: The Duration of the Study

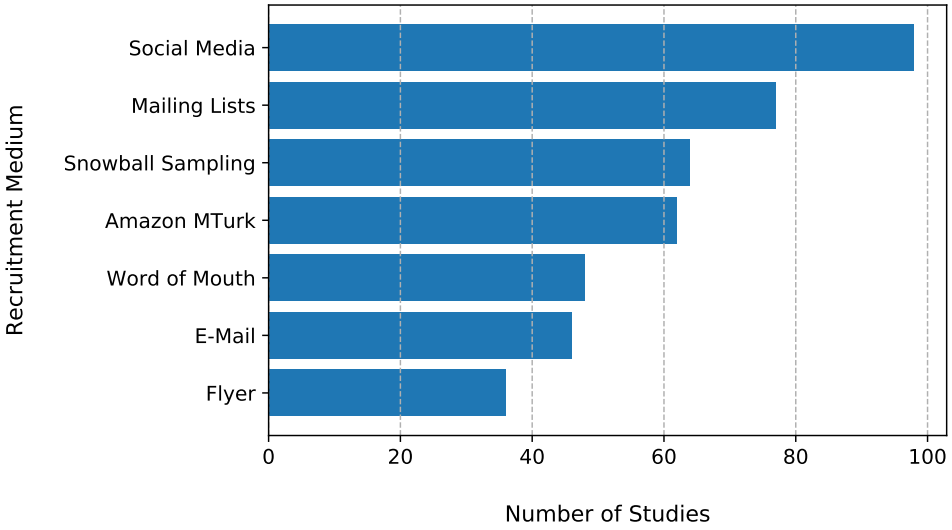

Figure 3.6: The Number of Studies according to the Recruitment Medium

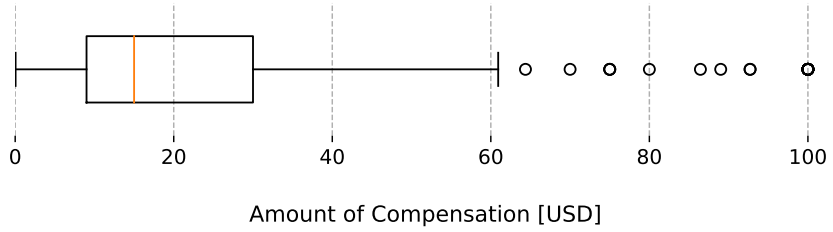

Figure 3.7: The Amount of Compensation

Chapter 4

Findings

In the following, the findings of the analysis are presented. First, the differences in compensation according to the analysis approach are described. In order to explain these differences, the following factors are investigated:

- Number of Participants
- Duration of the Study
- Number of Components
- The Recruitment Medium
- The Topic of the Study
- The Method
- The Type of Study
- The Type of Recording

4.1 The Analysis Approach

Compensation differs depending on the analysis approach

Figure 4.1 shows that the frequency of compensation differs depending on the analysis approach ($N = 1031$). While participants were compensated in 45.12 % of all quantitative and 43.54 % of all mixed-method studies, participation

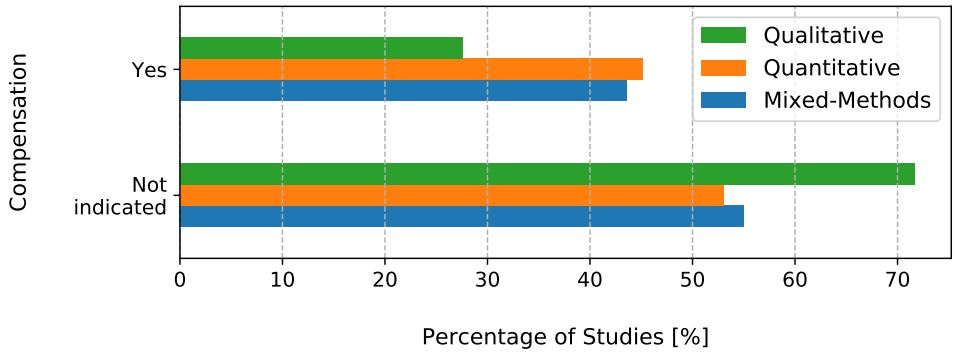

Figure 4.1: Frequency of Compensation according to the Analysis Approach

was only compensated in 27.53 % of all qualitative studies. The amount of compensation also differs according to the analysis approach ($N = 318$), which can be seen in figure 4.2. While the median amount is 25 USD for qualitative and 17.50 USD for mixed-method studies, participants only received a median amount of 8.10 USD for quantitative studies.

4.2 Number of Participants

To find reasons for the differences in compensation according to the analysis approach, the number of participants is investigated as a first factor. It is analyzed whether there are differences in the sample size between the various analysis approaches ($N = 992$) and whether there is a correlation between the number of participants and the amount of compensation ($N = 318$).

Sample size differs depending on the analysis approach

Figure 4.3 shows that the sample size differs depending on the analysis approach. In quantitative studies, the median number of participants is 30, while it is 18 for mixed-method and 11 for qualitative studies. Pearson's correlation coefficient reveals a weak negative correlation between the number of participants and the amount of compensation (r = -.12). Figure 4.4 shows the corresponding scatter plot.

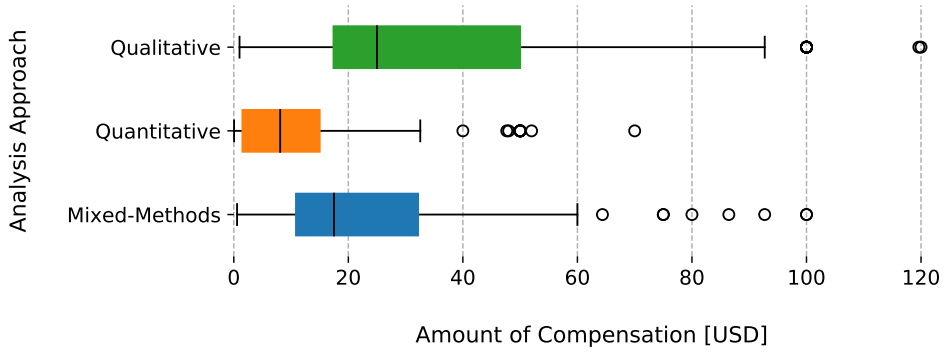

Figure 4.2: Amount of Compensation according to the Analysis Approach

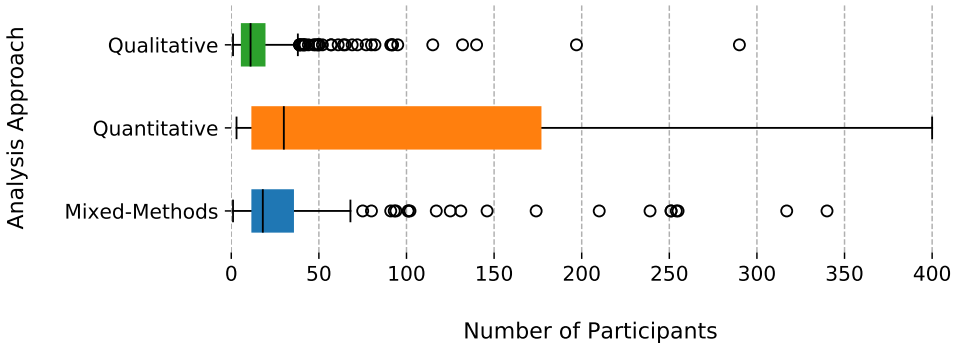

Figure 4.3: Number of Participants according to the Analysis Approach

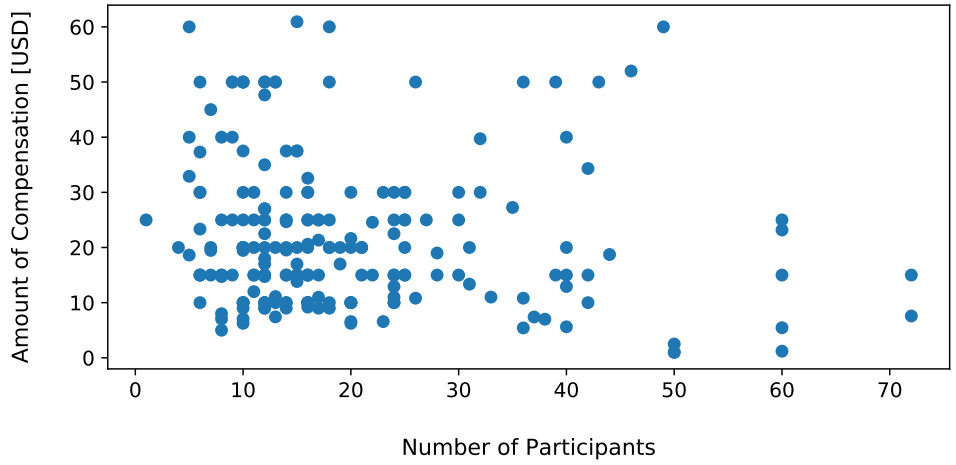

Figure 4.4: Number of Participants and Amount of Compensation

4.3 Duration of the Study

As the second factor, the duration of the study is investigated. Again, it is analyzed whether there are differences among the various analysis approaches ($N = 511$) and whether there is a correlation between the study duration and the amount of compensation ($N = 212$).

The duration differs depending on the analysis approach

Figure 4.5 shows that the duration differs depending on the analysis approach. Qualitative and mixed-method studies have a median duration of 60 minutes, whereas the median duration for quantitative studies is 40 minutes. Pearson's correlation coefficient reveals a strong positive correlation between the duration of a study and the amount of compensation (r = .59). The corresponding scatter plot is shown in figure 4.6.

4.4 Number of Components

The last metric factor that was investigated is the number of components. It is analyzed whether there are differences between the various analysis approaches ($N = 1025$) and whether there is a correlation between the number of components and the amount of compensation ($N = 318$).

The number of components differs depending on the analysis approach

Figure 4.7 shows that the number of components differs depending on the analysis approach. Qualitative and quantitative studies have a median number of components of one whereas the median number of components for mixed-method studies is two. Pearson's correlation coefficient reveals a weak positive correlation between the number of components of a study and the amount of compensation (r = .17).

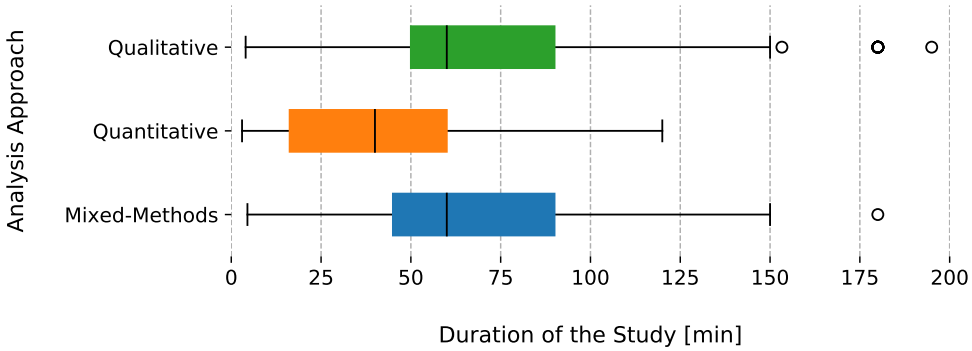

Figure 4.5: Duration of the Study according to the Analysis Approach

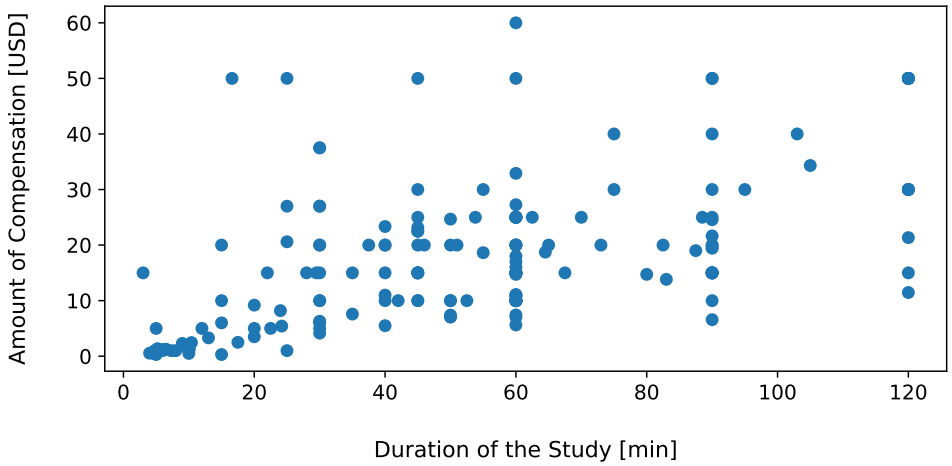

Figure 4.6: Duration of the Study and Amount of Compensation

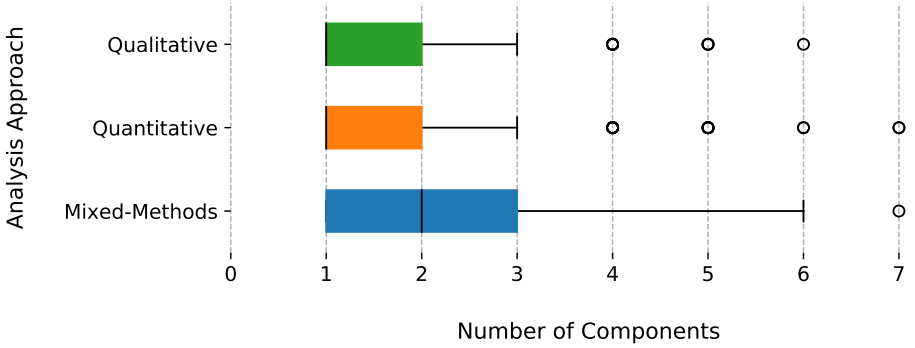

Figure 4.7: Number of Components according to the Analysis Approach

4.5 The Recruitment Medium

For the analysis of the recruitment medium, it is investigated whether the medium differs with regard to the analysis approach ($N = 410$) and whether the frequency and amount of compensation differ according to the recruitment medium (frequency: $N = 410$, amount: $N = 196$). In order to investigate whether the recruitment medium itself is a reason for these differences in compensation or whether there are other factors influencing the results, it is analyzed whether the duration of the study also differs according to the used recruitment medium ($N = 237$).

The recruitment medium differs depending on the analysis approach

Figure 4.8 shows that social media were used more often for the recruitment in qualitative studies (29.38 %) than for the recruitment in quantitative (19.12 %) and mixed-method studies (18.75 %). The same holds for snowball sampling (qualitative: 25.26 %, quantitative: 2.94 %, mixed-method: 13.75 %) and word of mouth (qualitative: 14.95 %, quantitative: 9.56 %, mixed-method: 7.50 %) . Amazon MTurk was most often used for the recruitment in quantitative studies (40.44 %), it was only used in 8.75 % of all mixed-method studies and was never used for qualitative ones. Mailing lists were used more often for the recruitment in mixed-method studies (27.50 %) than for the recruitment in qualitative (17.01 %) and quantitative ones (16.18 %). The same also holds for the recruitment via e-mail (mixed-method: 15.00 %, qualitative: 12.37 %, quantitative: 7.35 %) and flyer (mixed-method: 12.50 %, qualitative: 9.28 %, quantitative: 5.88 %).

Compensation differs depending on the recruitment medium

In figure 4.9, it can be seen that participants were most likely to receive compensation when recruited through Amazon MTurk (93.55 %), followed by the recruitment with flyers (69.44 %), mailing lists (54.55 %), word of mouth (52.08 %), and e-mail (50.00 %). Participants who were recruited through social media and snowball sampling were least likely to receive compensation. They were compensated in 39.80 % and 32.81 % respectively.

As figure 4.10 shows, the median amount of compensation is highest for people recruited with snowball sam-

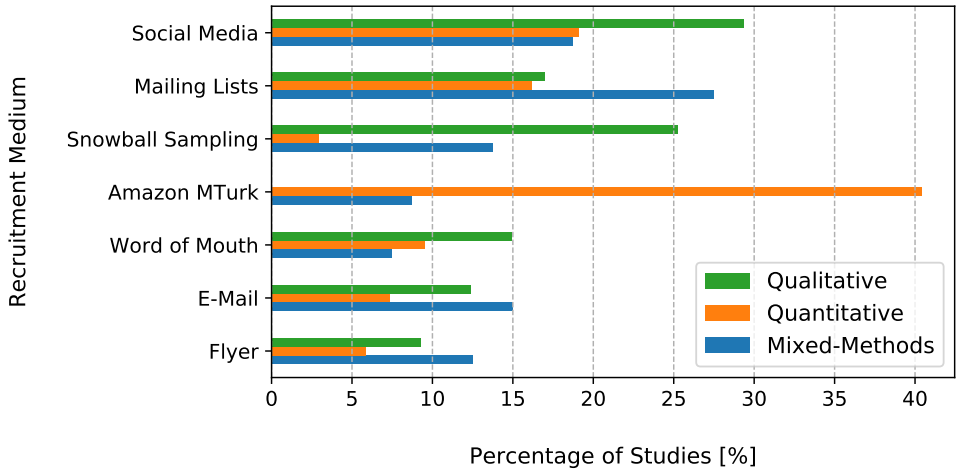

Figure 4.8: The Recruitment Medium and the Analysis Approach

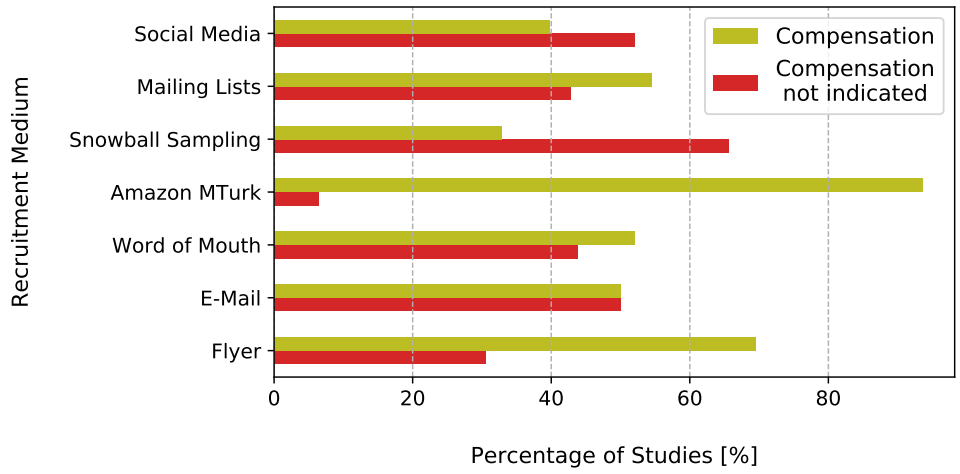

Figure 4.9: Frequency of Compensation according to the Recruitment Medium

pling ($\tilde{X} = 30$ USD) and lowest for participants recruited via Amazon MTurk ($\tilde{X} = 1.25$ USD). When recruited via social media, mailing lists, word of mouth, e-mail and flyer, participants got compensated with a median amount between 20 USD and 22.50 USD.

Studies for which participants were recruited via social media, mailing lists, snowball sampling and word of mouth

The duration differs depending on the recruitment medium

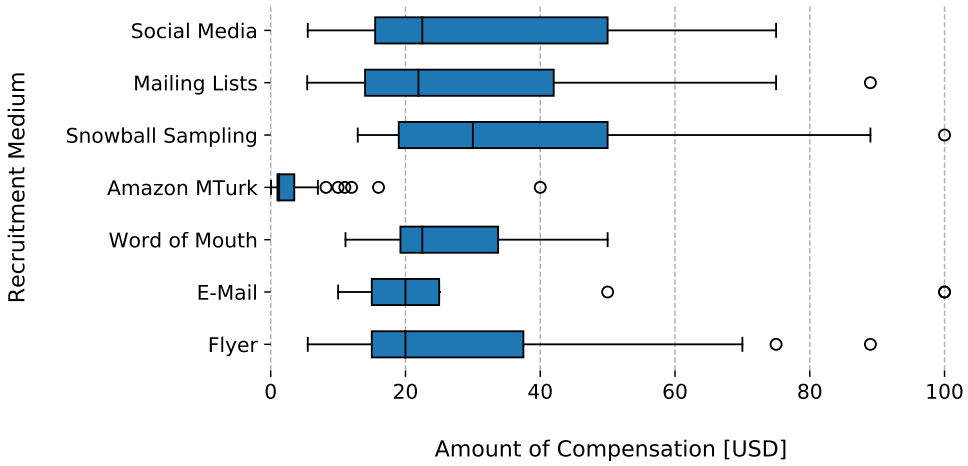

Figure 4.10: Amount of Compensation according to the Recruitment Medium

have a median duration of 60 minutes (see figure 4.11). The median duration for studies with e-mail recruitment is 51.75 minutes, while it is 51 minutes for studies with flyer recruitment. Studies in which Amazon MTurk was used have a median duration of ten minutes.

4.6 The Topic

As a next factor, the topic of the study was analyzed. For this, it is investigated whether the topic differs according to the analysis approach ($N = 1031$) and whether the frequency and amount of compensation differ with regard to the topic (frequency: $N = 1031$, amount: $N = 318$). To examine the influence of other factors on the received results, it is analyzed whether the use of Amazon MTurk ($N = 410$) and the duration of the study ($N = 511$) also differ among the various topics.

The topic differs depending on the analysis approach

Figure 4.12 shows that health/wellbeing related topics were more often addressed in qualitative studies (11.54 %) than in quantitative (2.44 %) and mixed-method ones (3.83 %), whereas mobile devices/wearables as well as AI/machine learning and haptics were more of-

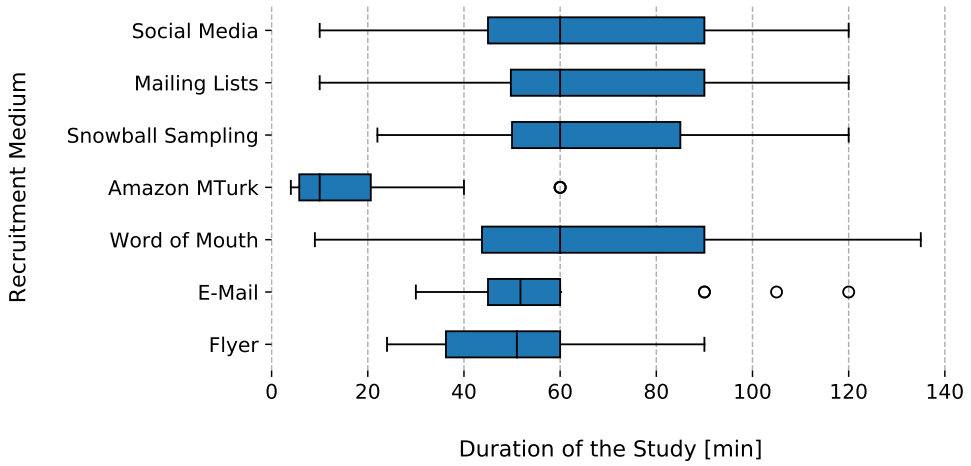

Figure 4.11: Duration of the Study according to the Recruitment Medium

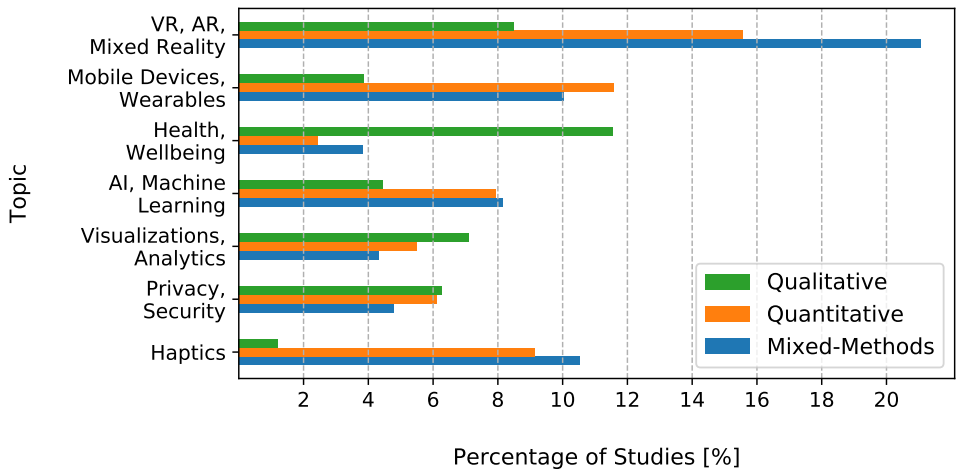

Figure 4.12: The Topic and the Analysis Approach

ten addressed in quantitative (mobile devices/wearables: 11.59 %, AI/machine learning: 7.93 %, haptics: 9.15 %) and mixed-method studies (mobile devices/wearables: 10.05 %, AI/machine learning: 8.13 %, haptics: 10.53 %) than in qualitative ones (mobile devices/wearables: 3.85 %, AI/machine learning: 4.45 %, haptics: 1.21 %). Visualizations/analytics as well as privacy/security related topics were more often addressed in qualitative (visual-

izations/analytics: 7.09 %, privacy/security: 6.28 %) and quantitative studies (visualizations/analytics: 5.49 %, privacy/security: 6.10 %) than in mixed-method studies (visualizations/analytics: 4.31 %, privacy/security: 4.78 %). VR/AR/mixed reality topics were most often addressed in mixed-method studies (21.05 %), followed by quantitative (15.55 %) and qualitative ones (8.50 %).

Compensation differs depending on the topic

In figure 4.13, it can be seen that participants were most likely to receive compensation when they took part in a privacy/security (63.93 %), visualizations/analytics (51.61 %) or AI/machine learning (49.23 %) related study, followed by studies concerning mobile devices/wearables (41.03 %) and haptics (32.76 %). Participants who took part in a health/wellbeing or VR/AR/mixed reality related study were least likely to receive compensation. They were compensated in 28.77 % and 28.47 % respectively.

As seen in figure 4.14, the median amount of compensation is highest for people who participated in health/wellbeing related studies ($\tilde{X} = 37.50$ USD), followed by the participation in studies related to mobile devices/wearables ($\tilde{X} = 20$ USD). For the participation in visualizations/analytics as well as AI/machine learning related studies, the median amount of compensation is lowest (visualizations/analytics: $\tilde{X} = 10$ USD, AI/machine learning: $\tilde{X} = 9.19$ USD). For the participation in VR/AR/mixed reality as well as privacy/security and haptics related studies, participants received a median amount of 15 USD.

When was Amazon MTurk used?

Figure 4.15 shows that Amazon MTurk was most often used for visualizations/analytics (17.74 %), privacy/security (16.39 %) as well as AI/machine learning related topics (15.38 %). For mobile devices/wearables related studies, it was used in 2.56 % and for health/wellbeing related studies in 1.37 %. Amazon MTurk was never used for studies related to VR/AR/mixed reality or haptics.

The duration differs depending on the topic

With a median duration of 81.25 minutes, health/wellbeing related studies lasted the longest, followed by studies related to VR/AR/mixed reality, mobile devices/wearables

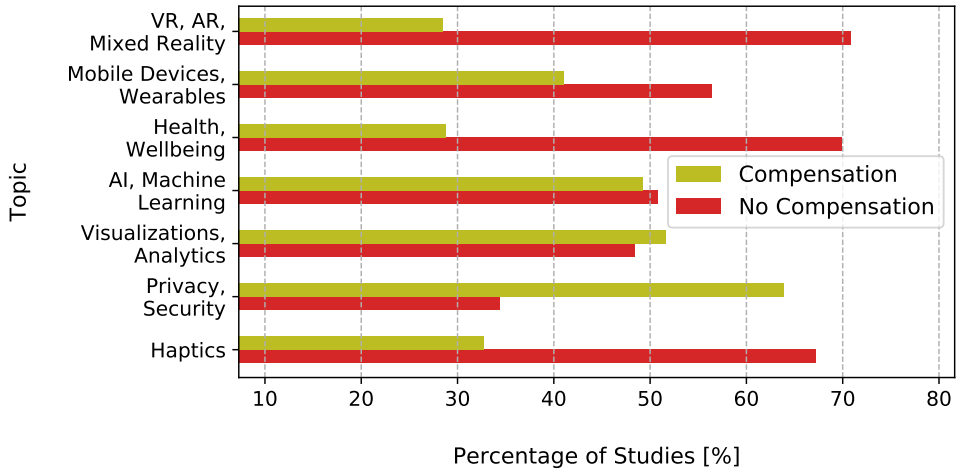

Figure 4.13: Frequency of Compensation according to the Topic

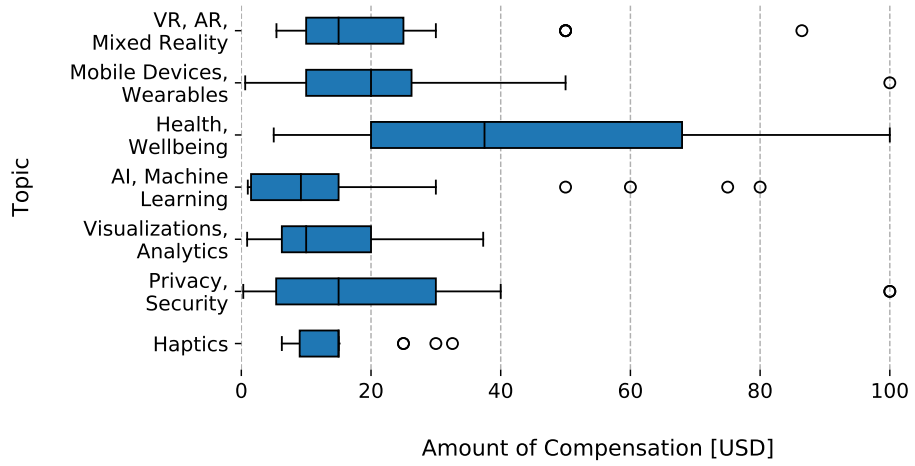

Figure 4.14: Amount of Compensation according to the Topic

or visualizations/analytics (see figure 4.16). Studies that address these topics have a median duration of 60 minutes. Privacy/security related studies have a median duration of 53 minutes. With a median duration of 45 minutes, studies related to AI/machine learning or haptics have the shortest median duration.

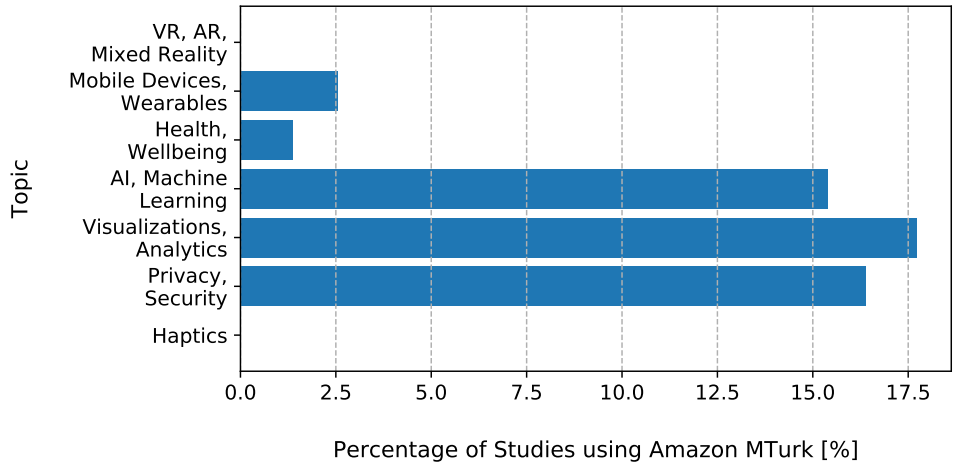

Figure 4.15: The Use of Amazon MTurk according to the Topic

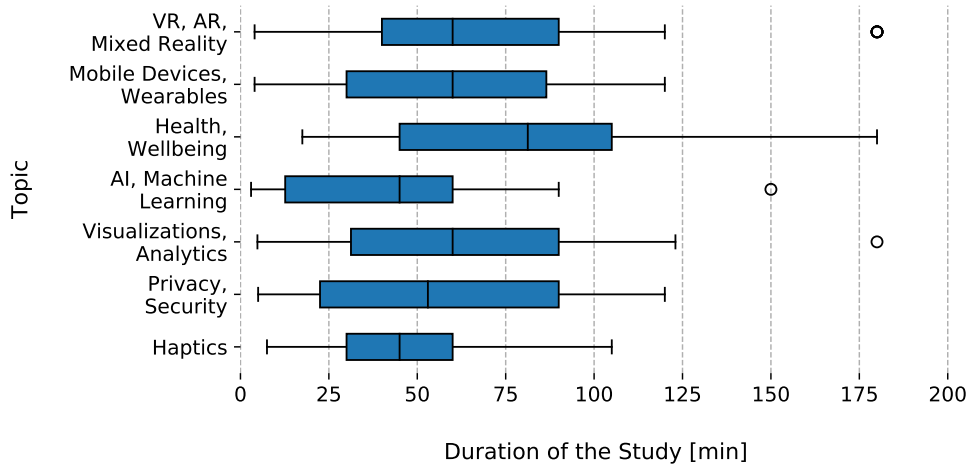

Figure 4.16: Duration of the Study according to the Topic

4.7 The Method

For the analysis of the method, it is analyzed whether the method differs according to the analysis approach ($N = 1031$). Afterwards, it is examined whether the frequency and amount of compensation differ with regard to the used method (frequency: $N = 1031$, amount: $N = 318$).

To investigate whether the method itself is a reason for the differences in compensation or whether there are other factors influencing the results, it is analyzed whether the use of Amazon MTurk ($N = 410$) and the duration of the study also differ among the various methods ($N = 511$).

The method differs depending on the analysis approach

Figure 4.17 shows that interviews were used more often for qualitative (57.09 %) and mixed-method studies (50.24 %) than for quantitative ones (3.96 %). The same holds for diary studies (qualitative: 1.42 %, mixed-method: 1.44 %, quantitative: 0.30 %). Experiments and surveys were used more often for quantitative (experiment: 71.95 %, survey: 57.93 %) and mixed-method studies (experiment: 74.16 %, survey: 60.29 %) than for qualitative ones (experiment: 14.17 %, survey: 14.57 %). In qualitative studies, more workshops were used (15.79 %) than for quantitative (0.61 %) and mixed-method studies (0.96 %). The same holds for observations (qualitative: 10.53 %, quantitative: 0.91 %, mixed-method: 4.31 %) and focus groups (qualitative: 6.68 %, quantitative: 0.00 %, mixed-method: 1.44 %).

Compensation differs depending on the method

In figure 4.18, it can be seen that participants were most likely to receive compensation when they took part in a diary study (63.64 %), followed by the participation in experiments (47.29 %), surveys (44.33 %), interviews (37.50 %) and focus groups (36.11 %). Participants who took part in a workshop or an observation were the least likely to receive compensation. They were compensated in 17.07 % of all workshops and 14.06 % of all observations.

Figure 4.19 shows that the median amount of compensation is highest for people who participated in workshops ($\tilde{X} = 75$ USD) and focus groups ($\tilde{X} = 60.93$ USD), followed by diary studies ($\tilde{X} = 43.75$ USD) and observations ($\tilde{X} = 33.61$ USD). For interviews ($\tilde{X} = 20$ USD), experiments ($\tilde{X} = 15$ USD) and surveys ($\tilde{X} = 15$ USD), the median amount of compensation is comparatively low.

When was Amazon MTurk used?

Only for experiments and surveys, participants got recruited via Amazon MTurk. 22.86 % of all surveys and 24.68 % of all experiments were posted on the crowdsourcing website.

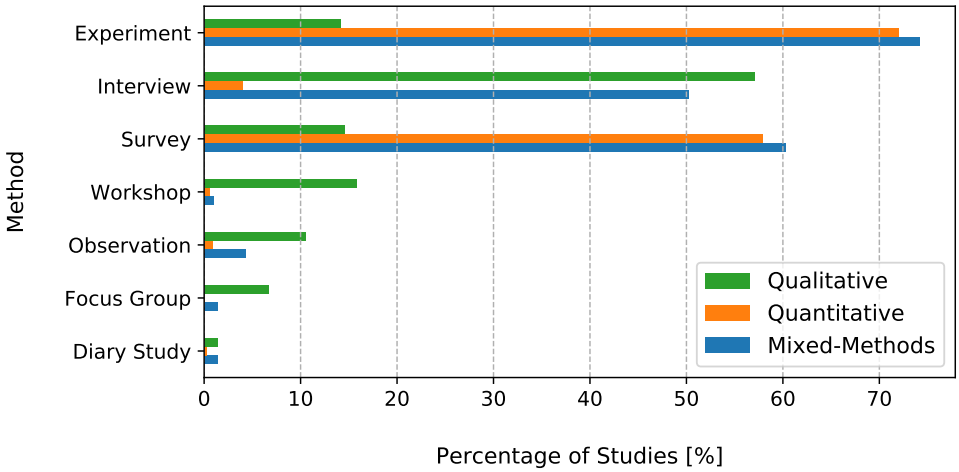

Figure 4.17: The Method and the Analysis Approach

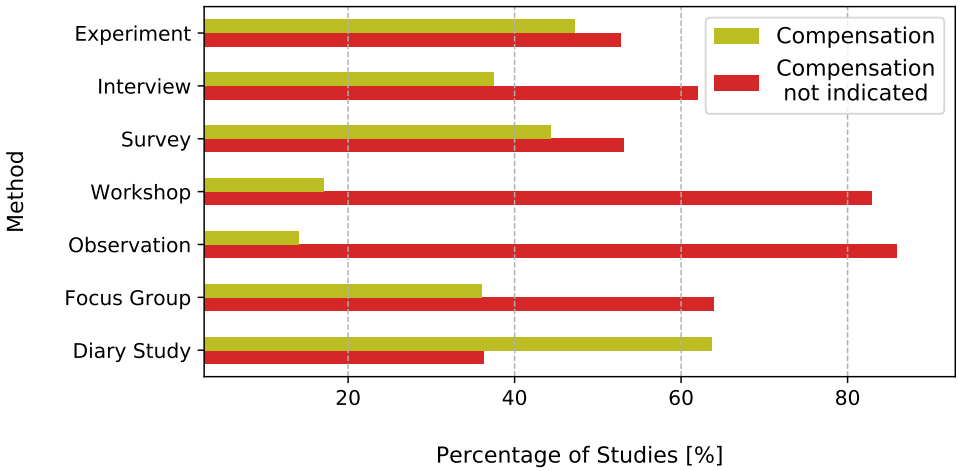

Figure 4.18: Frequency of Compensation according to the Method

The duration differs depending on the method

As it can be seen in figure 4.20, the median duration of workshops ($\tilde{X} = 97.5$ min), observations ($\tilde{X} = 90$ min) and focus groups ($\tilde{X} = 82.5$ min) is higher than the median duration of experiments ($\tilde{X} = 60$ min), interviews ($\tilde{X} = 60$ min) and surveys ($\tilde{X} = 55$ min). Diary studies are not included in this analysis because of insufficient data samples.

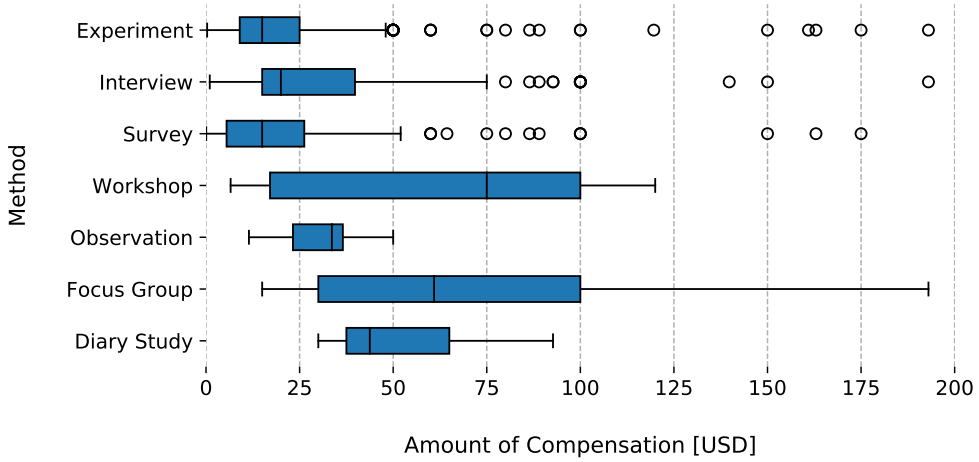

Figure 4.19: Amount of Compensation according to the Method

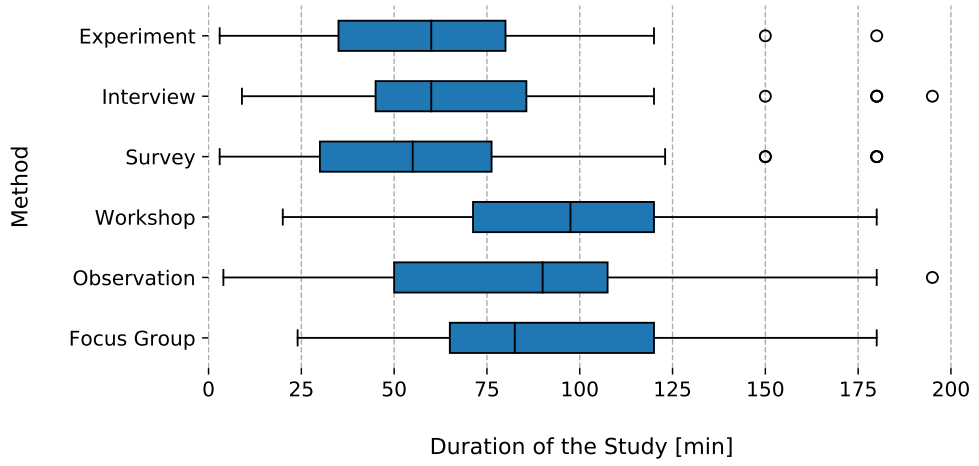

Figure 4.20: Duration of the Study according to the Method

4.8 The Type of Study

The next factor that is examined is the type of study. After analyzing whether the type of study differs among the various analysis approaches ($N = 905$), it is investigated whether the frequency and amount of compensation differ according to the type of study (frequency: $N = 905$, amount:

$N = 290$). In order to investigate whether the study type itself is a reason for the differences in compensation or whether there are other factors influencing the results, it is analyzed whether the use of Amazon MTurk ($N = 364$) and the duration of the study also differ according to the type of study ($N = 440$).

The study type differs depending on the analysis approach

In figure 4.21, it can be seen that qualitative studies took place in the field more often (82.60 %) than mixed-method (22.87 %) and quantitative studies (10.03 %), whereas quantitative studies took place online more often (35.60 %) than mixed-method (19.15 %) and qualitative studies (15.93 %). Mixed-method (60.64 %) and quantitative studies (54.69 %) took place in the lab more often than qualitative ones (7.60 %).

Compensation differs depending on the study type

Figure 4.22 shows that participants were most likely to receive compensation in online studies (52.13 %) and least likely to receive compensation in field studies (27.01 %). Regarding lab studies, they received compensation in 43.31 %.

As it can be seen in figure 4.23, the median amount of compensation is highest for people who participated in field studies ($\tilde{X} = 30$ USD) and lowest for participants of online studies ($\tilde{X} = 5.49$ USD). Participation in lab studies was compensated with a median amount of 15 USD.

When was Amazon MTurk used?

Amazon MTurk was used in 34.94 % of all online studies and in 2.65 % of all field studies. It was never used for lab studies.

The duration differs depending on the study type

Field studies have a median duration of 65 minutes, whereas the median duration is 60 minutes for lab and 30 minutes for online studies (see figure 4.24).

4.9 The Type of Recording

Lastly, the type of recording is investigated. It is analyzed whether the recording differs according to the analysis approach ($N = 1027$) and whether the amount of compensation differs with regard to the type of recording ($N = 318$).

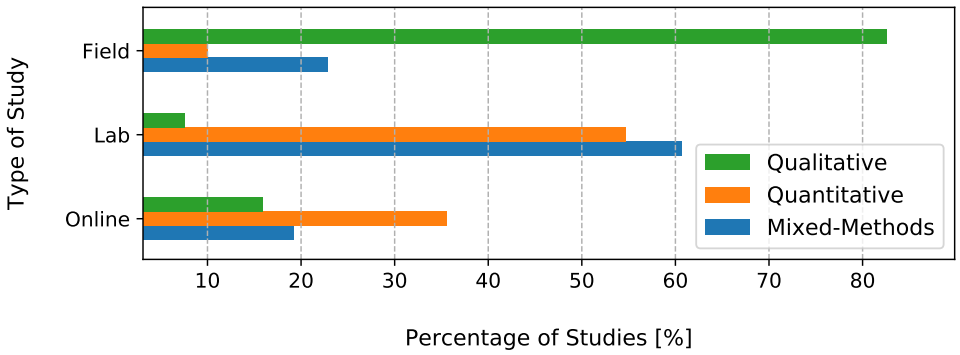

Figure 4.21: The Type of Study and the Analysis Approach

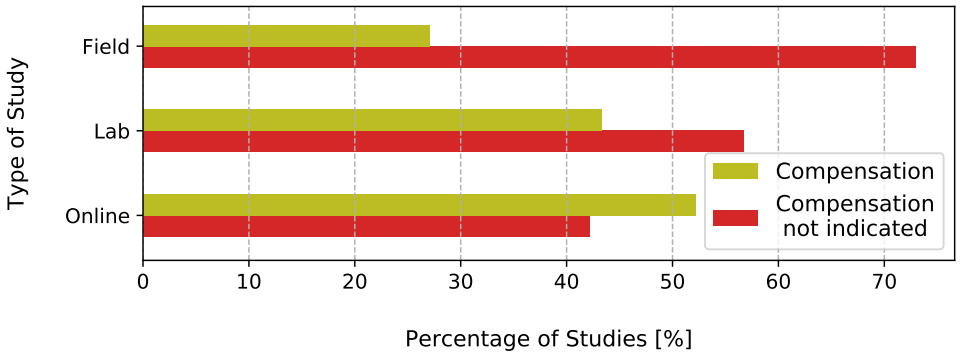

Figure 4.22: Frequency of Compensation according to the Study Type

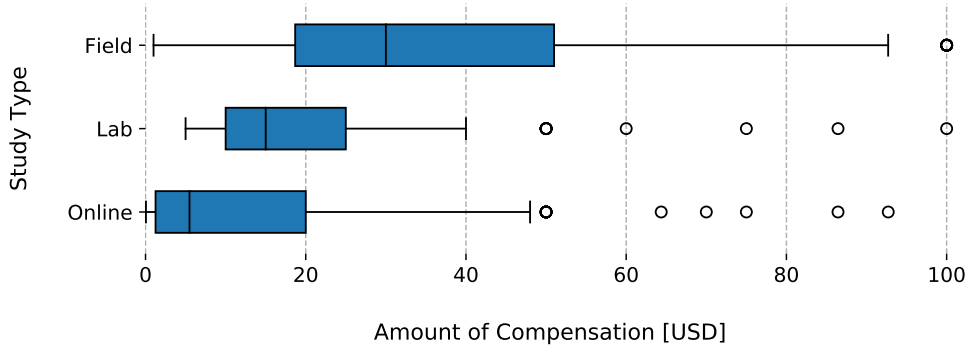

Figure 4.23: Amount of Compensation according to the Study Type

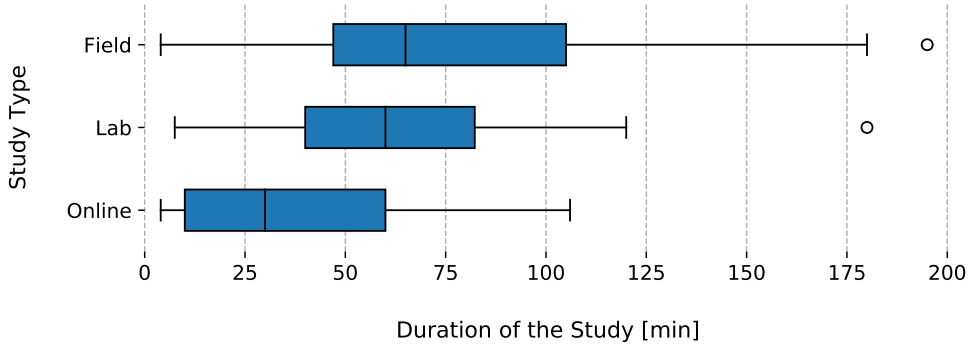

Figure 4.24: Duration of the Study according to the Study Type

The recording type differs depending on the analysis approach

Figure 4.25 shows that in most quantitative and mixed-method studies, researchers did not use any kind of recording (quantitative: 80.67 %, mixed-method: 57.21 %). If recording was used in quantitative studies, it was mostly video recording (5.52 %). Audio or audio and video recording were each used in only 0.92 % of all quantitative studies. If recording was used in mixed-method studies, it was mostly either only audio (13.96 %) or only video recording (13.46 %). In the rarest of cases, both, audio and video were used (6.25 %). In qualitative studies, participants got recorded more frequently. The most common type of recording for these studies is audio (39.35 %), followed by audio and video (9.53 %) and only video recording (9.33 %).

Compensation differs depending on the recording type

In figure 4.26, it can be seen that the median amount of compensation is highest for people recorded by audio and video ($\tilde{X}$ = 41.46 USD) and lowest for participants who were not recorded at all ($\tilde{X}$ = 10 USD). Participants who were either only audio or only video recorded, were compensated with a median amount of 24.57 USD and 21.64 USD respectively.

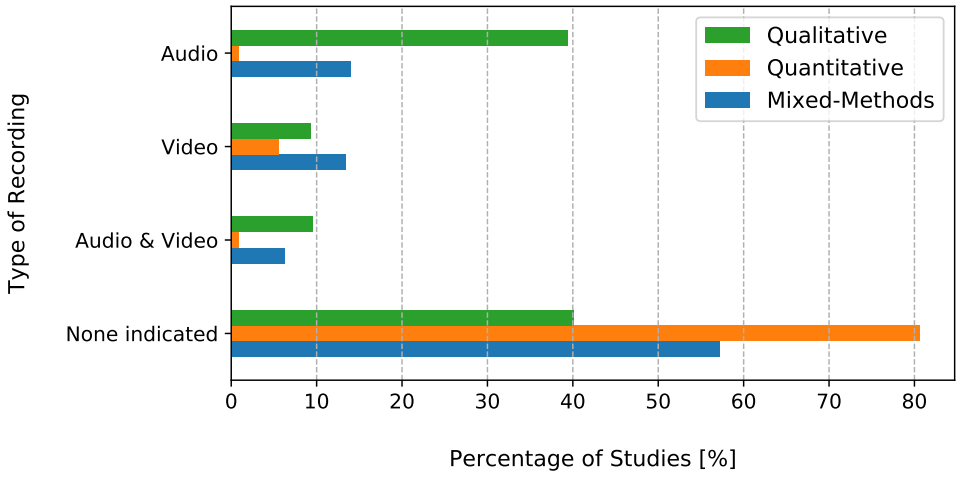

Figure 4.25: The Type of Recording and the Analysis Approach

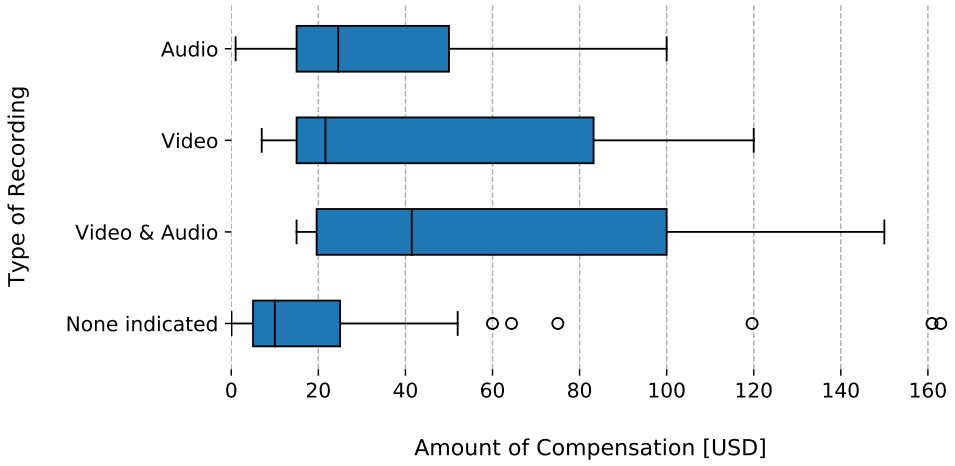

Figure 4.26: Amount of Compensation according to the Type of Recording

Chapter 5

Discussion

In the following, the findings of the analysis are discussed. Initially, the frequency of compensation is examined, followed by the discussion of the amount of compensation.

5.1 The Frequency of Compensation

It was found that the frequency of compensation differs for qualitative, quantitative and mixed-method studies. Several factors were investigated to find explanations for the fact that participants were compensated more frequently in quantitative and mixed-method studies than in qualitative ones.

Participants recruited via Amazon MTurk are most frequently compensated

First of all, it is not surprising that the number of participants differs according to the analysis approach. Caine already showed that the mean sample size is higher for quantitative than for qualitative studies [Caine, 2016]. Because of these disparities in sample size, the differences in the use of recruitment media among the analysis approaches were to be expected. Since on Amazon MTurk, a large number of participants can be recruited in a very short time [Kittur et al., 2008], it is not surprising that this is the most widely used recruitment medium for quantitative studies. On Amazon MTurk, participants get paid for every

HIT conducted [Hitlin, 2016]. Consequentially, it is obvious that the frequency of compensation is highest for participants recruited on the crowdsourcing website. Naderifar et al. pointed out that snowball sampling is a meaningful method in qualitative research [Naderifar et al., 2017], which has been confirmed for CHI 2019. It is the second most often used recruitment medium for qualitative studies. Only social media were used more often. Since the frequency of compensation is lowest for social media and snowball sampling and highest for Amazon MTurk, the recruitment medium can be seen as a factor that could explain the differences in compensation depending on the analysis approach.

Participation in privacy/security, visualizations/analytics and AI/machine learning related studies are most frequently compensated

As a second factor, the topic of the study was investigated. It was found that quantitative and mixed-method studies were most likely to address topics related to VR/AR/mixed reality, which is counted as one of the most popular topics for years [Liu et al., 2014]. In qualitative studies, on the other hand, mostly health/wellbeing related topics were addressed. A possible explanation for this finding is the high sensitivity of health data. Participants might be more likely to share this type of data in a personal study setting, i.e., in qualitative studies [Brown, 2003]. In both, studies related to VR/AR/mixed reality and studies related to health/wellbeing, participants were least likely to be compensated considering seven of the most addressed topics in CHI 2019. This indicates that the topic is not one of the main factors to be considered in order to explain the differences in compensation among the various analysis approaches. The highest frequency of compensation was found for studies related to privacy/security, visualizations/analytics and AI/machine learning. Studies addressing one of these topics were exactly the ones that were posted on Amazon MTurk most frequently, which highlights the recruitment medium as an important factor in explaining the differences in the frequency of compensation.

Participation in experiments and surveys is most frequently compensated

Regarding the method, it was found that experiments and surveys were mostly used in quantitative and mixed-method studies, whereas for example, workshops and observations were most often used in qualitative ones. These findings match the results of Caine because she reported

that the mean sample size for surveys and experiments is higher than the mean sample size for observations [Caine, 2016]. Considering the differences in sample size, her results indicate that surveys and experiments are primarily quantitative research methods while observations are primarily qualitative ones. Since participation in experiments and surveys was compensated way more frequently than participation in workshops and observations, the method is another factor that could explain the differences in compensation according to the analysis approach. The higher frequency of compensation for experiments and surveys compared to workshops and observations can be explained by the fact, that more people have to be recruited for these data collection methods. Since it might be difficult to find a large number of intrinsically motivated participants, incentives are used to improve the rate of participation [Fiore et al., 2014]. Hitlin already pointed out that Amazon MTurk is mainly used to conduct surveys and experiments [Hitlin, 2016], which has been confirmed for CHI 2019. For none of the other research methods, participants were recruited via Amazon MTurk. This shows that it cannot be stated that the method by itself is an explanation for the differences in compensation. There might be other underlying factors, like for example the recruitment medium, that influence the results.

Participation in online studies is most frequently compensated

Lastly, the study type was analyzed. It was found that online studies were mostly quantitative, whereas field studies were mostly qualitative. This also reflects the findings of Caine who found that the mean sample size is higher for remotely conducted studies than for in-person ones [Caine, 2016]. Since participants taking part in an online study were compensated most frequently and people participating in a field study were least likely to be compensated, the type of study can also be seen as a factor to explain the differences in compensation depending on the analysis approach. However, once more, it can be hypothesized that not the study type by itself is an explanation for the differences in compensation. One third of all online studies was posted on Amazon MTurk which is an explanation for the higher frequency of compensation for this type of study.

5.2 The Amount of Compensation

Furthermore, it was found that also the amount of compensation differs for qualitative, quantitative and mixed-method studies. Several factors were analyzed to find explanations for the fact that participants were compensated with the highest amount in qualitative and with the lowest amount in quantitative studies.

Negative correlation between the number of participants and the amount of compensation

It was already discussed that the number of participants differs according to the analysis approach, which was also shown in prior research [Caine, 2016]. The negative correlation between the number of participants and the amount of compensation can be explained by the fact that of a certain amount of money that can be used for compensating the participants, less is left for each individual when more people participate. The findings show that sample size is to be seen as a factor that could explain the differences in the amount of compensation depending on the analysis approach.

Participants recruited via snowball sampling are compensated with the highest amount

Furthermore it was discussed that the differences in the use of recruitment media among the analysis approaches can be explained by the different requirements for sample size. It was highlighted that Amazon MTurk is the most widely used recruitment medium for quantitative studies, whereas snowball sampling is often used for the recruitment in qualitative ones [Naderifar et al., 2017]. Since participants recruited via snowball sampling received the highest median amount of compensation and participants recruited via Amazon MTurk were compensated with the lowest median amount, the recruitment medium is not only a factor that could explain the differences in the frequency but also in the amount of compensation depending on the analysis approach. Kittur and Hitlin have already shown that participants recruited via Amazon MTurk usually receive a very small amount of compensation [Kittur et al., 2008, Hitlin, 2016]. In addition, Hitlin pointed out that most tasks on Amazon MTurk have a very short duration [Hitlin, 2016]. Indeed, this was also proven for CHI 2019. Studies in which Amazon MTurk is used have a way shorter median duration than studies using any other recruitment medium.

This implies that not the recruitment medium by itself is an explanation for the differences in compensation but that there are other underlying factors, like the duration of the study, that influence the results.

Positive correlation between the study duration and the amount of compensation

It was found that there is a strong correlation between the duration of a study and the amount of compensation, which reflects the findings of Latterman and Merz [Latterman and Merz, 2001]. Since qualitative and mixed-method studies have a longer median duration than quantitative ones, the duration is to be seen as another factor that could explain the differences in the amount of compensation among the analysis approaches.

Positive correlation between the number of components and the amount of compensation

In order to explain the differences in duration according to the analysis approach, it was investigated whether qualitative and mixed-method studies are more complex than quantitative ones. For this, the number of components was considered. The results show that mixed-method studies are more complex than qualitative and quantitative studies, which can be explained by the fact that mixed-method studies include methods to collect data for both, statistical and non-statistical analysis. A correlation between the number of components and the amount of compensation was already shown by Latterman and Merz [Latterman and Merz, 2001] and was also proven for CHI 2019. However, it is only a weak correlation which indicates that the number of components is not one of the main factors that could explain the differences in the amount of compensation.

Participation in workshops and focus groups is compensated with the highest amount

To examine whether it is less about the number but more about the type of components, the method was investigated. As discussed earlier, experiments and surveys were mostly used in quantitative and mixed-method studies, whereas for example, workshops and focus groups were most often used in qualitative ones. Since participants who took part in a workshop or focus group received the highest median amount of compensation and people who participated in an experiment or survey were compensated with the lowest median amount, the method is seen as another factor that can explain the differences in the amount of compensation depending on the analysis approach. The findings reflect the results of Latterman and Merz. They have

also shown that participants are compensated with a higher amount by taking part in a focus group than by filling out a survey [Latterman and Merz, 2001]. However, the median amount of compensation in focus groups more than doubled comparing the results from 2001 to the findings in CHI 2019. This might be due to the change of study characteristics over time. For example, Barkhuus and Rode stated that more recent studies often last longer [Barkhuus and Rode, 2007]. In order to explain the differences in the amount of compensation depending on the method used, the duration for the different methods was investigated. In fact, it was found that workshops and focus groups have a higher median duration than experiments and surveys. This indicates, that not the method by itself is an explanation for the differences in the amount of compensation but that the duration is an important factor to consider. Apart from the duration, the level of anonymity might be an explanation for the differences found. Participants might expect comparably high amounts of compensation for workshops and focus groups because taking part in one of these methods means revealing private information in front of other participants. In comparison, experiments and surveys allow for a higher anonymity.

Participants recorded by audio and video are compensated with the highest amount

Regarding the level of anonymity, recording is another factor to investigate. Recording was most often used in qualitative and least often used in quantitative studies. Participants recorded by audio and video received the highest median amount of compensation, followed by participants recorded by audio only or video only. In studies in which no recording was used, participants were compensated with the lowest median amount of compensation. The findings can be explained by considering the level of anonymity. Participants recorded by audio and video reveal a lot more information to the study director than participants who are not recorded at all. The use and type of recording can thus be seen as another factor that could explain the differences in compensation among the analysis approaches.

Participation in field and lab studies is compensated with the highest amount

A study is also more anonymous when it takes place online than when it takes place in the field or in the lab. It was already discussed that online studies are mostly quantita-

tive, whereas field studies are mostly qualitative. Since the median amount of compensation is higher for field and lab studies than for online ones, the study type is another factor that could explain the differences in compensation among the analysis approaches. However, once more, the duration of the study seems to be an influencing factor, since field and lab studies have a longer median duration than online studies. Apart from the duration and the level of anonymity, the findings can be explained by the fact that online studies allow for a higher flexibility. Participants do not have to be at a specific place at a specific time, which is the case for field and lab studies.

Participation in health/wellbeing related studies is compensated with the highest amount

Lastly, the topic of the study was investigated. It is noticeable that health/wellbeing is the only topic that was addressed way more often in qualitative than in quantitative or mixed-method studies. Brown already pointed out that qualitative research methods are especially important for health related topics [Brown, 2003]. Regarding the amount of compensation, people participating in health/wellbeing related studies received a higher amount than participants of studies that are related to other topics. This characterizes the topic of the study as another factor to explain the differences in compensation depending on the analysis approach. On the one hand, the comparatively high amount for health/wellbeing related studies can be explained by the fact that personal health data is highly sensitive. On the other hand, it was found, that health/wellbeing studies have a higher median duration than studies with other topics, again highlighting the duration as one of the main influencing factors.

5.3 Limitations of the Method

Findings rely on authors' statements

The main drawback of a literature review is the fact that findings rely on what the authors stated. For example, failure to indicate that incentives have been used does not mean that compensation has not been offered.

Extraction and classification of keywords

Furthermore, there are some limitations to the method used to extract and classify the keywords for the topic. First,

the keywords vary in number and detail between papers. Second, the classification was done manually which means that keywords may have been grouped that do not fit together as well as others, or keywords that fit perfectly into one of the groups may be missing.

Amounts of compensation were not normalized and partially converted

Regarding the amount of compensation, it should be taken into account that amounts have not been normalized based on the hourly wage of the respective country. In addition, the exchange rates have most likely changed within the five-month period of data collection. However, most amounts were given in USD, meaning that this limitation is probably not too significant.

Chapter 6

Summary and Future Work

6.1 Summary and Contributions

In summary, it was found that the differences in the frequency and amount of compensation according to the analysis approach can be explained by many factors that differ for qualitative, quantitative and mixed-method studies.

Participants recruited via Amazon MTurk are compensated most frequently

The frequency differs based on the recruitment medium, the topic, the method, and the type of study, whereby participants recruited via Amazon MTurk, participants who took part in privacy/security, visualizations/analytics or AI/machine learning related studies as well as participants who took part in an experiment, survey or online study were compensated with the highest frequency. Of these factors, the recruitment medium has proven to be one of the main influencing factors since it also explains the higher frequency of compensation for privacy/security, visualizations/analytics and AI/machine learning related studies as well as for experiments, surveys and online studies.

The highest amounts of compensation are offered in studies with longer durations

The amount of compensation differs based on the number of participants, the duration of the study, the number of components, the recruitment medium, the method,

the type of recording, the study type and the topic of the study. Higher amounts were offered for studies with fewer participants, longer durations and more components. Furthermore, participants were compensated with highest median amounts when they were recruited via snowball sampling, participated in a workshop or focus group, were recorded with audio and video, took part in a field study or participated in a study related to health/wellbeing topics. Of these factors, the duration has proven to be one of the main influencing factors since workshops and focus groups, as well as field studies and studies related to health/wellbeing are those with the longest median duration compared to studies that use other methods and types or address different topics.

Contributions of the thesis

Researchers in the field of HCI could benefit from these findings by considering them when planning the recruitment in future studies. They could use the results as an orientation and for example, adjust the amount of incentives to the length and complexity of the study while also taking other factors into consideration. On the other hand, they can question the findings and, if applicable, consciously deviate from current practices, which in the best case could lead to an improvement of the recruitment process. By reusing the source code, other conferences can be taken into account for the critical evaluation or to reveal trends over time.

6.2 Future Work

Analysis of the type of compensation

Apart from investigating the frequency and amount of compensation, the data set can be used to give insights about current practices concerning the type of compensation (e.g., money, voucher, gift cards). Several factors could be investigated to explain possible differences in the types used. In addition, the data set allows for the analysis of other factors that have not been considered in this thesis, e.g., the age of the participants. This way, further explanations for the findings can be found which would lead to a broader overview of current practices. It could be considered to conduct a multiple regression to get a better impres-

sion of the degree of influence of single factors.

Machine learning algorithms for the keyword extraction and classification

To get around the limitations of the method used to extract the topic of a study, machine learning algorithms should be applied for both, keyword extraction and classification. For the keyword extraction, it would be best to consider larger parts of the papers or even the paper as a whole in order to obtain a greater number of keywords with the same level of detail.

Reusable source code to analyze other conferences

In further research, it would also be interesting to investigate other conferences using the same source code to compare the results with the findings of this thesis. This way, it might be possible to reveal trends over time or to identify differences to other fields of research.

Bibliography

Louise Barkhuus and Jennifer A Rode. From mice to men-24 years of evaluation in chi. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, volume 10, 2007.

Patrick Biernacki and Dan Waldorf. Snowball sampling: Problems and techniques of chain referral sampling. *Sociological methods & research*, 10(2):141–163, 1981.

Phil Brown. Qualitative methods in environmental health research. *Environmental Health Perspectives*, 111(14):1789–1798, 2003.

Kelly Caine. Local standards for sample size at chi. In *Proceedings of the 2016 CHI conference on human factors in computing systems*, pages 981–992, 2016.

Tina Christensen, Anders H Riis, Elizabeth E Hatch, Lauren A Wise, Marie G Nielsen, Kenneth J Rothman, Henrik Toft Sørensen, and Ellen M Mikkelsen. Costs and efficiency of online and offline recruitment methods: a web-based cohort study. *Journal of medical Internet research*, 19 (3):e58, 2017.

Andrew T Fiore, Coye Cheshire, Lindsay Shaw Taylor, and GA Mendelsohn. Incentives to participate in online research: an experimental examination of" surprise" incentives. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 3433–3442, 2014.

Yvonne Garbers and Udo Konradt. The effect of financial incentives on performance: A quantitative review of individual and team-based financial incentives. *Journal of occupational and organizational psychology*, 87(1):102–137, 2014.

Christine Grady et al. Payment of clinical research subjects. *The Journal of clinical investigation*, 115(7):1681–1687, 2005.

Kasper M Hansen. The effects of incentives, interview length, and interviewer characteristics on response rates in a cati-study. *International Journal of Public Opinion Research*, 19(1):112–121, 2007.

Kayla Herbell and Jaclene A Zauszniewski. Facebook or twitter?: Effective recruitment strategies for family caregivers. *Applied Nursing Research*, 41:1–4, 2018.

Paul Hitlin. Research in the crowdsourcing age, a case study. *Pew Research Center*, 11, 2016.

Gary Hsieh and Rafał Kocielnik. You get who you pay for: The impact of incentives on participation bias. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, pages 823–835, 2016.

Forrest Huang, John F Canny, and Jeffrey Nichols. Swire: Sketch-based user interface retrieval. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–10, 2019.

Stephen Hutt, Joseph F Grafsgaard, and Sidney K D'Mello. Time to scale: Generalizable affect detection for tens of thousands of students across an entire school year. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–14, 2019.

Ainsley James, Bev Taylor, and Karen Francis. Researching with young people as participants: Issues in recruitment. *Contemporary nurse*, 47(1-2):36–41, 2014.

Bridget Kelly, Marjorie Margolis, Lauren McCormack, Patricia A LeBaron, and Dhuly Chowdhury. What affects people's willingness to participate in qualitative research? an experimental comparison of five incentives. *Field Methods*, 29(4):333–350, 2017.

Aniket Kittur, Ed H Chi, and Bongwon Suh. Crowdsourcing user studies with mechanical turk. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 453–456, 2008.

Malcolm Koo and Harvey Skinner. Challenges of internet recruitment: a case study with disappointing results. *Journal of Medical Internet Research*, 7(1):e6, 2005.

Jessica Latterman and Jon F Merz. How much are subjects paid to participate in research? *American journal of bioethics*, 1(2):45–46, 2001.

Yong Liu, Jorge Goncalves, Denzil Ferreira, Bei Xiao, Simo Hosio, and Vassilis Kostakos. Chi 1994-2013: mapping two decades of intellectual progress through co-word analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 3553–3562, 2014.

Yuewen Liu and Juan Feng. Does money talk? the impact of monetary incentives on user-generated content contributions. *The Impact of Monetary Incentives on User-Generated Content Contributions (May 10, 2016)*, 2016.

Arminda Guerra Lopes. Using research methods in human computer interaction to design technology for resilience. *JISTEM-Journal of Information Systems and Technology Management*, 13(3):363–388, 2016.

Qingguo Ma, Jia Jin, Liang Meng, and Qiang Shen. The dark side of monetary incentive: how does extrinsic reward crowd out intrinsic motivation. *Neuroreport*, 25(3): 194–198, 2014.

Winter Mason and Duncan J Watts. Financial incentives and the" performance of crowds". In *Proceedings of the ACM SIGKDD workshop on human computation*, pages 77–85, 2009.

Mahin Naderifar, Hamideh Goli, and Fereshteh Ghaljaie. Snowball sampling: A purposeful method of sampling in qualitative research. *Strides in Development of Medical Education*, 14(3):1–6, 2017.

Julie Pallant. Survival manual. *A step by step guide to data analysis using SPSS*, 2011.

Maxine X Patel, Victor Doku, and Lakshika Tennakoon. Challenges in recruitment of research participants. *Advances in Psychiatric Treatment*, 9(3):229–238, 2003.

Elizabeth Ripley, Francis Macrina, Monika Markowitz, and Chris Gennings. Who's doing the math? are we really compensating research participants? *Journal of Empirical Research on Human Research Ethics*, 5(3):57–65, 2010.

Anam Sohail. Challenges in recruiting participants for studies in hci. `https://hci.rwth-aachen.de/publications/sohail2020a.pdf`, 2020. Accessed: 2020-09-12.

Jaime Waters. Snowball sampling: A cautionary tale involving a study of older drug users. *International Journal of Social Research Methodology*, 18(4):367–380, 2015.

Index

Typeset September 14, 2020