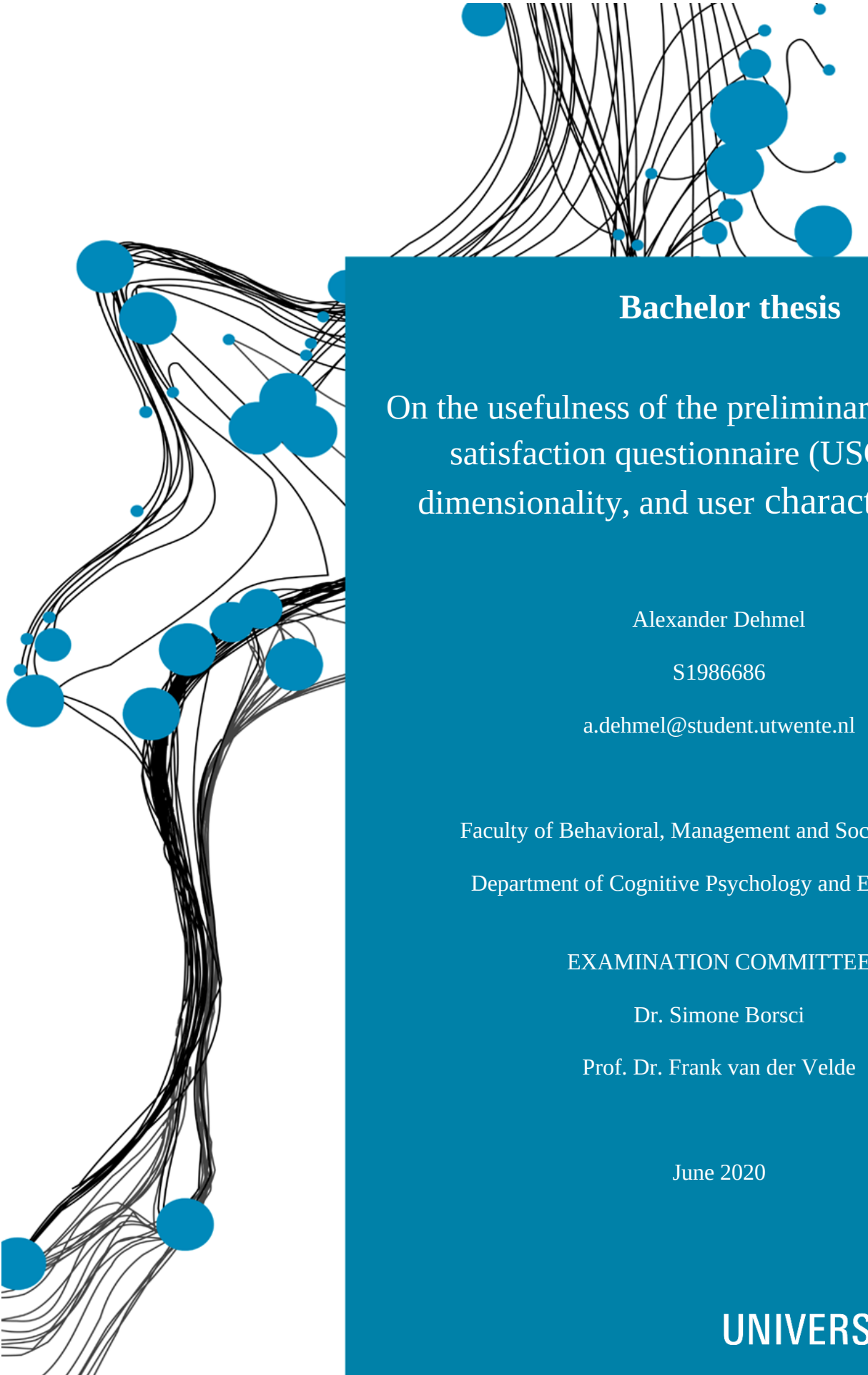

Bachelor thesis

On the usefulness of the preliminary usability satisfaction questionnaire (USQ), its dimensionality, and user characteristics

Alexander Dehmel

S1986686

a.dehmel@student.utwente.nl

Faculty of Behavioral, Management and Social Sciences

Department of Cognitive Psychology and Ergonomics

EXAMINATION COMMITTEE

Dr. Simone Borsci

Prof. Dr. Frank van der Velde

June 2020

UNIVERSITY OF TWENTE.

Abstract

This study investigated the ability of the preliminary Usability Satisfaction Questionnaire (USQ), to measure user satisfaction with chatbots. We explored its concurrent validity by conducting a correlational analysis between the USQ and an established questionnaire of usability, the UMUX-LITE. Furthermore, we ran a principal component analysis to investigate its dimensionality and proposed a condensed version that we compared with previous results. Lastly, we investigated the impact of participants' gender, first-time usage of chatbots, geekism, and institution-based trust on the scores of the USQ via linear regression analyses. For this purpose, thirty-nine participants, mainly students from the University of Twente, had to interact with five randomly assigned chatbots by solving two information-retrieval tasks and then rated the chatbots' usability. Twenty-four of the participants also filled out the institution-based trust and geekism questionnaires. We found a positive correlation between the USQ and the UMUX-LITE, a PCA suggested a 5-component structure with 32 items, and linear regression analyses revealed no significant effects of the four independent variables. The results further contributed to the reasoning that the preliminary questionnaire can be seen as a suitable basis to develop a standardised measurement tool of chatbot usability due to demonstrations of its psychometric abilities.

Keywords: chatbots, usability, user satisfaction, questionnaire, psychometrics

Table of contents

1. Introduction 5
1.1 The rise of chatbots *5*
1.2 Previous work *8*
1.3 The aim of this study *9*
2. Methods 12
2.1 Participants *12*
2.2 Materials *13*
2.3 Procedure *14*
2.4 Data analysis *15*
3. Results 18
3.1 Correlation between USQ and UMUX-LITE *18*
3.2 Principal component analysis of the USQ *19*
3.3 Linear regression of demographic characteristics *25*
4. Discussion 27
4.1 Main findings *27*
4.2 Limitations *33*
4.3 Future recommendations *34*
4.4 Conclusion *35*
References 36
Appendix A 45
Revised list of chatbot features *45*
Appendix B 47
Preliminary Usability Satisfaction questionnaire *47*
UMUX-LITE *52*
Appendix C 53
Qualtrics Survey Flow *53*
Appendix D 88
Geekism scale *88*
Institution-based trust questionnaire *90*

Appendix E 94
Spss-Syntax *94*
Appendix F 105
Informed Consent (before Covid-19) *105*
Informed Consent via Qualtrics (during Covid-19) *106*
Appendix G 109
Chatbot tasks *109*
Appendix H 113
Oblique rotated factor loadings *113*
The final version of oblique rotated factor loadings *115*

1. Introduction

1.1 The rise of chatbots

In recent years, a new trend has shaped the landscape of human-computer interaction. Chatbots are technical dialogue systems which are capable of communicating with a user despite the absence of any human operators (McTear, 2017). Companies from various industries such as education or online-marketing make use of this technology to provide customers with information, guidance, or to sell their products (Ciechanowski, Przegalinska, Magnuski, & Gloor, 2018). It appears that chatbots have partially replaced human customer services - further enhanced by their constant availability without any time restrictions (Brandtzaeg & Følstad, 2017; Hald, 2018). This increasing relevance across industries can be related to recent technological advances, namely in artificial intelligence, deep learning, or natural language processing (McTear, 2017). Additionally, users have become increasingly familiar with mobile messaging applications (Brandtzaeg & Følstad, 2017). This is relevant because a significant amount of chatbots is modelled around the idea of messaging interfaces to provide users with the opportunity of expressing themselves via a written format - just like chatting with a friend (Jain, Kota, Kumar, & Patel, 2018a; Przegalinska, Ciechanowski, Stróż, Gloor, & Mazurek, 2019).

Researchers worked on the phenomenon of chatbots well before the rise of recent technology. Dating back to 1966, the chatbot *ELIZA* emulated a Rogerian psychotherapist and tried to trick users into thinking that they are talking to an actual human being than artificial intelligence (Jain et al., 2018a; Przegalinska et al., 2019). Seventy years have passed since Alan Turing introduced the Turing test to determine a machines' cognitive abilities. If a certain amount of "judges" fail to distinguish the machine's performance from a real human being's, the program passes as "intelligent". Several experts like famous cognitivist Noam Chomsky

criticised Turing's reasoning for equating a simulation of human communicative abilities with intelligence, as Chomsky considered it only as one aspect of cognition (Chomsky, 2009). While the question of machines' "intelligence" has riddled philosophers, psychologists, and engineers for decades, substantial attention has been paid by developers to improve chatbot's abilities of "conversational intelligence" (Nilsson, 2009; Jain, Kumar, Kota, & Patel, 2018b).

Nevertheless, many chatbots fail to interact convincingly with users. One example was Microsoft's *Tay* which mimicked users' speech patterns to obtain advanced levels of sophistication. However, its adoption of inappropriate and insulting language of users led to its shutdown after just 16 hours (Brandtzaeg & Følstad, 2018). *Tay* is not the only example of a chatbot failing to create the notion of real human interaction, as many chatbots at this point are struggling to react "appropriately" to a given situation (Brandtzaeg & Følstad, 2018).

Two factors could explain this problem. Firstly, many chatbots are not capable of interpreting users' input with sufficient consideration of the context and previously said statements (McTear, 2017). A reason for this might be that many programs just follow simplified if-else statements, which are organised around a database (Khanna et al., 2015). Secondly, it is not always possible to predict how people might react to the chatbot's output or what they consider as a "good" conversation. It turns out that a substantial amount of users reported signs of frustration and scepticism because their needs and expectations were not met by the machine, which could explain the still existing preference of many people to rely on the interaction with a real human being (Araujo, 2018).

A considerable corpus of literature has addressed the question of what the needs of users are. According to Jenkins, Churchill, Cox, and Smith (2007), people prefer chatbots to be helpful and efficient in terms of information processing, as well as capable of concise use of language.

This is consistent with Brandtzaeg and Følstad (2018) who pointed out that users are more interested in effective deliverance of information and chatbots' abilities to solve problems instead of them employing realistic avatars or pretending to be human beings. A phenomenon further contributing to this reasoning is the so-called uncanny valley effect, which suggests that photorealistic designs of robots do not necessarily add to the amount of given sympathy and satisfaction by the users, but often raise doubts (Mori, MacDorman, & Kageki, 2012). Mori et al. (2012) have proposed one explanation of this effect: The expectations of the machine as a sophisticated, almost human-like program were violated, once its communicative limitations came to light, leading to frustration and repudiation. Overall, efficient and transparent communication of the machine's abilities seems to be vital for its success (Dybkjær & Bernsen, 2001).

A consequence of this is the importance to measure how well chatbots are meeting these expectations. While some measurements of usability already exist, they differ significantly across industries (Przegalinska et al., 2019). Many researchers consider the length and structure of a conversation between a user and a chatbot as an essential marker of usability, while others emphasise the chatbot's ability to provide personalised and relevant dialogue (Przegalinska et al., 2019). Maroengsit et al. (2019) mentioned several other practices like content evaluation of the chatbot's responses, expert evaluations, or methods based around user satisfaction measures. The latter has become a popular methodology in human-computer interaction. Users give feedback by rating their experience of the interaction with the application (Macleod, Bowden, Bevan, & Curson, 1997). The most often used tools for this purpose are questionnaires, which also differ by utilisation of, for instance, 3-point Likert scales or open-ended questions (Morris, Kouddous, Kshirsagar, & Schueller, 2018; Skjuve et al., 2019).

This variety of measurements suggest a lack of standardisation to assess chatbots' communicative abilities. This is a severe downside because standardised measurements across industries would be beneficial in terms of replicability and objectivity (Sauro & Lewis, 2012). Furthermore, many researchers perceive standardised assessment tools as more reliable than unstandardised ones (Hornbæk, 2006). Several instruments for assessing general usability have been developed for a range of contexts. Whether it is the System Usability Scale, the CSUQ, the UMUX as well as its shorter version, the UMUX-LITE – they all have shown signs of sufficient reliability and validity across samples and domains (Balaji & Borsci, 2019). However, according to Tariverdiyeva and Borsci (2019), these questionnaires miss the ability to provide diagnostic insights into relevant aspects and factors of chatbot interaction. Therefore, they suggested the need to develop a tool tailored explicitly for user satisfaction concerning chatbots.

1.2 Previous work

Consequently, Tariverdiyeva and Borsci (2019) conducted a literature review to obtain features which they considered relevant for a measurement tool of chatbot usability. They came up with an initial list of 18 essential features, which was later reduced to 14 by consequent works and additionally assessed with a focus group study (e.g. Balaji & Borsci, 2019) (see Appendix A). Ultimately, this resulted in a preliminary questionnaire to measure user satisfaction with chatbots - the Usability Satisfaction Questionnaire (USQ), which consists of 42 items (see Appendix B).

For the USQ to be a standardised assessment tool, special attention has to be paid to its psychometric qualities. One repeatedly employed strategy is the questionnaire's degree of correlation with already established measurements to demonstrate indications of its concurrent validity (Berkman & Karahoca, 2016). For example, a consequent study by Boecker and Borsci

(2019) reported a significant correlation between the USQ and the UMUX-LITE (see Appendix B). They rated this as an essential insight for the questionnaire's development as a usability measurement tool.

To compensate for other questionnaires' lack of sufficient, diagnostic insights into aspects of user satisfaction with chatbots (see Balaji & Borsci 2019), previous studies have also paid attention to the USQ's dimensional structure. Such analyses of dimensionality, for example, factor or principal component analysis, are valuable for the development of a questionnaire by revealing insights of underlying constructs, thus demonstrating construct validity (Brown, 2010).

For instance, Balaji and Borsci (2019) conducted exploratory as well as confirmatory factor analyses with the suggestion for a 4-factor solution, while Waldera and Borsci (2019) yielded a 9-factor model containing 25 items. In contrast, Boecker and Borsci (2019) conducted a principal component analysis and proposed a 5-component structure with 27 items.

1.3 The aim of this study

In this study, we wanted to replicate the findings of previous studies by conducted a correlational analysis between the USQ and UMUX-LITE to provide further confidence in the questionnaire's validity and applicability to assess user satisfaction for the domain of chatbots. Therefore, the first research question was the following: *1. What is the relationship between the scores of the USQ and the UMUX-LITE for assessing the interaction with chatbots?*

Besides, we conducted a principal component analysis to explore the dimensionality of the USQ, propose a condensed version of the questionnaire, and critically discuss the results in comparison to previous findings. Thus, the second research question asked: 2. *What are the underlying dimensions of the Usability Satisfaction Questionnaire in comparison to previous studies?*

Besides, we wanted to investigate the USQ's sensitivity, which so far has been disregarded by previous studies. A standardised questionnaire of user satisfaction across various samples and industries should be sensitive to existing differences between chatbot systems without being too much affected by other variables (Cairns, 2013). This is especially important for the domain of human-computer interaction, where not the differences between users, but differences between systems should be the main emphasis (Berkman & Karahoca, 2016). Therefore, the third part of this research was the exploration of the impact of four different variables on the USQ scores.

First of all, participants' gender was repeatedly tested for its impact on questionnaires like the System Usability Scale and the UMUX-LITE (Bangor, Kortum, & Miller, 2008). Furthermore, while the majority of previous studies suggests an interplay between chatbots' and users' gender (see Nass, Moon, & Green, 1997), it has been stated by Hsiao-Chen and Yi-Chieh (2019) to pay attention specifically to the impact of users' gender, as it plays an essential role in the interaction with chatbots. Therefore, the third research question asked: *3. What is the effect of participants' gender on the scores of the Usability Satisfaction Questionnaire?*

In line with Jain et al. (2018) that 84 % of internet users have never interacted with a chatbot before, we also investigated the impact of first-time users regarding their scores of the USQ. These users have shown more signs of frustration during their initial encounters with chatbots, which might indicate the importance of familiarity for user satisfaction measures (Hackbarth, Grover, & Yi, 2003). Furthermore, for both UMUX and UMUX-LITE, significant effects of users' familiarity with the system have been found (Berkman & Karahoca, 2016). This might be relevant for a usability questionnaire since participants' scores could be the result of their experience with the software instead of a measure of usability satisfaction. Therefore, we

also considered a fourth research question: *4. What is the effect of first-time usage on scores of the Usability Satisfaction Questionnaire?*

In comparison, those users who are highly familiar with chatbots and technology might be equally interesting for questionnaire development. So-called geeks are technologically enthusiastic people who do not use a system solely to reach a goal but also experiment and interact with it in a "playful" manner (Schmettow, Noordzij, & Mundt, 2013). For them, technology becomes a significant object of interest which could be important for usability scores as the initial tool to reach a goal becomes the goal itself. These participants might react differently to "challenging" systems, driven by their intrinsic interest in technology (Schmettow et al., 2013). Therefore, an overly complicated chatbot might be perceived as tedious for a first-time user, but a geek could see it as a "challenge" to be solved, which could influence their USQ scores. The fifth research question thus asked: *5. What is the effect of geekism on scores of the Usability Satisfaction Questionnaire?*

The last aspect of this study concerned the context of chatbots. They are not isolated pieces of technology but embedded within a specific environment, for instance, a company's website (Araujo, 2018). McKnight, Choudhury, and Kacmar (2002) reported that the average user had declined the provision of personal information at least once due to significant distrust towards a website or respective vendor. Many usability studies focus on a micro-level analysis by conceptualising communication as a process between two individuals. At the same time, the embedding environment, in this case, the internet, is often only perceived as a contributing factor (Bachmann & Inkpen, 2011). That could be problematic because a negative bias regarding, for example, sharing private data might play an essential role in perceived trust towards a system (Bachmann & Inkpen, 2011). McKnight et al. (2002) emphasised the importance of the whole

sociological domain of the internet, which they conceptualised as "institution-based trust". This construct is more than a measure of trust towards specific internet vendors but instead describes users' perception of the internet as a whole. Such an impact could influence participants to perceive a chatbot in a certain way, not only due to its inherent qualities but because of past experiences with websites. Thus, the sixth research question asked: *6. What is the effect of institution-based trust on USQ scores?*

Overall, the objective of this study was built upon previous findings and explored the USQ's relationship with an established measurement tool of general usability, the questionnaire's dimensional structure, and its psychometric sensitivity for assessing chatbot usability by investigating the impact of four different variables. A questionnaire which shows signs of psychometric quality is an essential step towards building a consistent and standardised measurement tool of user satisfaction towards chatbots (Berkman & Karahoca, 2016).

2. Methods

2.1 Participants

We recruited 39 participants using the "SONA" system of the University of Twente as well as convenience sampling. This participant pool consisted of two different sets: While 24 participants were recruited from the lead researcher of this study, we also used the data of 15 people from a comparable study by Neumeister and Borsci (2020)[1], who conducted similar research. These 39 participants consisted of 19 males and 20 females with a mean age of 25.77, and the respective nationalities were German (N = 30), Dutch (N = 6), German-Dutch (N = 1), English (N = 1), and French (N = 1). The only restrictions for participation were a minimum age

[1] While Neumeister and Borsci (2020) also aimed to replicate previous findings concerning the USQ, they also investigated the impact of the belief that a chatbot is controlled by a human being and used deceptive elements.

of 18 and a sufficient understanding of the English language. The participants who were recruited via the SONA system received two credits as an incentive.

2.2 Materials

For the procedure of this study, we used Qualtrics, a program which allows creating surveys of various kinds. It contained all relevant questionnaires, tasks, and links of the study (see Appendix C). To ensure replicability, its structure was mostly resembling the survey from Boecker and Borsci (2019).

We used four questionnaires for this study. The main objective was the Usability Satisfaction Questionnaire (USQ), a preliminary questionnaire consisting of 42 items with a 5-point Likert scale to measure the perceived usability of a chatbot. The scores range from 42 to 210. Also, we implemented the UMUX-Lite, a 2-item questionnaire with raw scores between 0 and 100 to quickly evaluate a system's perceived usability (Lewis, Utesch, & Maher, 2013). Besides we used the Geekism questionnaire, a 15-item questionnaire with a 5-point Likert scale measuring the enthusiasm of users towards technology as well as the Institution-based trust questionnaire which consists of 15 items and uses a 5-point Likert scale to measure the trust of users towards the internet (see Appendix D).

In addition to these questionnaires, the survey contained a demographic scale to gather data regarding participants demographic backgrounds like age, gender, and experience with chatbots. We also made use of various chatbots from the study by Boecker and Borsci (2019). However, as three of the previous chatbots were not working at the beginning of the study, we had to integrate three new chatbots as a replacement. Furthermore, two chatbots stopped working during the data collection phase, which therefore had to be replaced as well. Overall, a pool of 11 chatbots from different websites was available for every participant (see Appendix G).

The analysis of the data has been done with the help of the statistical program SPSS using descriptive techniques as well as relevant inferential statistics. Appendix E gives an overview of the respective syntax. Lastly, since the COVID-19-pandemic occurred shortly after the beginning of the study, we had to change the initial face-to-face meetings in the library of the University of Twente into a digital format. We used Skype for this purpose to enable communication with the participants. The program also allowed us to record the screen for potential future qualitative analyses. This way of communication was possible since the Qualtrics survey could be still used in its original form.

2.3 Procedure

Before the study started, we had to get approval from the universities ethical committee. Initially, the study took place in a library room of the University of Twente. However, we later had to change the procedure into a digital format via Skype due to the COVID-19-pandemic[2]. A study session took around one hour and was guided by the Qualtrics survey. After the participants gave written consent for voluntary participation (see Appendix F) and agreed to the recording of the screen, they filled in a demographic survey and a rating of their familiarity with chatbots. We presented them two different tasks and a link to a specific website containing the chatbot (see Appendix G). The tasks were mostly about information retrieval and a means to let the user interact with a chatbot. For every participant, five of the chatbots were randomly assigned with the help of the "randomiser" function of Qualtrics. Once they finished the tasks by either having them solved or giving up, users were asked to rate the tasks' difficulties and to fill out the USQ and UMUX-LITE to evaluate their satisfaction with the chatbot. After repeating these steps five times, the geekism and institution-based trust questionnaires were filled out by

[2] The COVID-19 disease is initiated by the Corona-virus "SARS-CoV-2" and caused a pandemic at the beginning of 2020, leading to various measures of caution like restrictions of face-to-face meetings or mobility.

the 24 participants who were recruited from the lead researcher of this study alone, since Neumeister and Borsci (2020) did not investigate geekism and institution-based trust.

While we mainly adopted the survey structure from Boecker and Borsci (2019), there were two noteworthy differences. Firstly, participants had to solve two tasks instead of one to increase the time spent per chatbot, and to collect more data for the assessment of chatbots. This was in line with Balaji and Borsci (2019) who reported that one task alone might not be enough to allow for sufficient interaction with the chatbots. Furthermore, Borsci, Federici, Bacci, Gnaldi and Bartolucci (2015) reported an effect of the time that users spend with a system on the outcomes of usability assessment tools. Secondly, for cases of websites' or chatbots' malfunctioning, the survey would allow us to skip the current chatbot and to offer a replacement. The same would be applied for those participants without a Facebook account, as they would have been incapable of interacting with three of our which were embedded in Facebook (see Appendix C). Such a feature was especially useful for the digital continuation of the study via Skype, as the participants were able to "skip" a chatbot themselves without extra effort from the researcher's side.

2.4 Data analysis

Before analysing the data with SPSS, we rescaled the raw scores for both UMUX-LITE and USQ between 0 and 1 for compatibility purposes. Furthermore, we reverted items 10 and 11 because the agreement to a statement like *"I had to rephrase my input multiple times for the chatbot to be able to help me"* seemed to represent something negative in terms of chatbot interaction. We considered this to be important as the majority of items were oriented towards a more positive direction to measure users' satisfaction with chatbot's.

The initial step of the analysis was the exploration of the relationship between the rescaled scores of the USQ and the UMUX-LITE to establish an indication of the USQ's concurrent validity (Cairns, 2013). We chose a correlational analysis for this purpose and checked the assumption of normality by conducting a Shapiro-Wilk test to decide which correlation coefficient would be appropriate for the data set. Depending on this, we applied either a Pearson correlation or Kendall's Tau. The results were then tested for statistical significance by calculating 97.5% confidence intervals using bootstrapping with 9999 replicates.

Besides, we conducted a principal component analysis to explore the questionnaire's dimensionality and to make suggestions for a condensed version. While this was in line with Boecker and Borsci (2019), it contrasted prior studies which used factor analysis. However, Preacher & MacCallum (2003) have pointed out that both analyses are suitable for the exploration of the underlying dimensional structure as well as data reduction purposes. Especially principal component analysis can be beneficial for the latter and provides valuable insights for the questionnaire's construct validity (Cairns, 2013; Goldberg, 1990). However, despite replicating the study by Boecker and Borsci (2019), we decided to exclude the results of the previous focus group study, because a PCA is mainly a measure based on linear item combinations, instead of making a priori assumptions, for instance, to decide to not remove certain features before the actual analysis (Jolliffe & Cadima, 2016).

Initial considerations concerned the PCA's appropriateness for the given data and the number of extracted components. The Kaiser-Meyer-Olkin Criterion (KMO) should be at least 0.5 to be acceptable (Kaiser, 1974). Furthermore, Bartlett's test of sphericity should be statistically significant to justify the continuation of the principal component analysis. The number of extracted components depended on the Kaiser criterion to consider only those with

Eigenvalues bigger than 1. We further consulted a scree plot, but only as additional insight, as it has been criticised for its subjective nature (Osborne & Costello, 2005; Hayton, Allen, & Scarpello, 2004). The last decision marker concerned the rotation of the analysis. An oblique rotation (oblimin) was used similarly to Boecker and Borsci (2019) since components in the social sciences are almost always assumed to correlate with each other to some degree. Therefore, orthogonal rotations might result in a loss of information (Costello & Osborne, 2005).

During the analysis, we removed items with a communality under .2 from the analysis, as those seem to be might not be sufficiently explained by underlying components (Costello & Osborne, 2005). Additionally, following Field (2013), we suppressed item loadings less than .3 at the start of the analysis. We considered a primary item loading of less than .5 as a reasonable cut-off point and removed those items which "crossloaded" with at least .4 on two different dimensions (Costello & Osborne, 2005; Howard, 2016). Lastly, whole components which did not contain at least three items exceeding a minimum loading of .5 were removed (Costello & Osborne, 2005). After conducting the principal component analysis, we computed the reliability for each of the obtained components by conducting Cronbach's Alpha as a measure of internal consistency (Schmitt, 1996). We considered a value of at least .7 as acceptable and deleted those items, whose removal would increase a scale's reliability (Blunch, 2008).

Lastly, we explored the impact of the variables gender, first-time usage, geekism, and institution-based trust on the USQ scores with simple linear regression analyses and tested the significance of the results via bootstrapping with 97.5 % confidence intervals. For this, we created the variable first-time usage by specifying participants as first-time users if they responded to the variable "prior usage" with "probably not" or "definitely not" (see Appendix C). In cases of uncertainty, the variable "familiarity" served as an additional decision marker.

Furthermore, gender and first-time usage were dummy-coded with male participants and first-time users as reference groups. Additionally, we checked relevant model assumptions of normality, linearity, and heteroscedasticity with the help of normal probability plots of residuals for the predicted variable and scatterplots for residual errors. While the assumption of independence was technically not met due to five repeated responses by every participant, we accepted this because studies have suggested that for repeated measures with all values of the independent variable being equal for every subject, the linear regression analysis still yields interpretable results without significant loss of information (Donner, 1984).

3. Results

3.1 Correlation between USQ and UMUX-LITE

Overall, 39 participants filled out the USQ and the UMUX-LITE five times, except for one participant who only interacted with four chatbots, resulting in 194 responses. No outliers were found to be excluded from the data set. The relevant descriptives like mean, standard deviation, minimum, and maximum of the responses are summarised in table 1. The scores for the USQ ranged from 96 to 196 ($M = 154.81$, $SD = 24.37$). The UMUX-LITE had a range between 0 and 100 with $M = 71.5$, $SD = 24.95$. The rescaled equivalents of all scores were ranging between 0 and 1.

None of the data was found to be normally distributed with Shapiro-Wilk, $W = .971$, $p < .01$, which led to the usage of Kendall's Tau as a correlational measure between the UMUX-LITE and the USQ.

Table 1

Descriptive statistics

Questionnaire	Type of score	Range	M	SD	Min.	Max.
UMUX-LITE	Raw scores	[0;100]	71.5	24.95	12.5	100
	Rescaled scores	[0;1]	.71	.25	.13	1
USQ	Raw scores	[42;210]	154.81	24.37	96.00	196.00
	Rescaled scores	[0;1]	.67	.14	.32	.92

Based on the results of the analysis, the two questionnaires correlated with $r = 0.71$, $p < .01$. The bootstrapping with 9999 samples confirmed the significance of the results, with 97.5 % [.65, .76].

3.2 Principal component analysis of the USQ

A principal component analysis with oblimin rotation was computed for all 42 items of the USQ. The Kaiser-Meyer-Olkin Criterion, KMO = .88 verified sampling adequacy. Besides, Bartlett's test of sphericity x^2 (861) = 5517.23, $p < .001$ was statistically significant, and the communalities for the majority of items were way over .3, which we considered as acceptable. The Kaiser criterion confirmed an initial 10-component solution as best fit, accounting for 72.08 % of the variance. This was backed up by a scree plot, even though a 3- or 5-component solution was also a possible interpretation based on visible "elbows" (figure 1). Therefore, since the scree plot showed some ambiguity, the Kaiser criterion of Eigenvalues over 1 led to the decision to extract ten components.

However, the pattern matrix of the output revealed that components 5, 6, 8, 9, and 10 were not containing a minimum of three items with loadings of at least .5. Therefore, we removed these components from the analysis. The resulting 5-component solution still explained

56. 5 % of the variance, but contained several items either not loading high enough on their primary component, having high "crossloadings", or no loadings at all (see Appendix H). Therefore, these items were removed one after another.

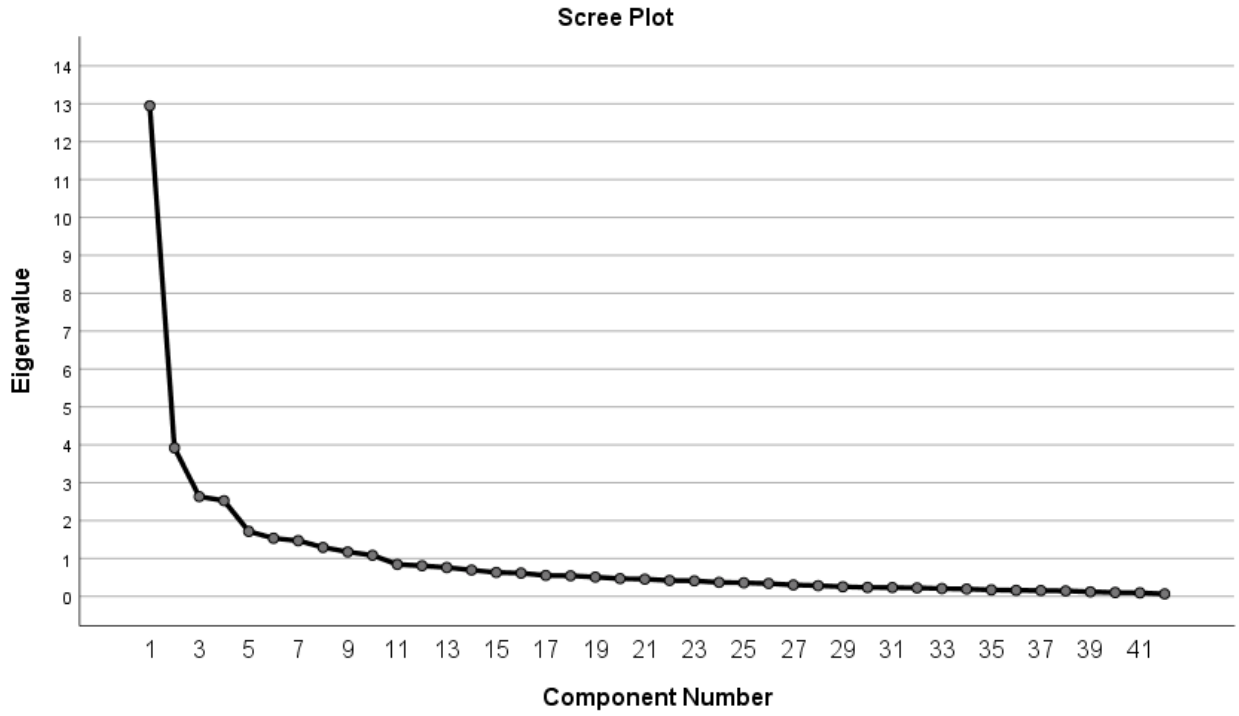

Figure 1. Scree plot of the PCA for 42 items

After eleven repetitions, a final 5-component solution had been found with 32 items all loading higher than .5 on their primary component (see Appendix H). In the process, items 7, 8, 9, 10, 11, 12, 15, 17, 18, and 36 were deleted. Three items were "crossloading" without being removed because their primary loadings were higher than .5 and the alternative loadings did not exceed .4. Consequent checks of internal consistency showed that most scales had a sufficient Cronbach's alpha, $\alpha = .7$ or higher, except the fifth one being below, $\alpha = .68$. The only possible improvement could have been made for the fourth component with, $\alpha = .83$, by removing item 20, leading to an increased value of, $\alpha = .89$. However, this would have resulted in the

component's deletion due to fewer than three items with loadings over .5. Therefore, the reliability of α =.83 was considered to be sufficient, and the item was not deleted.

The final results (see table 2) suggested a 5-component structure with the first component (items 16, 22, 23, 24, 25, 26, 27, 28, 29, 30, 34, 35, 37, 38, and 39) called "*quality and quantity of information*". We decided this because those items featuring *"maxim of relation"*, *"relevant information", "relevant service", "recognition and facilitation of goal", "understandability"* and *"perceived credibility"* seemed to describe conversational quality while quantity was represented via items labelled as *"maxim of quantity"*). In similar fashion like Boecker and Borsci (2019), we called the second component *"ease of getting started"* which was represented by items 1 to 6 featuring "*visibility*" and *"ease of getting started"*. Component three was labelled *"response time"*, similar to the feature represented by the three items 40, 41, and 42. We did the same for the fourth component *"perceived privacy and security"* with items 19, 20, and 21. The fifth component was called *"keeping track of context"* and included the features *"graceful responses"*, *"ongoing conversation"*, and *"awareness of context"* which were represented by the items 13, 14, 31, 32, and 33.

Table 2

Labels of components[a]

Components	Item	text of item	Feature
Quality and quantity of information	USQ_28	The amount of received information was neither too much nor too less.	Maxim of quantity
	USQ_29	The chatbot gives me the appropriate amount of information.	Maxim of quantity
	USQ_25	The chatbot gave relevant information during the whole conversation.	Maxim of relation
	USQ_26	The chatbot is good at providing me with a helpful response at any point of the process.	Maxim of relation
	USQ_30	The chatbot only gives me the information I need.	Relevant information
	USQ_27	The chatbot provided relevant information as and when I needed it.	Relevant information
	USQ_39	It appeared that the chatbot provided accurate and reliable information.	Perceived credibility
	USQ_37	I feel like the chatbot's responses were accurate.	Perceived credibility

	USQ_38	I believe that the chatbot only states reliable information.	Perceived credibility
	USQ_22	I felt that my intentions were understood by the chatbot.	Recognition and facilitation of goal
	USQ_23	The chatbot was able t guide me to my goal.	
	USQ_24	I find that the chatbot understands what I want and helps me to achieve my goal.	
	USQ_34	I found the Chatbot's responses clear.	Understandability
	USQ_35	The chatbot only states understandable answers.	
	USQ_16	The chatbot guided me to the relevant service.	Relevant Service
Ease of getting started	USQ_4	The chatbot was easy to access.	Visibility
	USQ_5	The chatbot's function was easily detectable.	Visibility

	USQ_6	It was easy to find the chatbot.	Visibility
	USQ_2	It was easy for me to understand how to start the interaction with the chatbot.	Ease of starting a conversation
	USQ_1	It was clear how to start a conversation with the chatbot.	
	USQ_3	I find it easy to start a conversation with the chatbot.	
Response time	USQ_40	The time of the response was reasonable.	Response time
	USQ_41	My waiting time for a response from the chatbot was short.	Response time
	USQ_42	The chatbot is quick to respond.	Response time
Perceived privacy and security	USQ_19	The interaction with the chatbot felt secure in terms of privacy.	Perceived privacy and security
	USQ_20	I believe the chatbot informs me of any possible privacy issues.	Perceived privacy and security
	USQ_21	I believe that this chatbot maintains my privacy.	Perceived privacy and security
Keeping track of context	USQ_13	The interaction with the chatbot felt like	Ongoing conversation

	an ongoing conversation.	
USQ_14	The chatbot was able to keep track of context.	Ability to maintain themed discussion
USQ_31	The chatbot could handle situations in which the line of conversation was not clear.	Graceful responses
USQ_32	The chatbot explained gracefully when it could not help me.	
USQ_33	When the chatbot encountered a problem, it responded appropriately.	

[a] labels mainly taken from Boecker and Borsci (2019)

3.3 Linear regression of demographic characteristics

Overall, we registered 194 responses for the relevant independent variables of the linear regression analyses (see table 3). Gender was distributed with 99 female and 95 male responses. Furthermore, 74 replies were registered to be provided by first-time users. Besides, we listed 120 responses for the variables geekism and institution-based trust from the 24 participants who were recruited specifically for this study. The geekism scores ranged from -25.00 to 22 with M = -1, SD = 10.35, while the scores for institution-based trust varied between 36.00 and 90.00 with M = 67.33, SD = 16.44.

Table 3

Demographic variables

Variable	Responses	Mean	SD	Min	Max
Male	95				
Female	99				
First-time user	74				
Non-first-time user	120				
Geekism	120	-1.00	10.35	-25.00	22.00
Institution-based trust	120	67.33	16.44	36.00	90.00

A normal probability plot of residuals for the predicted variable and the scatterplot of residuals against the predicted values indicated that the assumptions of normality, homoscedasticity, and linearity were met.

For gender, the regression equation was found to be not significant, $F(1,192) = .4$, $p = .525$) with $R^2 = .002$. Participants' predicted USQ score was equal to 153.72 + 2.23 when the participants were male with 97.5% bootstrapping [-5.56, 10.11], which suggested that male participants had 2.23 higher scores in comparison to female participants.

For first-time usage, the regression equation was found to be not significant, $F(1, 192) = .004$, $p = .951$) and $R^2 = .004$. Therefore, participants' predicted USQ scores were equal to 154.72 + .22 with 97.5 % bootstrapping [-7.83, 8.1] when treated as first-time users. This suggests that first-time users score .22 higher regarding their USQ score than non-first-time users.

For geekism, no significant regression was found, $F(1, 118) = 1.31$, $p = .254$), with an $R^2 = .011$. Participants' predicted USQ score was equal to 155.17 - .25 on the geekism scale with 97.5 % bootstrapping [-.74, .22]. Therefore, for every decrease in geekism, the USQ scores dropped with .25.

Regarding institution-based trust, the regression also predicted no significant effect on the USQ scores, $F(1, 118) = 1.05$, $p = .308$) and $R^2 = .009$. USQ scores were equal to 145.79 - .14 in institution-based trust with 97.5 % bootstrapping [-.14, .44], which suggested that for every decrease in institution-based trust, the USQ scores drop with a slope of -.14.

4. Discussion

4.1 Main findings

The first research question asked about the relationship between the scores of the USQ and UMUX-LITE, which turned out to be a positive correlation. This is an indication of the questionnaire's criterion validity by comparing it with an established measurement of general usability (Cairns, 2013; Lewis, Utesch, & Maher, 2013). Cairns (2013) emphasised the importance of validity for a new questionnaire. Thus, uncertainty whether the USQ measures usability would be a severe downside for its development. However, this study, as well as previous endeavours like Boecker and Borsci (2019), delimited such concerns. That is especially important for the assessment of chatbots. Cameron et al. (2018) have conceptualised them as a new type of interface in comparison to traditional systems due to chatbots' interactive nature. New interfaces require new methods of measurement, as established questionnaires might not be sufficient to explore all relevant aspects of the interaction between users and the system (Holmes et al., 2019). The current findings are contributing to this endeavour and confirm that the

preliminary questionnaire can be used as a basis to establish a standardised measurement for the assessment of chatbots.

The second research question asked about the underlying dimensions of the Usability Satisfaction Questionnaire, which resulted in the proposal of a condensed 5-component version with 32 items. Consequent reliability analyses suggested sufficient internal consistency for all components. Despite the significant overlap, the proposed component structure also differed in some regard to prior findings, which are presented in table 4.

Table 4

Dimensionality propositions of previous studies

Boecker and Borsci (2019)		Balaji and Borsci (2019)		Waldera and Borsci (2019)	
Component	Items	Factor	Items	Factor	Items
General usability	8, 10, 11, 12, 14, 22, 23, 24, 26, 27, 29, 31, 37	Response Quality	7, 15, 18, 24, 25, 30, 33, 34, 37	Perceived credibility, implementation & understanding the User's intent	16, 17, 18, 23, 24, 37, 38, 39
Ease of getting started	2, 3, 4, 5, 6	Communication Quality	1, 2, 4, 5, 10, 11	Accessibility & Starting the conversation	1, 2, 3, 4, 5, 6
Perceived privacy and security	19, 20, 21	Perceived privacy	21	Perceived Privacy & Security	19, 20, 21
Response time	40, 41, 42	Perceived Speed	41	Response time	40, 41, 41

Articulateness	33, 35, 36	Handling unexpected situations	32
		Expectation setting	8
		Ability to maintain themed discussion	13
		Understandability	35
		Flexibility of Linguistic input	11

The component *"perceived privacy"* (items 19, 20, and 21) seemed to describe the ability of the chatbot to maintain a quality conversation in terms of privacy concerns. While most studies came up with an identical solution, Balaji and Borsci (2019) recommended only to use item 21 *"I believe that this chatbot maintains my privacy"*. The length of the preliminary questionnaire might justify this suggestion to avoid repetitiveness and boredom by users (Wanous, Reichers, & Hudy, 1997). Further proof of this approach is a critical assessment of item 20 *"I believe the chatbot informs me of any possible privacy issues"* as a double-barreled item. According to Vellis (1991), double-barreled items describe more than one concept and should be avoided due to the difficulty of interpretations. Item 20 could be interpreted to represent both the chatbot's ability to make privacy-related statements as well as the existence of any privacy-related issues as such. Furthermore, reliability analyses suggested the removal of item 20. Overall, we agree with Balaji and Borsci (2019) to reduce this component's item structure to some extent, even though our proposed dimension mostly resembled prior findings.

The component *"response time"* (items 40, 41, and 42) showed excellent reliability and was repeatedly proposed throughout studies. A dimension that considers the time to give an appropriate response is also supported by literature since past research implies that users prefer chatbots that are efficient in terms of information processing (Brandtzaeg & Følstad, 2018).

Therefore, Balaji and Borsci (2019) suggested only to use item 41 *"the chatbot is quick to respond"*, which is conceptually similar to item 42. Both ask whether a chatbot delivers quick responses. However, it might be more suitable to consider item 40 *"The time of the response was reasonable"* as this component's representation because it does not just provide a measure of speed, but an assessment of the response's appropriateness. While the chatbot should not take too long to formulate an output, a quick response time alone will not necessarily increase perceived usability (Gnewuch, Morana, & Maedche, 2017). Therefore, the findings suggest a dimension of *"response time"*, but future studies need to figure out whether this should be a measure of speed or the response's appropriateness.

Another component was called *"ease of getting started"* with the features *"visibility"* and *"ease of starting a conversation"*. While there is significant overlap with previous findings, Balaji and Borsci (2019) decided to combine these features with items 10 and 11, both assessing "*flexibility of linguistic input"*. We removed these two items because they had low component loadings. However, this might have been problematic because it led to the removal of the feature *"flexibility of linguistic input"* which is difficult to justify since the necessity to rephrase your input can be considered a potential source of frustration (Hackbarth et al., 2003). Hence, it might be advisable to keep the two items in the questionnaire. However, combining them with other items of the dimension *"ease of getting started"* can be seen as critical because a chatbot's accessibility and visibility have been reported as essential in terms of user satisfaction and even whether a chatbot is used at all (Kuligowska, 2015; Følstad, Nordheim, & Bjørkli, 2018). This seems to be different in comparison to a chatbot's ability to react with flexibility to users' input. Therefore, it is proposed to assume a component which mainly assesses the chatbot's accessibility and visibility before the actual conversation. Additionally, future research is

required to conceptualise a suitable place for items 10 and 11 in the preliminary questionnaire's dimensionality.

While the first three components showed considerable overlap with prior findings, the fourth component *"keeping track of context"* varied to some degree. Including the features "*graceful responses", "ongoing conversation"*, and *"ability to maintain themed discussion"*, it suggests an underlying dimension, which describes the ability of the chatbot to react appropriately to the given context. Furthermore, items 22 and 24 (*"recognition and facilitation of goal"*), as well as item 23 *("relevant service"*), were crossloading on this component. Such results have not been present in previous studies. For instance, Boecker and Borsci (2019) suggested a component called *"articulateness"* including items featuring *"graceful responses"* and *"understandability"*. They justified this decision by emphasising the importance of unambiguous communication patterns during chatbot interaction (Gnewuch, Morana, Adam, & Maedche, 2018). However, a chatbot's understanding often depends on the context (Kirakowski, Odonnell, & Yiu, 2009). Without such a given context, for instance, the users' goals, their direct input, or the website's content, every statement of the user would be only analysed in isolation (Jain et al., 2018a). Brandtzaeg and Følstad (2018) stated that users' goals, as well as the relevance of the chatbot's service, are vital factors to consider in chatbot usability. The component's reliability of less than .7 certainly raises questions but should not be overinterpreted either, since modest reliabilities are reasonable to work within the beginning stages of questionnaire development (Nunally, 1978). Therefore, there seem to be implications for future studies to explore the possibility of a context-based dimension.

Lastly, the component *"quality and quantity of information"* pointed towards a measure of content quality. The first noticeable observation was the considerable amount of 15 items

overall. Robinson (2017) suggests that the right number of items per scale depends on the balance between parsimony and sufficient framework coverage, which implies the necessity to shorten the item structure of this component to some extent. For instance, the feature *"perceived credibility"* could be represented by only using one or two items. Besides, there was a significant overlap with Balaji and Borsci (2019) who labelled the respective factor as *"response quality"*, thus ending up with a comparable interpretation. That contrasts Boecker and Borsci (2019) who proposed a vaguer component *"general usability"*. Nevertheless, they also emphasised the need for future studies to explain the interplay of the item structure for such "general usability". The uncertainty of this dimension across studies becomes even more evident, given the amount of variation of features. For example, previous studies included the features "*expectation setting"* and *"graceful response"* for this specific dimension, while elements within the current study were missing in prior works. This suggests that despite some agreement of a qualitative dimension, unclarity remains what this quality represents.

The last part of this study concerned research questions three to six which explored the impact of participants gender, first-time usage, geekism, and institution-based trust on the USQ scores. We found no significant influence for either variable, which provides further proof of the questionnaire's suitability to measure chatbot usability without being too sensitive towards other factors or constructs (Cairns, 2013). This notion of sensitivity is essential for standardised measurements of usability (Berkman & Karahoca, 2016).

We based our decision to consider gender and first-time usage by following the proceedings of other questionnaires. For example, the UMUX-LITE was unaffected by participants gender but showed signs of sensitivity towards the user's experience (Berkman & Karahoca, 2016). The finding of this study suggests that the preliminary questionnaire might be

suitable for both genders. Regarding the familiarity of users, we conceptualised the dichotomous variable first-time usage, based on the suggestion to consider people, who have never used a chatbot before, as they might react differently in comparison to more familiar users (Jain, 2018b). However, experience with a system captures a range of levels beyond the sole difference between familiar and unfamiliar users. Therefore, we only explored the impact of one aspect of "chatbot-expertise". Nonetheless, we have indicated the questionnaire's suitability for users without any chatbot experience as well as more experienced ones. Additionally, the results regarding geekism and institution-based trust were promising, as they added towards the questionnaire's quality of sensitivity because neither an interest in technology nor a bias towards the internet seemed to affect the USQ scores.

4.2 Limitations

Sample and selection bias. Due to its replicable nature, the study was based on similar samples as previous endeavours, in this case, students from the University of Twente and convenience sampling. While this is useful in terms of replicability (see Asendorpf et al., 2013), it also creates challenges regarding the generalisability of results. Besides, non-significant findings for variables like geekism could be explained by a lack of discriminative ability. An alternative might have been to recruit people who explicitly consider themselves as geeks. Second, it might have been useful to not only look out for geeks but instead to find people who show high interest in chatbots specifically.

Violation of independence. During the linear regression analyses, we violated the assumption of independence by treating all five assessments of every participant as an individual response. We based this decision on given literature like Donner (1984). However, it is still a violation and therefore listed as a limitation of this study.

COVID-19. Due to the outbreak of COVID-19, we had to change the initially intended form to an online version. While participants were still able to assess the chatbots in the intended manner and communicated with the lead researcher via Skype, a potential side effect on perceived usability satisfaction regarding the chatbots is assumed to be minor. No theoretical framework suggests that assessing a chatbot at home or in a library room would have such a statistically significant impact on the USQ scores in terms of psychometric measures. In real life, we would expect that most people interact with a chatbot from their home instead of a public place like a library. Therefore, an online format could even come closer to real encounters between users and chatbots. However, it is still listed as a limitation, since our agenda was a replication of previous studies and thus should have happened with similar conditions.

Interaction time with chatbots. The current approach to "enforce" interaction with chatbots entailed users to solve tasks. Balaji and Borsci (2019) raised concerns about this approach as it would be hindering to explore a chatbot in its entirety, especially when only having one task at hand. While this study as well as Neumeister and Borsci (2020) both tried to facilitate this by increasing the number of tasks, it could have still been a hindering factor for the assessment of user satisfaction with chatbots, as two tasks might still provide not enough interaction time to explore a chatbot adequately.

4.3 Future recommendations

We have demonstrated concurrent validity by providing evidence of the USQ's correlation with the UMUX-LITE. This was a useful decision, especially given the length of the preliminary questionnaire. The UMUX-LITE offers a brief assessment of usability and is thus convenient as a complementary tool next to the longer USQ to avoid users becoming tired during the process, which might have affected the results (Wanous et al., 1997). We still advise future

studies to explore the USQ's relationship with other established measurements of usability across different samples to provide more insights into its concurrent validity. We also demonstrated signs of construct validity by exploring the USQ's dimensional structure as it is necessary to have a clear understanding of the underlying dimensions and which items are essential (Brown, 2010). However, despite a considerable amount of overlap across studies, differences still exist. How important is the dimension of the context for chatbot interaction? Is the dimension *"response time"* a notion of quickness or its "appropriateness" according to participants subjective experience? What do the dimensions *"response quality"* or *"general usability"* entail? These questions have to be answered by future studies.

4.4 Conclusion

The USQ aims to be a multifaceted tool that covers all relevant aspects of chatbot usability. This requires stable psychometric qualities like validity, reliability, and sensitivity (Cairns, 2013). Several studies have demonstrated evidence for these performance indices by finding correlations with established measurements, exploring the questionnaire's dimensional structure, as well as components' internal consistency. This study replicated relevant results but also pointed out some differences. Furthermore, it explored the questionnaire's sensitivity to assess chatbot usability by testing the impact of four different variables. The development of the Usability Satisfaction Questionnaire is far from over, as we expect future changes regarding, for example, its item structure. Nonetheless, the preliminary questionnaire is based on a strong foundation with sound psychometric qualities. Therefore, we consider the USQ as a compelling candidate to be a standardised measurement of chatbot usability.

References

Araujo, T. (2018). Living up to the chatbot hype: The influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions. *Computers in Human Behavior, 85*, 183-189. doi:10.1016/j.chb.2018.03.051

Asendorpf, J., Conner, M., De Fruyt, F., De Houwer, J., Denissen, J., Fiedler, K., ... Wicherts, J. (2013). Recommendations for Increasing Replicability in Psychology. *European Journal of Personality, 27*, 108-119. doi:10.1002/per.1919.

Bachmann, R., & Inkpen, A. C. (2011). Understanding Institutional-based Trust Building Processes in Inter-organizational Relationships. *Organization Studies, 32*(2), 281–301. doi:10.1177/0170840610397477

Balaji, D., & Borsci, S. (2019). *Assessing User Satisfaction with Information Chatbots: A Preliminary Investigation* (Master's thesis). Retrieved from University of Twente Student Theses. (77182).

Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An Empirical Evaluation of the System Usability Scale. *International Journal of Human-Computer Interaction, 24*(6), 574–594. doi:10.1080/10447310802205776

Berkman, M., & Karahoca, D. (2016). Re-Assessing the Usability Metric for User Experience (UMUX) Scale. *Journal of Usability Studies, 11*, 89-109. Retrieved from https://uxpajournal.org/de/assessing-usability-metric-umux-scale/

Blunch, N. J. (2008). *Introduction to structural equation modelling using SPSS and AMOS.* Thousand Oaks, CA: Sage Publications Ltd.

Borsci, S., Federici, S., Bacci, S., Gnaldi, M., & Bartolucci, F. (2015). Assessing user satisfaction in the era of user experience: Comparison of the SUS, UMUX, and UMUX-

LITE as a function of product experience. *International Journal of Human-Computer Interaction, 31*(8), 484–495. doi:10.1080/10447318.2015.1064648

Brandtzaeg, P., & Følstad, A. (2018). Chatbots: changing user needs and motivations. *Interactions*, *25*, 38-43. doi:10.1145/3236669

Brandtzaeg, P., & Følstad, A. (2017). *Why people use chatbots*. Pager presented at the Fourth International Conference on Internet Science (INSCI), Thessaloniki, Greece. Abstract retrieved from https://www.researchgate.net/publication/318776998_Why_people_use_chatbots

Brown, J.D. (2010). How are PCA and EFA used in language test and questionnaire development? *Jalt, 14*(2), 30-35. Retrieved from http://hosted.jalt.org/test/PDF/Brown33.pdf

Boecker, N. & Borsci, S. (2019). *Usability of information-retrieval chatbots and the effects of avatars on trust* (Bachelor's thesis). Retrieved from University of Twente Student Theses. (78097).

Cairns, P. (2013). A commentary on short questionnaires for assessing usability. *Interacting with Computers*, *25*(4), 312–316. doi:10.1093/iwc/iwt01

Cameron, G., Cameron, D., Megaw, G., Bond, R., Mulvenna, M., O'Neill, S., … McTear, M. (2018). *Back to the Future: Lessons from Knowledge Engineering Methodologies for Chatbot Design and Development.* Paper presented at the 32nd International BCS Human Computer Interaction Conference (HCI), Belfast, Ireland. doi: 10.14236/ewic/HCI2018.153

Chomsky, N. (2009). Turing on the "Imitation Game". In R. Epstein, G. Roberts, & G. Beber (Eds.), *Parsing the Turing Test*, (pp. 103–106). doi:10.1007/978-1-4020-6710-5_7

Ciechanowski, L., Przegalinska, A., Magnuski, M., & Gloor, P. (2018). In the Shades of the Uncanny Valley: An Experimental Study of Human-Chatbot Interaction. *Future Generation Computer Systems*. doi:10.1016/j.future.2018.01.055

Costello, A. B., & Osborne, J. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation, 10*, 1-9. doi:10.4135/9781412995627.d8

Donner, A. (1984). Linear regression analysis with repeated measurements. *Journal of Chronic Diseases, 37*(6), 441–448. doi:10.1016/0021-9681(84)90027-4

Dybkjær, L., & Bernsen, N. O. (2001). *Usability evaluation in spoken language dialogue systems*. Paper presented at the Workshop on Evaluation for Language and Dialogue Systems - Volume 9, Toulouse, France. doi: https://doi.org/10.3115/1118053.1118055

Field, A. (2013). *Discovering Statistics using IBM SPSS Statistics*: Sage Publications Ltd.

Følstad, A., Nordheim, C.B., & Bjørkli, C.A. (2018). What Makes Users Trust a Chatbot for Customer Service? An Exploratory Interview Study. In S. Bodrunova (Ed.), *Lecture Notes in Computer Science: Vol. 11193. The Fifth International Conference on Internet Science (INSCI)* (pp. 194-208). doi:10.1007/978-3-030-01437-7_16

Gnewuch, U., Morana, S., & Maedche, A. (2017). *Towards Designing Cooperative and Social Conversational Agents for Customer Service*. Paper presented at the Proceedings of the International Conference on Information System (ICIS), Seoul, South Korea. Retrieved from https://www.researchgate.net/publication/320015931_Towards_Designing_Cooperative_and_Social_Conversational_Agents_for_Customer_Service

Gnewuch, U., Morana, S., Adam, M. T. P., & Maedche, A. (2018). *Faster is Not Always Better: Understanding the Effect of Dynamic Response Delays in Human-Chatbot Interaction.* Paper presented at the 26th European Conference on Information Systems (ECIS), Portsmouth, United Kingdom. Retrieved from https://www.researchgate.net/publication/324949980_Faster_Is_Not_Always_Better_Understanding_the_Effect_of_Dynamic_Response_Delays_in_Human-Chatbot_Interaction

Goldberg, L. R. (1990). An alternative "Description of personality": The Big-Five factor structure. *Journal of Personality and Social Psychology, 59,* 216-1229. doi:10.1037//0022-3514.59.6.1216

Hackbarth, G., Grover, V., & Yi, M. (2003). Computer playfulness and anxiety: Positive and negative mediators of the system experience effect on perceived ease of use. *Information & Management, 40,* 221-232. doi:10.1016/S0378-7206(02)00006-X

Hald, G. (2018, February 16). 7 Benefits of using chatbots to drive your business goals [Web log post]. Retrieved from https://medium.com/botsupply/7-benefits-of-using-chatbots-to-drive-your-businessgoals-5a3a5e809951.

Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor Retention Decisions in Exploratory Factor Analysis: A Tutorial on Parallel Analysis. *Organizational Research Methods, 7,* 191-205. doi:10.1177/1094428104263675

Holmes, S., Moorhead, A., Bond, R., Zheng, H., Coates, V., & McTear, M. (2019). *Usability testing of a healthcare chatbot: Can we use conventional methods to assess conversational user interfaces?* Paper presented at the 31st European Conference, Belfast, UK. doi: 10.1145/3335082.3335094

Hornbæk, K. (2006). Current practice in measuring usability: Challenges to usability studies and research. *International Journal of Human-Computer Studies, 64*(2), 79–102. doi:10.1016/j.ijhcs.2005.06.002

Howard, M. C. (2016). A Review of Exploratory Factor Analysis Decisions and Overview of Current Practices: What We Are Doing and How Can We Improve? *International Journal of Human–Computer Interaction, 32*(1), 51-62. doi:10.1080/10447318.2015.1087664

Hsiao-Chen, Y., & Yi-Chieh, C. (2019). *The Effects of Chatbot Gender on User Trust and Perception towards Shopping Chatbots.* Paper presented at the Asian Conference on the Social Sciences, Tokyo, Japan. Retrieved from: https://25qt511nswfi49iayd31ch80-wpengine.netdna-ssl.com/wp-content/uploads/papers/acss2019/ACSS2019_51692.pdf

Jain, M., Kota, R., Kumar, P., & Patel, S. N. (2018a). *Convey: Exploring the Use of a Context View for Chatbots*. Paper presented at the Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Montreal QC, Canada. doi:10.1145/3173574.3174042

Jain, M., Kumar, P., Kota, R., & Patel, S. N. (2018b). *Evaluating and Informing the Design of Chatbots*. Paper presented at the Proceedings of the 2018 Designing Interactive Systems Conference, Hong Kong, China. doi:10.1145/3196709.3196735

Jenkins, M.-C., Churchill, R., Cox, S., & Smith, D. (2007). Analysis of User Interaction with Service Oriented Chatbot Systems. Paper presented at the Human-Computer Interaction. HCI Intelligent Multimodal Interaction Environments, Berlin, Heidelberg. doi:10.1007/978-3-540-73110-8_9

Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 374*(2065), 20150202. doi:10.1098/rsta.2015.0202

Kaiser, H. F. (1974). An Index of Factorial Simplicity. *Psychometrika, 39,* 31-36. doi:10.1007/BF02291575

Khanna, A., Pandey, B., Vashishta, K., Kalia, K., Bhale, P., & Das, T. (2015). A Study of Today's A.I. through Chatbots and Rediscovery of Machine Intelligence. *International Journal of u- and e-Service, Science and Technology, 8,* 277-284. doi:10.14257/ijunesst.2015.8.7.28

Kuligowska, K. (2015). Commercial Chatbot: Performance Evaluation, Usability Metrics and Quality Standards of Embodied Conversational Agents. *Professionals Center for Business Research, 2*(02), 1-16. doi:10.18483/pcbr.22

Lewis, J. R., Utesch, B. S., & Maher, D. E. (2013). UMUX-LITE. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13. New York, USA: ACM Press. doi:10.1145/2470654.2481287

Macleod, M., Bowden, R., Bevan, N., & Curson, I. (1997). The music performance measurement method. *Behaviour and Information Technology, 16*, 1-27. doi:10.1080/014492997119842

Maroengsit, W., Piyakulpinyo, T., Phonyiam, K., Pongnumkul, S., Chaovalit, P., & Theeramunkong, T. (2019). *A Survey on Evaluation Methods for Chatbots*. Paper presented at the Proceedings of the Seventh International Conference on Information and Education Technology, Aizu-Wakamatsu, Japan. doi:10.1145/3323771.3323824

McKnight, D. H., Choudhury, V., & Kacmar, C. (2002). Developing and Validating Trust Measures for e-Commerce: An Integrative Typology. *Information Systems Research, 13*(3), 334-359. doi:10.1287/isre.13.3.334.81

McTear, M. F. (2017). The Rise of the Conversational Interface: A New Kid on the Block? Future and Emerging Trends in Language Technology. *Machine Learning and Big Data, 38–49.* doi:10.1007/978-3-319-69365-1_3

Mori, M., MacDorman, K. F., & Kageki, N. (2012). The Uncanny Valley [From the Field]. *IEEE Robotics & Automation Magazine, 19*(2), 98-100. doi:10.1109/MRA.2012.2192811

Morris, R. R., Kouddous, K., Kshirsagar, R., & Schueller, S. M. (2018). Towards an Artificially Empathic Conversational Agent for Mental Health Applications: System Design and User Perceptions. *Journal of medical Internet research, 20*(6), e10148. doi:10.2196/10148

Nass, C., Moon, Y., & Green, N. (1997). Are Machines Gender Neutral? Gender-Stereotypic Responses to Computers With Voices. *Journal of Applied Social Psychology, 27*(10), 864-876. doi:10.1111/j.1559-1816.1997.tb00275.x

Neumeister, S., & Borsci, S. (2020). *Testing of a usability assessment tool for chatbots: Investigating the effect of believing that a chatbot is a human* (Unpublished bachelor's thesis). University of Twente, Netherlands.

Nilsson, N. J. (2009). *The Quest for Artificial Intelligence*. New York: Cambridge University Press.

Kirakowski, J., Odonnell, P., & Yiu, A. (2009). Establishing the Hallmarks of a Convincing Chatbot-Human Dialogue. In I. Maurtua (Ed.), *Human-Computer Interaction.* Retrieved

from https://www.intechopen.com/books/human-computer-interaction/establishing-the-hallmarks-of-a-convincing-chatbot-human-dialogue

Preacher, K. J., & MacCallum, R. C. (2003). Repairing Tom Swift's Electric Factor Analysis Machine. *Understanding Statistics, 2*(1), 13–43. doi:10.1207/s15328031us0201_02

Przegalinska, A., Ciechanowski, L., Stróż, A., Gloor, P., & Mazurek, G. (2019). In bot we trust: A new methodology of chatbot performance measures. *Business Horizons, 72*(6), 799-797. doi:10.1016/j.bushor.2019.08.005

Robinson, M. A. (2017). Using multi-item psychometric scales for research and practice in human resource management. *Human Resource Management, 57*(3), 739–750. doi:10.1002/hrm.21852

Sauro, J., & Lewis, J. R. (2012). Chapter 8 - Standardized Usability Questionnaires. In J. Sauro & J. R. Lewis (Eds.), *Quantifying the User Experience* (pp. 185-240). Boston: Morgan Kaufmann.

Schmettow, M., Noordzij, M. L., & Mundt, M. (2013). *An implicit test of UX: Individuals Differ in What they Associate with Computers*. Paper presented at the Proceedings thirty-first annual CHI conference on Human factors in computing systems, Paris, France. Abstract retrieved from https://www.researchgate.net/publication/235950931_An_Implicit_Test_of_UX_Individuals_Differ_in_What_They_Associate_with_Computers

Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment, 8*(4), 350-353. doi:10.1037/1040-3590.8.4.350

Tariverdiyeva, G., & Borsci, S. (2019). *Chatbots' Perceived Usability in Information Retrieval Tasks: An Exploratory Analysis* (Master's thesis). Retrieved from University of Twente Student Theses. (66604).

Waldera, L., & Borsci, S. (2019). *Development of a Preliminary Measurement Tool of User Satisfaction for Information-Retrieval Chatbots* (Bachelor's thesis). Retrieved from University of Twente Student Theses. (56604).

Wanous, J. P., Reichers, A. E., & Hudy, M. J. (1997). Overall job satisfaction: How good are single-item measures? *Journal of Applied Psychology*, *82*(2), 247–252. doi:10.1037/0021-9010.82.2.247

Appendix A

Table A1

Revised list of chatbot features

Features	Description of features
Ease of starting a conversation	How easy it is to start interacting with the chatbot
Accessibility	The ease with which the user can access the chatbot
Expectation setting	The extent to which the chatbot sets expectations for the interaction with an emphasis on what it can and cannot do
Communication effort	The ease with which the chatbot understands a range of user input
Ability to maintain themed discussion	The ability of the chatbot to maintain a conversational theme once introduced and keep track of context
Reference to service	The ability of the chatbot to make references to the relevant service
Perceived privacy	The extent to which the user feels the chatbot protects one's privacy
Recognition and facilitation of user's goal and intent	The ability of the chatbot to understand the user's intention and help them accomplish their goal
Relevance	The ability of the chatbot to provide information that is relevant and appropriate to the user's request
Maxim of quantity	The ability of the chatbot to respond in an informative way without adding too much information
Graceful breakdown	The ability of the chatbot to respond appropriately when it encounters a situation it cannot handle
Understandability	The ability of the chatbot to communicate clearly and in an easily understandable manner
Perceived credibility	The extent to which the user believes the chatbot's

	responses to be correct and reliable
Perceived speed	The ability of the chatbot to respond timely to user's requests

Appendix B

Table B1

Preliminary Usability Satisfaction questionnaire

Statement	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
It was clear how to start a conversation with the chatbot.					
It was easy for me to understand how to start the interaction with the chatbot.					
I find it easy to start a conversation with the chatbot.					
The chatbot was easy to access.					
The chatbot function was easily detectable.					
It was easy to find the chatbot.					
Communicating with the chatbot was clear.					
I was immediately made aware of what					

information the chatbot can give me.

It is clear to me early on about what the chatbot can do.

I had to rephrase my input multiple times for the chatbot to be able to help me.

I had to pay special attention regarding my phrasing when communicating with the chatbot.

It was easy to tell the chatbot what I would like it to do.

The interaction with the chatbot felt like an ongoing conversation.

The chatbot was able to keep track of context.

The chatbot maintained relevant conversation.

The chatbot guided me to the relevant service.

The chatbot is using hyperlinks to guide me to my goal.

The chatbot was able to make references to the website or service when appropriate.

The interaction with the chatbot felt secure in terms of privacy.

I believe the chatbot informs me of any possible privacy issues.

I believe that this chatbot maintains my privacy.

I felt that my intentions were understood by the chatbot.

The chatbot was able to guide me to my goal.

I find that the chatbot understands what I want

and helps me achieve my goal.

The chatbot gave relevant information during the whole conversation

The chatbot is good at providing me with a helpful response at any point of the process.

The chatbot provided relevant information as and when I needed it.

The amount of received information was neither too much nor too less

The chatbot gives me the appropriate amount of information

The chatbot only gives me the information I need

The chatbot could handle situations in which the line of conversation was not clear

The chatbot explained gracefully when it could not help me

When the chatbot encountered a problem, it responded appropriately

I found the chatbot's responses clear.

The chatbot only states understandable answers.

The chatbot's responses were easy to understand.

I feel like the chatbot's responses were accurate.

I believe that the chatbot only states reliable information.

It appeared that the chatbot provided accurate and reliable information.

The time of the response was reasonable.

My waiting time for a response from the chatbot was short.					
The chatbot is quick to respond.					

Table B2

UMUX-LITE

Statement	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
The system's capabilities meet my requirements					
The system is easy to use					

Appendix C

Qualtrics Survey Flow

For the first chatbot, the flow is completely visible. Due to redundancy, for the rest of the chatbots, only the specific items containing each specific task were included. Every participant was randomly presented with five of the shown chatbots.

Start of Block: Introduction

Welcome to the study "Chatbots as a tool for the future? An analysis of the USQ and user characteristics". In this study, you will assess five different chatbots by trying to complete two tasks for each of them. After each chatbot, you will fill in two questionnaires, the USQ - the focus of this study and the UMUX-LITE.

Before you begin, I would like you to answer a couple of questions regarding your demographics.

Page Break

Participant ID

_____

Page Break

End of Block: Introduction

Start of Block: Demographics

Gender

O Male

O Female

O Prefer not to say

Age

_____

Nationality

O Dutch

O German

O If other, please specify: _____

Field of study

O Psychology

O Communication science

O If other, please specify: _____

	Extremely familiar	Very familiar	Moderately familiar	Slightly familiar	Not familiar at all
How familiar are you with chatbots and/or other conversational interfaces?	O	O	O	O	O

	Definitely yes	Probably	Unsure	Probably not	Definitely not
Have you used a chatbot or a conversational interface before?	o	o	o	o	o

	Daily	4 - 6 times a week	2 - 3 times a week	Once a week	Rarely	Never
How often do you use it?	o	o	o	o	o	o

Page Break

End of Block: Demographics

Start of Block: Enddemogr

Now, the assessment of the chatbots will begin. Access the chatbots via the links, read the tasks carefully, and try to do them. If you feel incapable to solve a task that is not a problem. Just continue with the study then. In some cases, when you feel stuck you can reload the page to restart the conversation with a chatbot.

Page Break

End of Block: Enddemogr

Start of Block: Amtrak

Chatbot: Amtrak

The chatbot can be found at https://www.amtrak.com/home

Please access the chatbot now.

Page Break

Please do the following task on this chatbot:

You would like to travel from Boston to Washington D.C. while being in the USA. You want to use Amtrak's chatbot to book the shortest trip possible on the eigth of October. Your departure station is Back Bay Station.

Page Break

On a scale of 1 (very difficult) to 10 (very easy), how easy did you find this task?

	1	2	3	4	5	6	7	8	9	10	
	1	2	3	4	5	6	7	8	9	10	
Very difficult	O	O	O	O	O	O	O	O	O	O	Very easy

Page Break

Please do the following task on this chatbot:

You have planned a trip to the USA. You are planning to travel by train from Boston to Washington D.C. You want to stop in New York to meet an old friend for a few hours and see the city. You want to use Amtrak's chatbot to find out how much it will cost to temporarily store your luggage at the station.

Page Break

On a scale of 1 (very difficult) to 10 (very easy), how easy did you find this task?

	1	2	3	4	5	6	7	8	9	10	
	1	2	3	4	5	6	7	8	9	10	
Very difficult	O	O	O	O	O	O	O	O	O	O	Very easy

Page Break

Based on the chatbot you just interacted with, respond to the following statements.

	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree

It was clear how to start a conversation with the chatbot.	O	O	O	O	O
It was easy for me to understand how to start the interaction with the chatbot.	O	O	O	O	O
I find it easy to start a conversation with the chatbot.	O	O	O	O	O
The chatbot was easy to access.	O	O	O	O	O
The chatbot function was easily detectable.	O	O	O	O	O
It was easy to find the chatbot.	O	O	O	O	O
Communicating with the chatbot was clear.	O	O	O	O	O

I was immediately made aware of what information the chatbot can give me.	O	O	O	O	O
It is clear to me early on about what the chatbot can do.	O	O	O	O	O
I had to rephrase my input multiple times for the chatbot to be able to help me.	O	O	O	O	O
I had to pay special attention regarding my phrasing when communicating with the chatbot.	O	O	O	O	O
It was easy to tell the chatbot what I would like it to do.	O	O	O	O	O
The interaction with the chatbot felt like an	O	O	O	O	O

ongoing conversation.					
The chatbot was able to keep track of context.	O	O	O	O	O
The chatbot maintained relevant conversation.	O	O	O	O	O
The chatbot guided me to the relevant service.	O	O	O	O	O
The chatbot is using hyperlinks to guide me to my goal.	O	O	O	O	O
The chatbot was able to make references to the website or service when appropriate.	O	O	O	O	O
The interaction with the chatbot felt secure in terms of privacy.	O	O	O	O	O

I believe the chatbot informs me of any possible privacy issues.	O	O	O	O	O
I believe that this chatbot maintains my privacy.	O	O	O	O	O
I felt that my intentions were understood by the chatbot.	O	O	O	O	O
The chatbot was able to guide me to my goal.	O	O	O	O	O
I find that the chatbot understands what I want and helps me achieve my goal.	O	O	O	O	O
The chatbot gave relevant information during the whole conversation	O	O	O	O	O

The chatbot is good at providing me with a helpful response at any point of the process.	O	O	O	O	O
The chatbot provided relevant information as and when I needed it.	O	O	O	O	O
The amount of received information was neither too much nor too less	O	O	O	O	O
The chatbot gives me the appropriate amount of information	O	O	O	O	O
The chatbot only gives me the information I need	O	O	O	O	O

The chatbot could handle situations in which the line of conversation was not clear	O	O	O	O	O
The chatbot explained gracefully when it could not help me	O	O	O	O	O
When the chatbot encountered a problem, it responded appropriately	O	O	O	O	O
I found the chatbot's responses clear.	O	O	O	O	O
The chatbot only states understandable answers.	O	O	O	O	O
The chatbot's responses were easy to understand.	O	O	O	O	O

I feel like the chatbot's responses were accurate.	O	O	O	O	O
I believe that the chatbot only states reliable information.	O	O	O	O	O
It appeared that the chatbot provided accurate and reliable information.	O	O	O	O	O
The time of the response was reasonable.	O	O	O	O	O
My waiting time for a response from the chatbot was short.	O	O	O	O	O
The chatbot is quick to respond.	O	O	O	O	O

Page Break

Based on the chatbot you just interacted with, respond to the following statements.

	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
This system's capabilities meet my requirements.	O	O	O	O	O
This system is easy to use.	O	O	O	O	O

End of Block: Amtrak

Start of Block: Emirates Holidays

Chatbot: Emirates Holidays

The chatbot can be found at https://www.emiratesholidays.com/gb_en/

Please access the chatbot now.

Page Break

Please do the following task with this chatbot:

You just woke up and realised that you forgot that it is your partner's birthday. Desperately, you are thinking about a birthday present and your idea is a holiday together in Paris. You visit the Emirates Holiday's page and use Emirates Holidays' chatbot to book a holiday from the fourth of September until the ninth of September to Paris for two persons. Your departure airport is London Heathrow (LHR). Everything else is not important, as you just need a present for today.

Page Break

On a scale of 1 (very difficult) to 10 (very easy), how easy did you find this task?

	1	2	3	4	5	6	7	8	9	10	
	1 (1)	2 (2)	3 (3)	4 (4)	5 (5)	6 (6)	7 (7)	8 (8)	9 (9)	10 (10)	
Very difficult	O	O	O	O	O	O	O	O	O	O	Very easy

Page Break

Please do the following task with this chatbot:

You arrived in Paris and there seems to be a problem with your hotel reservation. You try to call someone at Emirates Holiday, but it is 11 pm on Friday, so you cannot reach anyone. Hence, you ask Emirates Holidays' chatbot when the customer service opens on Saturday.

Page Break

Start of Block: ATO

Chatbot: ATO

The chatbot can be found at http://www.ato.gov.au/

Please access the chatbot now.

Page Break

Please do the following task with this chatbot:

You moved to Australia from the Netherlands recently. You want to know when the deadline is to lodge/submit your tax return using ATO's chatbot to find out.

Page Break

On a scale of 1 (very difficult) to 10 (very easy), how easy did you find this task?

	1	2	3	4	5	6	7	8	9	10	
	1	2	3	4	5	6	7	8	9	10	
Very difficult	O	O	O	O	O	O	O	O	O	O	Very easy

Page Break

Please do the following task with this chatbot:

You are a student and are wondering whether you have to lodge a tax return using the ATO's chatbot.

Page Break

On a scale of 1 (very difficult) to 10 (very easy), how easy did you find this task?

	1	2	3	4	5	6	7	8	9	10	
	1	2	3	4	5	6	7	8	9	10	
Very difficult	O	O	O	O	O	O	O	O	O	O	Very

		easy

Page Break

Chatbot: HubSpot

The chatbot can be found at https://www.hubspot.com/?survey=123

Please access the chatbot now.

Page Break

Please do the following task on this chatbot:

You have your own company and would like to grow your business even more. A former colleague recommended you HubSpot. However, you do not want to sign up for anything (even if it is free). You use HubSpot's chatbot to purely get information and get educated without using any tools. A collection of news/articles/tips would be great.

Page Break

On a scale of 1 (very difficult) to 10 (very easy), how easy did you find this task?

	1	2	3	4	5	6	7	8	9	10	
	1	2	3	4	5	6	7	8	9	10	
Very difficult	O	O	O	O	O	O	O	O	O	O	Very easy

Page Break

Please do the following task on this chatbot:

Now, you are convinced that HubSpot can help your own business. Your focus is on improving your customer service. Before you sign up for something, you would like to know how HubSpot can improve your customer service. You use HubSpot's chatbot to get more information about this.

Page Break

On a scale of 1 (very difficult) to 10 (very easy), how easy did you find this task?

	1	2	3	4	5	6	7	8	9	10	
	1	2	3	4	5	6	7	8	9	10	
Very difficult	O	O	O	O	O	O	O	O	O	O	Very easy

Page Break

Start of Block: UT

Chatbot: University of Twente

The chatbot can be found at https://www.utwente.nl/en/education/master/chat/

Please access the chatbot now.

Page Break

Please do the following task on this chatbot:

You are a Chinese student who would like to do a master's degree at the University of Twente. Your name is Jackie/Lin and your Email address is abc@def.com. You are interested in doing your master's in Nanotechnology in September 2021. You did your bachelor at the Utwente in the Netherlands. You ask the Utwente chatbot what options for a scholarship are available.

Page Break

On a scale of 1 (very difficult) to 10 (very easy), how easy did you find this task?

	1	2	3	4	5	6	7	8	9	10	
	1	2	3	4	5	6	7	8	9	10	
Very difficult	O	O	O	O	O	O	O	O	O	O	Very easy

Page Break

You are a German student who would like to do a master's degree at the University of Twente. Your name is Alan/Sabine and your Email address is abc@def.com. You are interested in doing your master's in computer science in February 2022. You did your bachelor's at the Jacobs University in Bremen. You ask the Utwente chatbot about deadlines and the admission process.

Page Break

On a scale of 1 (very difficult) to 10 (very easy), how easy did you find this task?

	1	2	3	4	5	6	7	8	9	10	

	1	2	3	4	5	6	7	8	9	10	
Very difficult	O	O	O	O	O	O	O	O	O	O	Very easy

Page Break

Start of Block: HSBC

Chatbot: HSBC UK

The chatbot can be found at https://www.hsbc.co.uk/

Please access the chatbot now.

Page Break

Please do the following task on this chatbot:

You live in the Netherlands but are travelling to Turkey for two weeks. During your travel, you would like to be able to use your HSBC credit card overseas at payment terminals and ATMs. You want to use HSBC's chatbot to find out the relevant procedure.

Page Break

On a scale of 1 (very difficult) to 10 (very easy), how easy did you find this task?

	1	2	3	4	5	6	7	8	9	10	
	1	2	3	4	5	6	7	8	9	10	

Very difficult	O	O	O	O	O	O	O	O	O	O	Very easy

Page Break

Please do the following task on this chatbot:

You have recently moved from Amsterdam to London and would like to know how you can change your address for your HSBC card, using the chatbot of HSBC UK.

Page Break

On a scale of 1 (very difficult) to 10 (very easy), how easy did you find this task?

	1	2	3	4	5	6	7	8	9	10	
	1	2	3	4	5	6	7	8	9	10	
Very difficult	O	O	O	O	O	O	O	O	O	O	Very easy

Page Break

Start of Block: Absolut

Chatbot: Absolut

The chatbot can be found at https://www.absolut.com/en/

Please access the chatbot now.

Page Break

Please do the following task on this chatbot:

You want to buy a bottle of Absolut vodka to share with your friends for the evening. One of your friends cannot consume gluten. You want to use Absolut's chatbot to find out if Absolut Lime contains gluten or not.

Page Break

On a scale of 1 (very difficult) to 10 (very easy), how easy did you find this task?

	1	2	3	4	5	6	7	8	9	10	
	1	2	3	4	5	6	7	8	9	10	
Very difficult	O	O	O	O	O	O	O	O	O	O	Very easy

Page Break

Please do the following task on this chatbot:

You want to buy a bottle of Absolut vodka for a good friend. But this friend is right now on a diet and tries to avoid sugar. You want to find information about the amount of sugar in the products of Absolut using Absolut's chatbot.

Page Break

On a scale of 1 (very difficult) to 10 (very easy), how easy did you find this task?

	1	2	3	4	5	6	7	8	9	10	
	1	2	3	4	5	6	7	8	9	10	
Very difficult	O	O	O	O	O	O	O	O	O	O	Very easy

Page Break

Start of Block: ManyChat

Chatbot: ManyChat

The chatbot can be found at https://www.facebook.com/messages/t/ManyChat

Please access the chatbot now.

Page Break

You want to integrate a chatbot on your companies' website. Therefore, you want to use ManyChat's chatbot to find video tutorials to learn the basics of ManyChat.

Page Break

On a scale of 1 (very difficult) to 10 (very easy), how easy did you find this task?

	1	2	3	4	5	6	7	8	9	10	
	1	2	3	4	5	6	7	8	9	10	
Very difficult	O	O	O	O	O	O	O	O	O	O	Very easy

Page Break

After using the Chatbot for a while, you are getting a little bored and want to have some fun. Let the ManyChat's chatbot tell a joke to you.

Page Break

On a scale of 1 (very difficult) to 10 (very easy), how easy did you find this task?

	1	2	3	4	5	6	7	8	9	10	
	1	2	3	4	5	6	7	8	9	10	
Very difficult	O	O	O	O	O	O	O	O	O	O	Very easy

Page Break

Start of Block: USCIS

Chatbot: USCIS

The chatbot can be found at http://www.uscis.gov/emma

Please access the chatbot now.

Page Break

Please do the following task on this chatbot:

You are a US citizen living abroad and want to vote in the upcoming federal elections. You want to use the USCIS chatbot to find out how.

Page Break

On a scale of 1 (very difficult) to 10 (very easy), how easy did you find this task?

	1	2	3	4	5	6	7	8	9	10	
	1	2	3	4	5	6	7	8	9	10	
Very difficult	O	O	O	O	O	O	O	O	O	O	Very easy

Page Break

Please do the following task on this chatbot:

You are planning to take a job in the USA. Since you are not a US citizen, you want to find out more about eligibility for a US-Green Card with the help of the USCIS chatbot.

Page Break

On a scale of 1 (very difficult) to 10 (very easy), how easy did you find this task?

	1	2	3	4	5	6	7	8	9	10	
	1	2	3	4	5	6	7	8	9	10	
Very difficult	O	O	O	O	O	O	O	O	O	O	Very easy

Page Break

Start of Block: NBC News

Chatbot: NBC News

The chatbot can be found at https://www.facebook.com/messages/t/NBCNews

Please access the chatbot now.

Page Break

Please do the following task on this chatbot:

You want to use the chatbot of NBC News to find out the most recent news regarding the environment.

Page Break

On a scale of 1 (very difficult) to 10 (very easy), how easy did you find this task?

	1	2	3	4	5	6	7	8	9	10	
	1	2	3	4	5	6	7	8	9	10	
Very difficult	O	O	O	O	O	O	O	O	O	O	Very easy

Page Break

Start of Block: Booking.com

Chatbot: Booking.com

The chatbot can be found at https://www.facebook.com/messages/t/1318400301782 50

Please access the chatbot now.

Page Break

Please do the following task on this chatbot:

You are traveling to London from the fifth of July to the ninth of July with your family. You want to use booking.com's chatbot to find a hotel room for you, your significant other, and your child in Central London that does not cost more than £500.

Page Break

On a scale of 1 (very difficult) to 10 (very easy), how easy did you find this task?

	1	2	3	4	5	6	7	8	9	10	
	1	2	3	4	5	6	7	8	9	10	
Very difficult	O	O	O	O	O	O	O	O	O	O	Very easy

Page Break

Please do the following task on this chatbot:

You have to attend an important business meeting from 18th to 19th of March in Amsterdam. You, therefore, are looking for a place to stay in the city centre of Amsterdam for not more than 200€ using booking.com's chatbot.

Page Break

On a scale of 1 (very difficult) to 10 (very easy), how easy did you find this task?

	1	2	3	4	5	6	7	8	9	10	
	1	2	3	4	5	6	7	8	9	10	
Very difficult	O	O	O	O	O	O	O	O	O	O	Very easy

Page Break

Start of Block: Endchatbots

Thank you very much for assessing the chatbots. I would finally like you to fill in two questionnaires regarding your trust towards the internet as well as your interest in technology. Afterwards, the study is finished.

Page Break

End of Block: Endchatbots

Start of Block: Institution-based Trust

I would also like you to give a couple of statements regarding your trust towards the Internet.

	Strongly disagree	moderately disagree	mildly disagree	neutral	mildly agree	moderately agree	strongly agree
1. I feel good about how things go when I do purchasing or other activities on the Internet.	O	O	O	O	O	O	O
2. I am comfortable making purchases on the Internet.	O	O	O	O	O	O	O

3. I feel that most Internet vendors would act in a customers' best interest.	O	O	O	O	O	O	O
4. If a customer required help, most Internet vendors would do their best to help.	O	O	O	O	O	O	O
5. Most Internet vendors are interested in customer well-being, not just their own well-being	O	O	O	O	O	O	O
6. I am comfortable relying on Internet vendors to meet their obligations.	O	O	O	O	O	O	O

7. I feel fine doing business on the Internet since Internet vendors generally fulfill their agreements.	O	O	O	O	O	O	O
8. I always feel confident that I can rely on Internet vendors to do their part when I interact with them.	O	O	O	O	O	O	O
9. In general, most Internet vendors are competent at serving their customers.	O	O	O	O	O	O	O
10. Most Internet vendors do a capable job at meeting customer needs.	O	O	O	O	O	O	O

11. I feel that most Internet vendors are good at what they do.	O	O	O	O	O	O	O
12. The Internet has enough safeguards to make me feel comfortable using it to transact personal business.	O	O	O	O	O	O	O
13. I feel assured that legal and technological structures adequately protect me from problems on the Internet.	O	O	O	O	O	O	O
14. I feel confident that encryption and other technological advances on the Internet make it safe for me to do business there.	O	O	O	O	O	O	O

15. In general, the Internet is now a robust and safe environment in which to transact business.	O	O	O	O	O	O	O

Page Break

End of Block: Institution-based Trust

Start of Block: Geekism

Furthermore, some statements regarding your interest and behaviour regarding Computers and technology in general.

	I totally disagree	I disagree	Cannot answer	I agree	I totally agree
1. I want to understand how computer parts and software work.	O	O	O	O	O
2. Complex procedures with technical devices put me off	O	O	O	O	O

3. I have sometimes modified a technical device or diverted it from its intended purpose.	O	O	O	O	O
4. I am motivated to optimize technical devices or configure them to my requirements.	O	O	O	O	O
5. I have, or I would make a project or work of mine publicly available on the Internet.	O	O	O	O	O
6. Some people would call me a computer freak,	O	O	O	O	O
7. I am not interested in the inner working or coding of software.	O	O	O	O	O

8. Challenging tasks with technical devices appeal to me.	O	O	O	O	O
9. I have good knowledge of computing devices.	O	O	O	O	O
10. I invest a lot of time and effort to explore computing devices.	O	O	O	O	O
11. I like acquiring more knowledge of technical devices.	O	O	O	O	O
12. I have more than once opened technical devices to see their insides.	O	O	O	O	O
13. Sometimes I use technical devices different to what they were intended for.	O	O	O	O	O

14. It puts me off when technical devices have too many settings options.	O	O	O	O	O
15. Usually, I need help when having trouble with a technical device.	O	O	O	O	O

Page Break

End of Block: Geekism

Start of Block: End

That is the end of the session. Thank you for participating. You will receive your SONA credits soon.

Study contact details:

Principal researcher

Alexander Dehmel

a.dehmel@student.utwente.nl

Co-investigator/supervisor

Dr. Simone Borsci

s.borsci@utwente.nl

Contact information for questions about your rights as a research participant

End of Block: End

Appendix D

Table D1

Geekism scale

Statement	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
1. I want to understand how computer parts and software work.					
2. Complex procedures with technical devices put me off					
3. I have sometimes modified a technical device or diverted it from its intended purpose.					
4. I am motivated to optimize technical devices or configure them to my requirements.					
5. I have, or I would make a project or work of mine publicly					

available on the Internet.

6. Some people would call me a computer freak,

7. I am not interested in the inner working or coding of software.

8. Challenging tasks with technical devices appeal to me.

9. I have good knowledge of computing devices.

10. I invest a lot of time and effort to explore computing devices.

11. I like acquiring more knowledge of technical devices.

12. I have more than once opened technical devices to see their insides.

13. Sometimes I use technical devices different to what they were intended for.							
14. It puts me off when technical devices have too many settings options.							
15. Usually, I need help when having trouble with a technical device.							

Table D2

Institution-based trust questionnaire

Statements	Strongly disagree	Moderately disagree	Mildly disagree	Neither agree nor disagree	Mildly agree	Moderately agree	Strongly agree
1. I feel good about how things go when I do purchasing or other activities on the Internet.							
2. I am comfortable							

making purchases on the Internet.	
3. I feel that most Internet vendors would act in a customers' best interest.	
4. If a customer required help, most Internet vendors would do their best to help.	
5. Most Internet vendors are interested in customer well-being, not just their own well-being	
6. I am comfortable relying on Internet vendors to meet their obligations.	
7. I feel fine doing business on the Internet since Internet vendors	

generally fulfill their agreements.

8. I always feel confident that I can rely on Internet vendors to do their part when I interact with them.

9. In general, most Internet vendors are competent at serving their customers.

10. Most Internet vendors do a capable job at meeting customer needs.

11. I feel that most Internet vendors are good at what they do.

12. The Internet has enough safeguards to make me feel comfortable using it to

transact personal business.	
13. I feel assured that legal and technological structures adequately protect me from problems on the Internet.	
14. I feel confident that encryption and other technological advances on the Internet make it safe for me to do business there.	
15. In general, the Internet is now a robust and safe environment in which to transact business.	

Appendix E

Spss-Syntax

```
#Preparation of dataset

COMPUTE usq_total=SUM(USQ_1, USQ_2, USQ_3, USQ_4, USQ_5, USQ_6,
    USQ_7, USQ_8, USQ_9, USQ_10, USQ_11, USQ_12, USQ_13,
    USQ_14, USQ_15, USQ_16, USQ_17, USQ_18, USQ_19, USQ_20,
    USQ_21, USQ_22, USQ_23, USQ_24, USQ_25, USQ_26, USQ_27,
    USQ_28, USQ_29, USQ_30, USQ_31, USQ_32, USQ_33, USQ_34,
   USQ_35, USQ_36, USQ_37, USQ_38, USQ_39, USQ_40, USQ_41,
    USQ_42).
EXECUTE.

COMPUTE usq_rescaled=(MEAN(USQ_1, USQ_2, USQ_3, USQ_4, USQ_5, USQ_6,
    USQ_7, USQ_8, USQ_9, USQ_10, USQ_11, USQ_12, USQ_13,
    USQ_14, USQ_15, USQ_16, USQ_17, USQ_18, USQ_19, USQ_20,
    USQ_21, USQ_22, USQ_23, USQ_24, USQ_25, USQ_26, USQ_27,
    USQ_28, USQ_29, USQ_30, USQ_31, USQ_32, USQ_33, USQ_34,
   USQ_35, USQ_36, USQ_37, USQ_38, USQ_39, USQ_40, USQ_41,
    USQ_42)-1)/4.
EXECUTE.

COMPUTE umux_total=((SUM(UMUX_1,UMUX_2)-2)/8)*100.
EXECUTE.

COMPUTE umux_rescaled=((SUM(UMUX_1,UMUX_2)-2)/8).
EXECUTE.

RECODE USQ_10 USQ_11 (1=5) (2=4) (3=3) (4=2) (5=1).
EXECUTE.

RECODE Gender (1=1) (2=0) INTO gender_dummy.
EXECUTE.

RECODE first_time_user (1=1) (2=0) INTO firsttime_dummy.
EXECUTE.

#exploring descriptive statistics for questionnaire scores

EXAMINE VARIABLES=usq_total umux_total usq_rescaled umux_rescaled
  /PLOT BOXPLOT HISTOGRAM NPPLOT
  /COMPARE GROUPS
```

```
  /STATISTICS DESCRIPTIVES
  /CINTERVAL 95
  /MISSING LISTWISE
  /NOTOTAL.

#correlational analysis between USQ and UMUX-LITE

BOOTSTRAP
  /SAMPLING METHOD=SIMPLE
  /VARIABLES INPUT=usq_rescaled umux_rescaled
  /CRITERIA CILEVEL=97.5 CITYPE=BCA  NSAMPLES=9999
  /MISSING USERMISSING=EXCLUDE.
NONPAR CORR
  /VARIABLES=usq_rescaled umux_rescaled
  /PRINT=KENDALL TWOTAIL NOSIG
  /MISSING=PAIRWISE.

#principal component analysis

FACTOR
  /VARIABLES USQ_1 USQ_2 USQ_3 USQ_4 USQ_5 USQ_6 USQ_7 USQ_8 USQ_9 USQ_10
USQ_11 USQ_12 USQ_13
    USQ_14 USQ_15 USQ_16 USQ_17 USQ_18 USQ_19 USQ_20 USQ_21 USQ_22 USQ_23 USQ_24
USQ_25 USQ_26 USQ_27
    USQ_28 USQ_29 USQ_30 USQ_31 USQ_32 USQ_33 USQ_34 USQ_35 USQ_36 USQ_37 USQ_38
USQ_39 USQ_40 USQ_41
    USQ_42
  /MISSING PAIRWISE
  /ANALYSIS USQ_1 USQ_2 USQ_3 USQ_4 USQ_5 USQ_6 USQ_7 USQ_8 USQ_9 USQ_10 USQ_11
USQ_12 USQ_13
    USQ_14 USQ_15 USQ_16 USQ_17 USQ_18 USQ_19 USQ_20 USQ_21 USQ_22 USQ_23 USQ_24
USQ_25 USQ_26 USQ_27
    USQ_28 USQ_29 USQ_30 USQ_31 USQ_32 USQ_33 USQ_34 USQ_35 USQ_36 USQ_37 USQ_38
USQ_39 USQ_40 USQ_41
    USQ_42
  /PRINT INITIAL CORRELATION KMO EXTRACTION ROTATION
  /FORMAT SORT BLANK(.30)
  /PLOT EIGEN
  /CRITERIA MINEIGEN(1) ITERATE(25)
  /EXTRACTION PC
  /CRITERIA ITERATE(25) DELTA(0)
  /ROTATION OBLIMIN
  /METHOD=CORRELATION.
```

```
FACTOR
  /VARIABLES USQ_1 USQ_2 USQ_3 USQ_4 USQ_5 USQ_6 USQ_7 USQ_8 USQ_9 USQ_10
USQ_11 USQ_12 USQ_13
    USQ_14 USQ_15 USQ_16 USQ_17 USQ_18 USQ_19 USQ_20 USQ_21 USQ_22 USQ_23 USQ_24
USQ_25 USQ_26 USQ_27
    USQ_28 USQ_29 USQ_30 USQ_31 USQ_32 USQ_33 USQ_34 USQ_35 USQ_36 USQ_37 USQ_38
USQ_39 USQ_40 USQ_41
    USQ_42
  /MISSING PAIRWISE
  /ANALYSIS USQ_1 USQ_2 USQ_3 USQ_4 USQ_5 USQ_6 USQ_7 USQ_8 USQ_9 USQ_10 USQ_11
USQ_12 USQ_13
    USQ_14 USQ_15 USQ_16 USQ_17 USQ_18 USQ_19 USQ_20 USQ_21 USQ_22 USQ_23 USQ_24
USQ_25 USQ_26 USQ_27
    USQ_28 USQ_29 USQ_30 USQ_31 USQ_32 USQ_33 USQ_34 USQ_35 USQ_36 USQ_37 USQ_38
USQ_39 USQ_40 USQ_41
    USQ_42
  /PRINT Initial EXTRACTION ROTATION
  /FORMAT SORT BLANK(.30)
  /CRITERIA FACTORS(5) ITERATE(25)
  /EXTRACTION PC
  /CRITERIA ITERATE(25) DELTA(0)
  /ROTATION OBLIMIN
  /METHOD=CORRELATION.

FACTOR
  /VARIABLES USQ_1 USQ_2 USQ_3 USQ_4 USQ_5 USQ_6 USQ_7 USQ_9 USQ_10 USQ_11
USQ_12 USQ_13
    USQ_14 USQ_15 USQ_16 USQ_18 USQ_19 USQ_20 USQ_21 USQ_22 USQ_23 USQ_24 USQ_25
USQ_26 USQ_27
    USQ_28 USQ_29 USQ_30 USQ_31 USQ_32 USQ_33 USQ_34 USQ_35 USQ_36 USQ_37 USQ_38
USQ_39 USQ_40 USQ_41
    USQ_42
  /MISSING PAIRWISE
  /ANALYSIS USQ_1 USQ_2 USQ_3 USQ_4 USQ_5 USQ_6 USQ_7 USQ_9 USQ_10 USQ_11
USQ_12 USQ_13
    USQ_14 USQ_15 USQ_16 USQ_18 USQ_19 USQ_20 USQ_21 USQ_22 USQ_23 USQ_24 USQ_25
USQ_26 USQ_27
    USQ_28 USQ_29 USQ_30 USQ_31 USQ_32 USQ_33 USQ_34 USQ_35 USQ_36 USQ_37 USQ_38
USQ_39 USQ_40 USQ_41
    USQ_42
  /PRINT Initial EXTRACTION ROTATION
  /FORMAT SORT BLANK(.30)
  /CRITERIA FACTORS(5) ITERATE(25)
  /EXTRACTION PC
```

```
  /CRITERIA ITERATE(25) DELTA(0)
  /ROTATION OBLIMIN
  /METHOD=CORRELATION.

FACTOR
  /VARIABLES USQ_1 USQ_2 USQ_3 USQ_4 USQ_5 USQ_6 USQ_7 USQ_10 USQ_11 USQ_12
USQ_13
    USQ_14 USQ_15 USQ_16 USQ_18 USQ_19 USQ_20 USQ_21 USQ_22 USQ_23 USQ_24 USQ_25
USQ_26 USQ_27
    USQ_28 USQ_29 USQ_30 USQ_31 USQ_32 USQ_33 USQ_34 USQ_35 USQ_36 USQ_37 USQ_38
USQ_39 USQ_40 USQ_41
    USQ_42
  /MISSING PAIRWISE
  /ANALYSIS USQ_1 USQ_2 USQ_3 USQ_4 USQ_5 USQ_6 USQ_7 USQ_10 USQ_11 USQ_12
USQ_13
    USQ_14 USQ_15 USQ_16 USQ_18 USQ_19 USQ_20 USQ_21 USQ_22 USQ_23 USQ_24 USQ_25
USQ_26 USQ_27
    USQ_28 USQ_29 USQ_30 USQ_31 USQ_32 USQ_33 USQ_34 USQ_35 USQ_36 USQ_37 USQ_38
USQ_39 USQ_40 USQ_41
    USQ_42
  /PRINT Initial EXTRACTION ROTATION
  /FORMAT SORT BLANK(.30)
  /CRITERIA FACTORS(5) ITERATE(25)
  /EXTRACTION PC
  /CRITERIA ITERATE(25) DELTA(0)
  /ROTATION OBLIMIN
  /METHOD=CORRELATION.

FACTOR
  /VARIABLES USQ_1 USQ_2 USQ_3 USQ_4 USQ_5 USQ_6 USQ_7 USQ_10 USQ_11 USQ_12
USQ_13
    USQ_14 USQ_15 USQ_16 USQ_19 USQ_20 USQ_21 USQ_22 USQ_23 USQ_24 USQ_25 USQ_26
USQ_27
    USQ_28 USQ_29 USQ_30 USQ_31 USQ_32 USQ_33 USQ_34 USQ_35 USQ_36 USQ_37 USQ_38
USQ_39 USQ_40 USQ_41
    USQ_42
  /MISSING PAIRWISE
  /ANALYSIS USQ_1 USQ_2 USQ_3 USQ_4 USQ_5 USQ_6 USQ_7 USQ_10 USQ_11 USQ_12
USQ_13
    USQ_14 USQ_15 USQ_16 USQ_19 USQ_20 USQ_21 USQ_22 USQ_23 USQ_24 USQ_25 USQ_26
USQ_27
    USQ_28 USQ_29 USQ_30 USQ_31 USQ_32 USQ_33 USQ_34 USQ_35 USQ_36 USQ_37 USQ_38
USQ_39 USQ_40 USQ_41
    USQ_42
```

```
  /PRINT Initial EXTRACTION ROTATION
  /FORMAT SORT BLANK(.30)
  /CRITERIA FACTORS(5) ITERATE(25)
  /EXTRACTION PC
  /CRITERIA ITERATE(25) DELTA(0)
  /ROTATION OBLIMIN
  /METHOD=CORRELATION.

FACTOR
  /VARIABLES USQ_1 USQ_2 USQ_3 USQ_4 USQ_5 USQ_6 USQ_7 USQ_10 USQ_11 USQ_12
USQ_13
    USQ_14 USQ_15 USQ_16 USQ_19 USQ_20 USQ_21 USQ_22 USQ_23 USQ_24 USQ_25 USQ_26
USQ_27
    USQ_28 USQ_29 USQ_30 USQ_31 USQ_32 USQ_33 USQ_34 USQ_35 USQ_36 USQ_37 USQ_38
USQ_39 USQ_40 USQ_41
    USQ_42
  /MISSING PAIRWISE
  /ANALYSIS USQ_1 USQ_2 USQ_3 USQ_4 USQ_5 USQ_6 USQ_7 USQ_10 USQ_11 USQ_12
USQ_13
    USQ_14 USQ_15 USQ_16 USQ_19 USQ_20 USQ_21 USQ_22 USQ_23 USQ_24 USQ_25 USQ_26
USQ_27
    USQ_28 USQ_29 USQ_30 USQ_31 USQ_32 USQ_33 USQ_34 USQ_35 USQ_36 USQ_37 USQ_38
USQ_39 USQ_40 USQ_41
    USQ_42
  /PRINT Initial EXTRACTION ROTATION
  /FORMAT SORT BLANK(.30)
  /CRITERIA FACTORS(5) ITERATE(25)
  /EXTRACTION PC
  /CRITERIA ITERATE(25) DELTA(0)
  /ROTATION OBLIMIN
  /METHOD=CORRELATION.

FACTOR
  /VARIABLES USQ_1 USQ_2 USQ_3 USQ_4 USQ_5 USQ_6 USQ_10 USQ_11 USQ_12 USQ_13
    USQ_14 USQ_15 USQ_16 USQ_19 USQ_20 USQ_21 USQ_22 USQ_23 USQ_24 USQ_25 USQ_26
USQ_27
    USQ_28 USQ_29 USQ_30 USQ_31 USQ_32 USQ_33 USQ_34 USQ_35 USQ_36 USQ_37 USQ_38
USQ_39 USQ_40 USQ_41
    USQ_42
  /MISSING PAIRWISE
  /ANALYSIS USQ_1 USQ_2 USQ_3 USQ_4 USQ_5 USQ_6 USQ_10 USQ_11 USQ_12 USQ_13
    USQ_14 USQ_15 USQ_16 USQ_19 USQ_20 USQ_21 USQ_22 USQ_23 USQ_24 USQ_25 USQ_26
USQ_27
```

```
   USQ_28 USQ_29 USQ_30 USQ_31 USQ_32 USQ_33 USQ_34 USQ_35 USQ_36 USQ_37 USQ_38
USQ_39 USQ_40 USQ_41
   USQ_42
  /PRINT Initial EXTRACTION ROTATION
  /FORMAT SORT BLANK(.30)
  /CRITERIA FACTORS(5) ITERATE(25)
  /EXTRACTION PC
  /CRITERIA ITERATE(25) DELTA(0)
  /ROTATION OBLIMIN
  /METHOD=CORRELATION.

FACTOR
  /VARIABLES USQ_1 USQ_2 USQ_3 USQ_4 USQ_5 USQ_6 USQ_10 USQ_11 USQ_13
   USQ_14 USQ_15 USQ_16 USQ_19 USQ_20 USQ_21 USQ_22 USQ_23 USQ_24 USQ_25 USQ_26
USQ_27
   USQ_28 USQ_29 USQ_30 USQ_31 USQ_32 USQ_33 USQ_34 USQ_35 USQ_36 USQ_37 USQ_38
USQ_39 USQ_40 USQ_41
   USQ_42
  /MISSING PAIRWISE
  /ANALYSIS USQ_1 USQ_2 USQ_3 USQ_4 USQ_5 USQ_6 USQ_10 USQ_11 USQ_13
   USQ_14 USQ_15 USQ_16 USQ_19 USQ_20 USQ_21 USQ_22 USQ_23 USQ_24 USQ_25 USQ_26
USQ_27
   USQ_28 USQ_29 USQ_30 USQ_31 USQ_32 USQ_33 USQ_34 USQ_35 USQ_36 USQ_37 USQ_38
USQ_39 USQ_40 USQ_41
   USQ_42
  /PRINT Initial EXTRACTION ROTATION
  /FORMAT SORT BLANK(.30)
  /CRITERIA FACTORS(5) ITERATE(25)
  /EXTRACTION PC
  /CRITERIA ITERATE(25) DELTA(0)
  /ROTATION OBLIMIN
  /METHOD=CORRELATION.

FACTOR
  /VARIABLES USQ_1 USQ_2 USQ_3 USQ_4 USQ_5 USQ_6 USQ_10 USQ_11 USQ_13
   USQ_14 USQ_15 USQ_16 USQ_19 USQ_20 USQ_21 USQ_22 USQ_23 USQ_24 USQ_25 USQ_26
USQ_27
   USQ_28 USQ_29 USQ_30 USQ_31 USQ_32 USQ_33 USQ_34 USQ_35 USQ_37 USQ_38 USQ_39
USQ_40 USQ_41
   USQ_42
  /MISSING PAIRWISE
  /ANALYSIS USQ_1 USQ_2 USQ_3 USQ_4 USQ_5 USQ_6 USQ_10 USQ_11 USQ_13
   USQ_14 USQ_15 USQ_16 USQ_19 USQ_20 USQ_21 USQ_22 USQ_23 USQ_24 USQ_25 USQ_26
USQ_27
```

```
   USQ_28 USQ_29 USQ_30 USQ_31 USQ_32 USQ_33 USQ_34 USQ_35 USQ_37 USQ_38 USQ_39
USQ_40 USQ_41
   USQ_42
  /PRINT Initial EXTRACTION ROTATION
  /FORMAT SORT BLANK(.30)
  /CRITERIA FACTORS(5) ITERATE(25)
  /EXTRACTION PC
  /CRITERIA ITERATE(25) DELTA(0)
  /ROTATION OBLIMIN
  /METHOD=CORRELATION.

FACTOR
  /VARIABLES USQ_1 USQ_2 USQ_3 USQ_4 USQ_5 USQ_6 USQ_11 USQ_13
   USQ_14 USQ_15 USQ_16 USQ_19 USQ_20 USQ_21 USQ_22 USQ_23 USQ_24 USQ_25 USQ_26
USQ_27
   USQ_28 USQ_29 USQ_30 USQ_31 USQ_32 USQ_33 USQ_34 USQ_35 USQ_37 USQ_38 USQ_39
USQ_40 USQ_41
   USQ_42
  /MISSING PAIRWISE
  /ANALYSIS USQ_1 USQ_2 USQ_3 USQ_4 USQ_5 USQ_6 USQ_11 USQ_13
   USQ_14 USQ_15 USQ_16 USQ_19 USQ_20 USQ_21 USQ_22 USQ_23 USQ_24 USQ_25 USQ_26
USQ_27
   USQ_28 USQ_29 USQ_30 USQ_31 USQ_32 USQ_33 USQ_34 USQ_35 USQ_37 USQ_38 USQ_39
USQ_40 USQ_41
   USQ_42
  /PRINT Initial EXTRACTION ROTATION
  /FORMAT SORT BLANK(.30)
  /CRITERIA FACTORS(5) ITERATE(25)
  /EXTRACTION PC
  /CRITERIA ITERATE(25) DELTA(0)
  /ROTATION OBLIMIN
  /METHOD=CORRELATION.

FACTOR
  /VARIABLES USQ_1 USQ_2 USQ_3 USQ_4 USQ_5 USQ_6 USQ_13
   USQ_14 USQ_15 USQ_16 USQ_19 USQ_20 USQ_21 USQ_22 USQ_23 USQ_24 USQ_25 USQ_26
USQ_27
   USQ_28 USQ_29 USQ_30 USQ_31 USQ_32 USQ_33 USQ_34 USQ_35 USQ_37 USQ_38 USQ_39
USQ_40 USQ_41
   USQ_42
  /MISSING PAIRWISE
  /ANALYSIS USQ_1 USQ_2 USQ_3 USQ_4 USQ_5 USQ_6 USQ_13
   USQ_14 USQ_15 USQ_16 USQ_19 USQ_20 USQ_21 USQ_22 USQ_23 USQ_24 USQ_25 USQ_26
USQ_27
```

```
   USQ_28 USQ_29 USQ_30 USQ_31 USQ_32 USQ_33 USQ_34 USQ_35 USQ_37 USQ_38 USQ_39
USQ_40 USQ_41
   USQ_42
  /PRINT Initial EXTRACTION ROTATION
  /FORMAT SORT BLANK(.30)
  /CRITERIA FACTORS(5) ITERATE(25)
  /EXTRACTION PC
  /CRITERIA ITERATE(25) DELTA(0)
  /ROTATION OBLIMIN
  /METHOD=CORRELATION.

FACTOR
  /VARIABLES USQ_1 USQ_2 USQ_3 USQ_4 USQ_5 USQ_6 USQ_13
   USQ_14 USQ_16 USQ_19 USQ_20 USQ_21 USQ_22 USQ_23 USQ_24 USQ_25 USQ_26 USQ_27
   USQ_28 USQ_29 USQ_30 USQ_31 USQ_32 USQ_33 USQ_34 USQ_35 USQ_37 USQ_38 USQ_39
USQ_40 USQ_41
   USQ_42
  /MISSING PAIRWISE
  /ANALYSIS USQ_1 USQ_2 USQ_3 USQ_4 USQ_5 USQ_6 USQ_13
   USQ_14 USQ_16 USQ_19 USQ_20 USQ_21 USQ_22 USQ_23 USQ_24 USQ_25 USQ_26 USQ_27
   USQ_28 USQ_29 USQ_30 USQ_31 USQ_32 USQ_33 USQ_34 USQ_35 USQ_37 USQ_38 USQ_39
USQ_40 USQ_41
   USQ_42
  /PRINT Initial EXTRACTION ROTATION
  /FORMAT SORT BLANK(.30)
  /CRITERIA FACTORS(5) ITERATE(25)
  /EXTRACTION PC
  /CRITERIA ITERATE(25) DELTA(0)
  /ROTATION OBLIMIN
  /METHOD=CORRELATION.

#consequent reliability analyses for newly proposed components of the condensed questionnaire

RELIABILITY
  /VARIABLES=USQ_28 USQ_25 USQ_29 USQ_30 USQ_27 USQ_39 USQ_37 USQ_26 USQ_38
USQ_24 USQ_22
  USQ_34 USQ_23 USQ_16 USQ_35
  /SCALE('ALL VARIABLES') ALL
  /MODEL=ALPHA
  /STATISTICS=DESCRIPTIVE
  /SUMMARY=TOTAL.

RELIABILITY
  /VARIABLES=USQ_1 USQ_2 USQ_3 USQ_4 USQ_5 USQ_6
```

```
  /SCALE('ALL VARIABLES') ALL
  /MODEL=ALPHA
  /STATISTICS=DESCRIPTIVE
  /SUMMARY=TOTAL.

RELIABILITY
  /VARIABLES=USQ_40 USQ_41 USQ_42
  /SCALE('ALL VARIABLES') ALL
  /MODEL=ALPHA
  /STATISTICS=DESCRIPTIVE
  /SUMMARY=TOTAL.

RELIABILITY
  /VARIABLES=USQ_19 USQ_20 USQ_21
  /SCALE('ALL VARIABLES') ALL
  /MODEL=ALPHA
  /STATISTICS=DESCRIPTIVE
  /SUMMARY=TOTAL.

RELIABILITY
  /VARIABLES=USQ_32 USQ_13 USQ_14 USQ_33 USQ_31
  /SCALE('ALL VARIABLES') ALL
  /MODEL=ALPHA
  /STATISTICS=DESCRIPTIVE
  /SUMMARY=TOTAL.

# Descriptives for gender, first-time usage, geekism and institution-based trust

FREQUENCIES VARIABLES= Gender first_time_user geekism_total trust_total
  /STATISTICS=STDDEV MEAN MINIMUM MAXIMUM
  /ORDER=ANALYSIS.

#linear regressions for gender, first-time usage, geekism, and institution-based trust

BOOTSTRAP
  /SAMPLING METHOD=SIMPLE
  /VARIABLES TARGET=usq_total INPUT=  gender_dummy
  /CRITERIA CILEVEL=97.5 CITYPE=BCA  NSAMPLES=9999
  /MISSING USERMISSING=EXCLUDE.
REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS CI(95) R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
```

```
  /DEPENDENT usq_total
  /METHOD=ENTER gender_dummy
  /SCATTERPLOT=(*ZRESID ,*ZPRED)
  /RESIDUALS HISTOGRAM(ZRESID) NORMPROB(ZRESID).

BOOTSTRAP
  /SAMPLING METHOD=SIMPLE
  /VARIABLES TARGET=usq_total INPUT=  firsttime_dummy
  /CRITERIA CILEVEL=97.5 CITYPE=BCA  NSAMPLES=9999
  /MISSING USERMISSING=EXCLUDE.
REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS CI(95) R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT usq_total
  /METHOD=ENTER firsttime_dummy
  /SCATTERPLOT=(*ZRESID ,*ZPRED)
  /RESIDUALS HISTOGRAM(ZRESID) NORMPROB(ZRESID).


BOOTSTRAP
  /SAMPLING METHOD=SIMPLE
  /VARIABLES TARGET=usq_total INPUT=  geekism_total
  /CRITERIA CILEVEL=97.5 CITYPE=BCA  NSAMPLES=9999
  /MISSING USERMISSING=EXCLUDE.
REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS CI(95) R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT usq_total
  /METHOD=ENTER geekism_total
  /SCATTERPLOT=(*ZRESID ,*ZPRED)
  /RESIDUALS HISTOGRAM(ZRESID) NORMPROB(ZRESID).

BOOTSTRAP
  /SAMPLING METHOD=SIMPLE
  /VARIABLES TARGET=usq_total INPUT=  trust_total
  /CRITERIA CILEVEL=97.5 CITYPE=BCA  NSAMPLES=99990
  /MISSING USERMISSING=EXCLUDE.
REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS CI(95) R ANOVA
```

```
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT usq_total
  /METHOD=ENTER trust_total
  /SCATTERPLOT=(*ZRESID ,*ZPRED)
  /RESIDUALS HISTOGRAM(ZRESID) NORMPROB(ZRESID).
```

Appendix F

F1 Informed Consent (before Covid-19)

Research project title: Chatbots as a tool for the future? An analysis of the USQ and user-characteristics

Student investigator: Alexander Dehmel

Research participant name:

Thank you for in this study. This research is part of my Bachelor thesis in Psychology at the University of Twente. The purpose of the research is to test and validate a preliminary questionnaire assessing user satisfaction with chatbot interaction. For this, you will interact with five chatbots and will perform to tasks for each. After that, the questionnaire as well as an item measuring task difficulty and your demographics will be filled out. Demographic data will be used to see if certain characteristics like previous experience with chatbots have a significant effect on the experienced user satisfaction regarding the chatbots. The test will take between 30 minutes up to one hour. Your test data will be processed anonymously. I do not anticipate that there are any risks associated with your participation, but you do have the right to stop the interview or withdraw from the research at any time. The research project has been reviewed and approved by the BMS Ethics Committee. Ethical procedures for academic research undertaken from Ethics Committee of the Faculty of Behavioural, Management and Social Sciences at the University of Twente, require that participants explicitly agree to be interviewed and how the information contained in their interviews will be used. This content form is necessary for us to ensure that you understand the purpose of your involvement and that you agree to the conditions of your participation. I would like you, therefore, to read the accompanying consent form and then sign this form to certify that you approve the following:

YOU WILL BE GIVEN A COPY OF THIS INFORMED CONSENT FORM

Please tick the appropriate boxes	**Yes**	**No**
Taking part in the study		
I consent voluntarily to be a participant in this study and understand that I can refuse to complete a task, and I can withdraw from the study at any time, without having to give a reason.	□	□
I understand that taking part in the study involves an audio-recording as well as a recording of the laptop screen. The data will be treated with discretion.	□	□
I agree with the collection of my age, gender, nationality, educational background and experience with chatbots. These data will be anonymized after this session.	□	□
I understand that the information I provide will be used for the Bachelor thesis of the lead researcher of this study.	□	□
I understand that personal information collected about me that can identify me, such as [e.g. my name or where I live], will not be shared beyond the study team.	□	□

Signatures

_____	_____	_____
Name of participant	Signature	Date

I have accurately read out the information sheet to the potential participant and, to the best of my ability, ensured that the participant understands that they are freely consenting.

_____	_____	_____
Name of researcher	Signature	Date

Study contact details

Principal researcher

Alexander Dehmel

a.dehmel@student.utwente.nl

Co-investigator/Supervisor

Dr. Simone Borsci

s.borsci@utwente.nl

Contact information for questions about your rights as a research participant

If you have questions about your rights as a research participant or wish to obtain information, ask questions, or discuss any concerns about this study with someone other than the researcher(s), please contact the Secretary of the Ethics Committee of the Faculty of Behavioural, Management and Social Sciences at the University of Twente by ethicscommittee-bms@utwente.nl

F2 Informed Consent via Qualtrics (during Covid-19)

Start of Block: informed consent

Informed consent

Student investigator: Alexander Dehmel

Surname of participant:

_____

Thank you for participating in this study. This research is part of my Bachelor thesis in Psychology at the University of Twente. The purpose of the research is to test and validate a preliminary questionnaire assessing user satisfaction with chatbot interaction. For this, you will interact with five chatbots and will perform two tasks for each. After that, the questionnaire as well as an item measuring task difficulty and your demographics will be filled out.

Demographic data will be used to see if certain characteristics like previous experience with chatbots have a significant effect on the experienced user satisfaction regarding the chatbots. The test will take between 30 minutes up to approximately one hour. Your test data will be processed anonymously. I do not anticipate that there are any risks associated with your participation, but you do have the right to stop the interview or withdraw from the research at any given time. The research project has been reviewed and approved by the BMS Ethics Committee.

Ethical procedures for academic research undertaken from Ethics Committee of the Faculty of Behavioral, Management and Social Sciences at the University of Twente, require that participants explicitly agree to be interviewed and how the information contained in their interviews will be used. This content form is necessary for us to ensure that you understand the purpose of your involvement and that you agree to the conditions of your participation. I would like you, therefore, to read the accompanying consent form and agree to certify that you approve the following:

I consent voluntarily to be a participant in this study and understand that I can refuse to complete a task and I can withdraw from the study at any time, without having to give a reason

- ○ Yes (1)
- ○ No (2)

I understand that taking part in the study involves an audio-recording as well as a recording of the screen, if possible. The data will be treated with discretion.

- ○ Yes (1)
- ○ No (2)

I agree with the collection of my age, gender, nationality, educational background and experience with chatbots. These data will be anonymized after this session.

- ○ Yes (1)
- ○ No (2)

I understand that the information I provide will be used for the Bachelor thesis of the lead researcher of this study.

- ○ Yes (1)
- ○ No (2)

I understand that personal information collected about me that can identify me, such as [e.g. my name or where I live], will not be shared beyond the study team.

- ○ Yes (1)
- ○ No (2)

I have accurately read the information sheet and was informed by the lead researcher in an appropriate manner. I understand that at the end of this survey, I will be given sufficient contact information in case of further questions.

- ○ Yes (1)
- ○ No (2)

Thank you for filling out the informed consent. The study will start now.

End of Block: informed consent

Appendix G

Table G1

Chatbot tasks

Chatbot	Task 1	Task 2	Link to website
ATO	You moved to Australia from the Netherlands recently. You want to know when the deadline is to lodge/submit your tax return using ATO's chatbot to find out.	You are a student and are wondering whether you have to lodge a tax return using the ATO's chatbot.	http://www.ato.gov.au/
HSBC UK	You live in the Netherlands but are travelling to Turkey for two weeks. During your travel, you would like to be able to use your HSBC credit card overseas at payment terminals and ATMs. You want to use HSBC's chatbot to find out the relevant procedure.	You have recently moved from Amsterdam to London and would like to know how you can change your address for your HSBC card, using the chatbot of HSBC UK.	https://www.hsbc.co.uk/
Absolut	You want to buy a bottle of Absolut vodka to share with your friends for the evening. One of your friends cannot consume gluten. You want to use Absolut's chatbot to find out if Absolut Lime contains gluten or not.	You want to buy a bottle of Absolut vodka for a good friend. But this friend is right now on a diet and tries to avoid sugar. You, therefore, want to find information about the amount of sugar in the products of Absolut using Absolut's chatbot.	https://www.absolut.com/en/
Booking.com	You are travelling to London from 5th July to 9th July with your family. You want to use booking.com's chatbot to find a hotel room for you, your	You have to attend an important business meeting from 18th to 19th	https://www.facebook.com/messages/t/131840030178250

	significant other and your child in Central London that does not cost more than 500€ in total	of March in Amsterdam. You, therefore, are looking for a place to stay in the city centre of London for not more than 200€ using the booking.com chatbot.	
USCIS	You are a US citizen living abroad and want to vote in the upcoming federal elections. You want to use the USCIS chatbot to find out how.	You are planning to take a job in the USA. Since you are not a US citizen, you want to find out more about eligibility for a US-Green Card with the help of the USCIS chatbot.	http://www.uscis.gov/emma
Emirates Holidays	You just woke up and realised that you forgot that it's your significant other's birthday. Desperately, you are thinking about a birthday present and your idea is a holiday together in Paris. You visit the Emirates Holidays page and use Emirates Holidays' chatbot to book a holiday from the 4th September until the 9th September to Paris for two persons. Your departure airport is London Heathrow (LHR). Everything else is not important, as you just need a present for today.	You arrived in Paris and there seems to be a problem with your hotel reservation. You try to call someone at Emirates Holiday, but it is 11 pm on Friday, so you cannot reach anyone. Hence, you ask Emirates Holidays' chatbot when the customer service opens on Saturday.	https://www.emiratesholidays.com/gb_en/
HubSpot	You have your own company and would like to grow your business even more. A former colleague recommends you HubSpot. However, you do not want to sign up for anything (even if it's free). You use	Now, you are convinced that HubSpot can help your own business. Your focus is on improving your own customer service. Before you sign up for something, you would like	https://www.hubspot.com/?survey=123

	HubSpot's chatbot to purely get information and get educated without using any tools. A collection of news/articles/tips would be great.	to know how HubSpot can improve your customer service. You use HubSpot's chatbot to get more information about this.	
Amtrak	You would like to travel from Boston to Washington D.C. while being in the USA. You want to use Amtrak's chatbot to book the shortest trip possible on the 8th of October. Your departure station is Back Bay Station.	You have planned a trip to the USA. You are planning to travel by train from Boston to Washington D.C. You want to stop at New York to meet an old friend for a few hours and see the city. You want to use Amtrak's chatbot to find out how much it will cost to temporarily store your luggage at the station.	https://www.amtrak.com/home
Utwente	You are a Chinese student who would like to do a master's degree at the University of Twente. Your name is Jackie/Lin and your Email address is abc@def.com. You are interested in doing your master in Nanotechnology in September 2021. You did your bachelor at the Utwente in the Netherlands. You ask the Utwente chatbot what options for a scholarship are available.	You are a German student who would like to do a master's degree at the University of Twente. Your name is Alan/Sabine and your Email address is abc@def.com. You are interested in doing your master's in computer science in February 2022. You did your bachelor's at the Jacobs University in Bremen. You ask the Utwente chatbot about	https://www.utwente.nl/en/education/master/chat/

		deadlines and the admission process.	
NBC News	You want to use the chatbot of NBC News to find out the most recent news regarding the environment.	Just out of curiosity, you are also interested in the most recent special coverage, using the chatbot of NBC News.	https://www.facebook.com/NBCNews/
ManyChat	You want to integrate a chatbot on your companies' website. Therefore, you want to use the ManyChat's chatbot to find video tutorials to learn the basics of ManyChat.	After using the Chatbot for a while, you are getting a little bored and want to have some fun. Let the ManyChat's chatbot tell a joke to you.	https://www.messenger.com/t/ManyChat

Appendix H

Table H1

Oblique rotated factor loadings[a]

Item	Description of item[b]	C1	*C2*	*C3*	*C4*	*C5*
USQ_28	Maxim of quantity	.85				
USQ_29	Maxim of quantity	.84				
USQ_30	Relevant information	.83				
USQ_25	Maxim of relation	.82				
USQ_39	Perceived credibility	.77				
USQ_37	Perceived credibility	.70				
USQ_38	Perceived credibility	.70			.30	
USQ_27	Relevant information	.68				
USQ_26	Maxim of relation	.63				
USQ_22	Recognition and facilitation of goal	.57				.30
USQ_24	Recognition and facilitation of goal	.56				
USQ_34	Understandability	.54				
USQ_35	Understandability	.51				
USQ_16	Relevant service	.50				
USQ_15	Relevant information	.47				
USQ_23	Relevant service	.47				.37
USQ_18	Reference to service	.41				

USQ_12	Flexibility of linguistic input	.40				.33
USQ_36	Understandability	.39				
USQ_17	Use of hyperlinks					
USQ_1	Ease of starting a conversation		.85			
USQ_5	Visibility		.85			
USQ_6	Visibility		.83			
USQ_4	Visibility		.82			
USQ_2	Ease of starting a conversation		.79			
USQ_3	Ease of starting a conversation		.73			
USQ_41	Response time			.98		
USQ_42	Response time			.96		
USQ_40	Response time			.92		
USQ_21	Perceived privacy and security				.74	
USQ_19	Perceived privacy and security				.70	
USQ_20	Perceived privacy and security				.64	
USQ_11	Flexibility of linguistic input	.33			-.57	
USQ_10	Flexibility of linguistic input				-.51	
USQ_32	Graceful responses					.69
USQ_13	Ongoing conversation					.65
USQ_33	Graceful responses					.64
USQ_14	Awareness of context					.56

USQ_31	Graceful responses					.51
USQ_7	Expectation setting	.31				.40
USQ_9	Expectation setting					.33
USQ_8	Expectation setting					
Eigenvalues		13.1	3.86	2.58	2.49	1.71
Percentage of variance		31.17	9.19	6.15	5.92	4.07

[a] factor loadings < .3 suppressed

[b] descriptions mainly taken from Boecker and Borsci (2019)

Table H2

The final version of oblique rotated factor loadings[a]

Item	Description of item[b]	C1	*C2*	*C3*	*C4*	*C5*
USQ_28	Maxim of quantity	.86				
USQ_29	Maxim of quantity	.86				
USQ_25	Maxim of relation	.85				
USQ_30	Relevant information	.81				
USQ_39	Perceived credibility	.76				
USQ_27	Relevant information	.72				
USQ_38	Perceived credibility	.72				
USQ_37	Perceived credibility	.71				
USQ_26	Maxim of relation	.67				
USQ_24	Recognition and facilitation of goal	.61				.31

USQ_22	Recognition and facilitation of goal	.58				.35
USQ_34	Understandability	.57				
USQ_23	Relevant service	.56				.37
USQ_16	Relevant service	.55				
USQ_35	Understandability	.54				
USQ_1	Ease of starting a conversation		.85			
USQ_4	Visibility		.83			
USQ_5	Visibility		.83			
USQ_6	Visibility		.81			
USQ_2	Ease of starting a conversation		.79			
USQ_3	Ease of starting a conversation		.71			
USQ_41	Response time			.97		
USQ_42	Response time			.95		
USQ_40	Response time			.91		
USQ_19	Perceived privacy and security				.84	
USQ_21	Perceived privacy and security				.84	
USQ_20	Perceived privacy and security				.79	
USQ_32	Graceful responses					.65
USQ_13	Ongoing conversation					.65
USQ_33	Graceful responses					.64

USQ_14	Awareness of context					.59
USQ_31	Graceful responses					.53
Eigenvalues		10.53	3.77	2.49	2.2	1.6
percentage of variance		32.9	11.77	7.78	6.88	5.03

[a] factor loadings < 3 suppressed

[b] descriptions mainly taken from Boecker and Borsci (2019)