Online Disclosure and Privacy Measures

How we protect our personal information online

R.V.R. van der Valk

SUPERVISORS

First supervisor:	prof. dr. ir. R.W. Helms
Second supervisor:	dr. F.J. Bex

“There is no such thing as an ending, or a beginning for that matter, everything is middle.”

- Eoin Colfer

With special thanks to my friends

Anke van Asten, Sarah Hoes, Vicky Trouerbach, Sanne Bleeker, Josje Hiddink, Lisette Mijnen, Guido de Groot and Ellis Smal

for their company and support during the long days in the University library.

Abstract

Online disclosure is a key element of many online platforms, as well as an important factor for numerous scientific studies. However, privacy concerns often limit the willingness of users and respondents to share their personal information. For this reason, privacy measures have been implemented on many online platforms. The relationship between online disclosure and these privacy measures still needed to be confirmed. Our study used an online survey assisted by a data collection tool to measure the respondents' privacy concern, online disclosure, knowledge about and use of privacy measures. Of these 4 factors, we found that people who were more concerned about their online privacy were less likely to disclose personal information, and more likely to have strict privacy settings. Moreover, we found that people with strict privacy settings were likely to share less information.

Table of Contents

1. Introduction

The past decade has seen the rise of several immensely popular online social media platforms. At the same time, concerns about privacy were brought into the spotlight and formed the basis for a public debate on online disclosure on these websites (Stutzman, Gross & Acquisti, 2013; Acoca, 2008). Considering that a social network thrives on content shared by its users, these concerns needed to be addressed in order to sustain the success of the platform (Nosko, Wood & Molema, 2010).

Since its initial launch, Facebook has implemented a number of privacy measures that aim to provide the user the control needed to determine who is allowed to see what information. These measures have changed over time (Acquisti & Gross, 2006; Appendix B), and will likely continue to do so. At the time of writing, users have control over a variety of sharing options, including general profile visibility, visibility specific personal details (such as gender and date of birth) and the visibility of their status updates.

1.1 Problem statement

The evolution in the availability of privacy measures was directly related to increasing concerns about the risks of online disclosure (Acquisti & Gross, 2006). As the platform grew more popular, it became clear that the information present on Facebook could be used in a number unintended, negative ways (Christofides, Muise & Desmarais, 2012). These include stalking, identity theft and bullying. Moreover, information and photos on the social media platform could have an adverse effect on the user's reputation in more formal settings. Unflattering pictures posted on Facebook might play a role when deciding who to hire for a job, or status updates about risky or irresponsible activities could be picked up by insurance agencies and affect their fees and responses to claims filed. Clearly, there is good reason for concerns about what is disclosed, and who can see this information.

Theoretically, the privacy measures put in place should take away some user objections about disclosing personal information (Tufekci, 2008). However, this relationship has not been proven, although it has been the object of study (Christofides, Muise & Desmarais, 2009). As such, it forms an interesting basis for our study.

1.2 Research questions

This research thesis is focussed on answering one main research question, namely: *what is the effect of the user's knowledge about and use of Facebook's privacy measures on their online disclosure behaviour?* In order to provide a complete and appropriate answer to this question, 4 sub questions were formulated. These are:

1. To what extent is knowledge about Facebook's privacy measures related to online disclosure behaviour?
2. To what extent is use of Facebook's privacy measures related to online disclosure behaviour?
3. To what extent are general privacy attitudes related to the user's knowledge about and use of Facebook's privacy measures?
4. To what extent are general privacy attitudes related to the user's disclosed information?

1.3 Research approach

Based on existing literature, a model of concepts will be constructed to explain how they are related. The 4 concepts in this model are privacy concern, online disclosure, knowledge about and use of privacy measures. Next, this model will be verified using the data collected with an online survey. Apart from the factors mentioned, the survey will ask respondents for a number of general demographics, which will serve as control variables. Using the data collected, linear regression tests will be used to verify the relationships in the model.

One important aspect to note is that the online survey will be assisted by a data collection tool, which will allow us to verify part of the data collected. As earlier studies found that a significant part of respondents were unable to correctly report on their disclosure behaviour (Acquisti & Gross, 2006), this tool will strengthen the validity of the data, adding to the overall trust in the study.

1.4 Relevance

Answering the question at hand could be of great value for science, business, and society alike. Starting with the societal contribution, the outcome of this thesis project can be of great interest both to individuals and companies. If a greater understanding of online disclosure behaviour is developed, it may be important for individuals to be aware of this. For instance, if it is shown that greater use of privacy measures increases the likelihood of disclosure, the decision to disclose should still be a conscious, rational decision, rather than based on a sense of trust. The information disclosed might still be used for other purposes by the social network itself, which might be what the user is trying to avoid by using these privacy measures.

On the other hand, companies may be able to use the findings in order to create a situation in which users are comfortable making the decision to disclose their information. Considering the interest in business intelligence that has grown over the past years (Sharda, Aronson & King, 2008), more and more data is sought after to produce the best results from these analyses. If users can be persuaded to disclose more information based on the findings, this would be very interesting for businesses.

Looking at the scientific contribution of the research thesis, many studies have been conducted into disclosure behaviour (a selection of which can be seen in section 4). The relationship between privacy options and disclosure behaviour, however, has not been researched thoroughly. In a paper by Stutzman, Capra & Thompson (2010), the relationship between privacy measures and disclosure behaviour was researched, but was partially inconclusive. This was likely due to a limited amount of information being collected about the participant's use of these measures (see section 4.4). Aside from this study, other papers have also hinted at a relationship (Stutzman, Gross & Acquisti, 2013). This research thesis aims to provide a clear and proper answer to this question.

1.5 Scope

This thesis will limit its focus to Facebook users. There are multiple reasons for this. Firstly, it is currently one of the largest and best-known social networks. Secondly, it provides clear categorisation of different items a person may choose to disclose. Lastly, the platform allows for the use of automated retrieval of personal information if a user logs in and gives permission to the application at hand, making it possible to implement the aforementioned tool. These factors combined make this social network an attractive basis for our study.

2. Related works

This section will discuss earlier studies that have focussed on online disclosure behaviour, or any of its related aspects. This will help to provide a context for this study, both in a sense of what is already known, but also in that of other factors that may affect the outcome of this study.

First, the method through which these papers were collected will be explained. In the following sections the findings of these papers will be explained. To start, several aspects of the papers found were gathered and analysed collectively in order to find trends and similarities. After analysing this, general studies on online disclosure behaviour will be discussed. The next section will focus on the related topic of privacy awareness. After this, online disclosure behaviour will be examined from 4 different perspectives, namely social, business, technical and legal. The last section of papers will discuss studies that have constructed, tested and confirmed models examining online disclosure behaviour. This chapter will then conclude with a brief discussion of what the information presented in this section means for the study as a whole.

2.1 Method

This related works chapter is based on the systematic literature review method of Free et al. (2013). The databases used were primarily Google Scholar and Scopus. The papers were found using various combinations of the keywords “disclosure”, “online disclosure”, “privacy”, “awareness”, “security” and “Facebook”. The complete list of queries used can be found in Appendix A. The resulting papers were then selected on relevance based on their placement in the results list, their title, their abstract and finally their text. The process can be seen in figure 2.1.

The papers used were selected on each step based on the criteria that they (significantly) add to the understanding of the concepts discussed, how they are influenced by other factors, or what aspects of them are studied from different perspectives. In the first step, papers found using online database searches were also selected based on relevance due to their placement on the results list. For each search, the list was examined in sets of 10 (this being the amount on a page). In case no relevant papers were found in this set, the remaining results were not examined. At very least, the first 4 results pages of each search were examined. Apart from the papers found this way, a small number of other relevant papers were included in order to paint a significantly more complete picture of the topic at hand. These were generally found using back-searching based on specific references in papers already included.

It should be noted that, unlike the study by Free et Al., all papers that made it through the selection process were included in the meta-analysis.

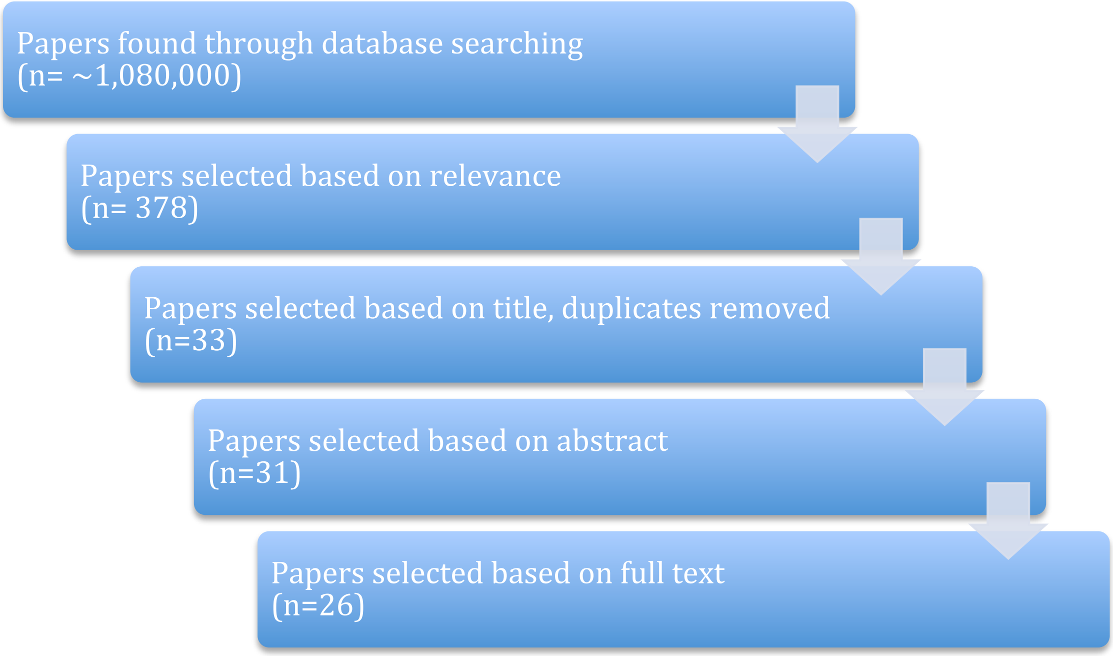

Papers found through database searching (n= ~1,080,000)
Papers selected based on relevance (n= 378)
Papers selected based on title, duplicates removed (n=33)
Papers selected based on abstract (n=31)
Papers selected based on full text (n=26)

Figure 2.1: Literature selection process

2.2 Meta-analysis of the literature found

Looking at all the literature found using the described method, some interesting observations can be made. In this section, several characteristics of the papers found and the studies they describe were collected so they could be examined together. This meta-analysis (Appendix A) shows a number of trends and similarities among the papers, which will be discussed in this section.

The first aspects studied were of a more superficial nature. Looking at the year of publication, not much of interest could be found. There seems to be a relatively equal division over the years 2005 through 2014. Also the country where the study took place was examined. This showed a strong interest from the United States, with more than half of the papers originating from this country. Together with Canada and UK, these countries were responsible for over three quarters of the papers found.

The papers were also compared in terms of the type of study that was conducted. As can be seen in Figure 2.2, the vast majority of papers found used quantitative methods to arrive at a conclusion. A small number of studies also used qualitative methods to form a basis for their later quantitative studies.

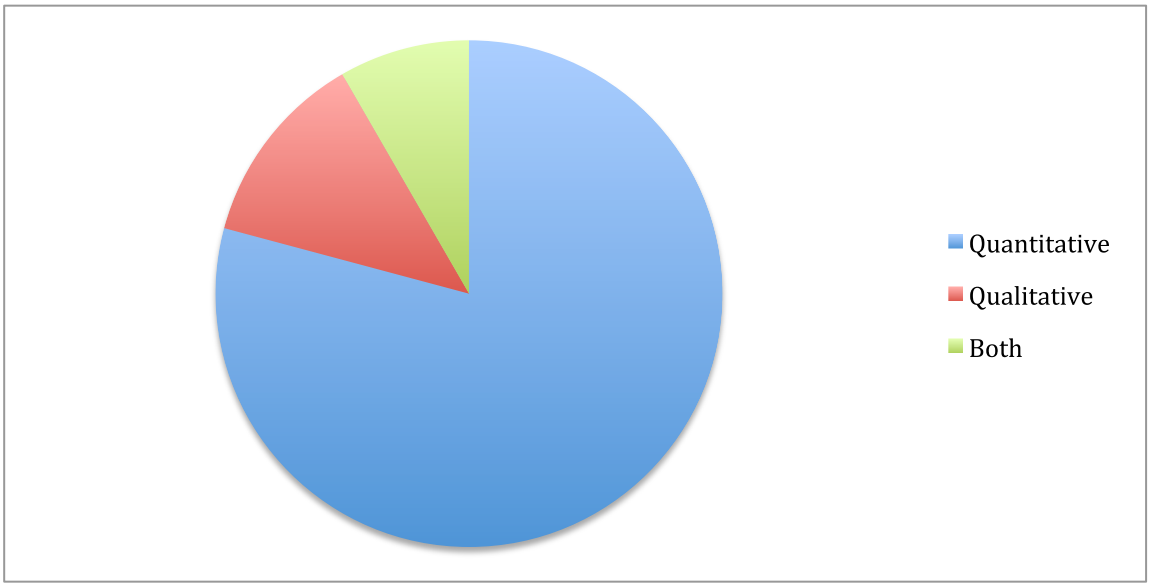


Figure 2.2: Methods used in studies found in the literature study.

The types of statistical tests were also compared for the papers found. As can be seen in Figure 2.3, there is an almost equal division between papers researching a correlation as there are papers researching regression. The methods used in these tests varied somewhat, correlations relying on parametric and non-parametric tests.

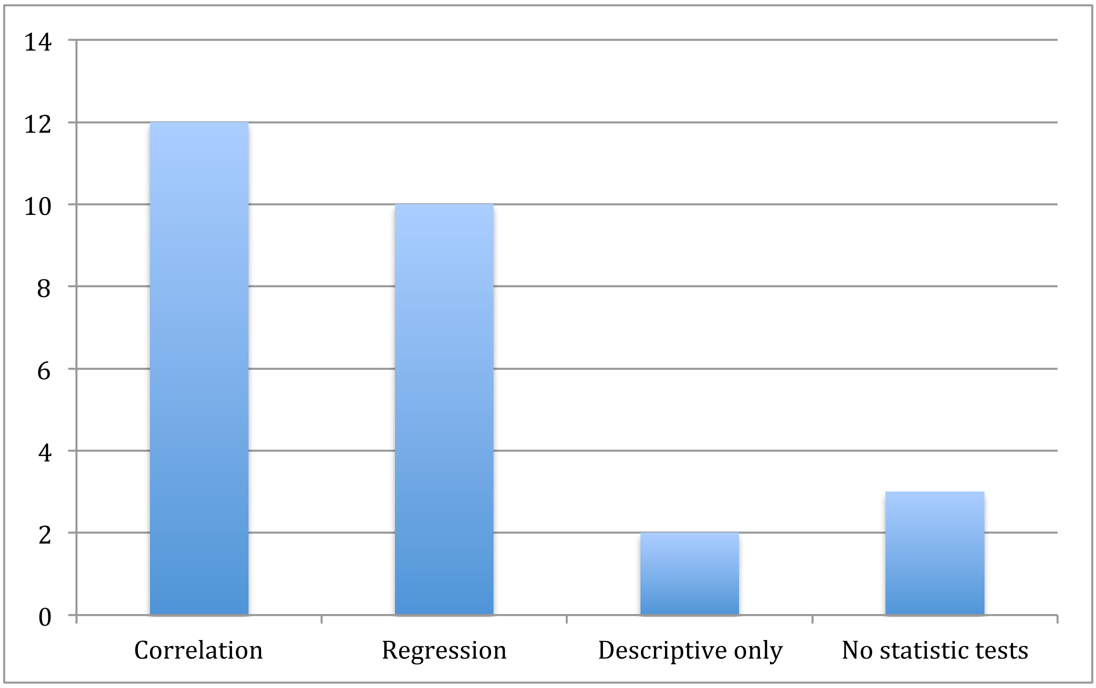


Figure 2.3: Statistical tests used in studies found in the literature study.

The last aspect of the papers that was compared were the research participants. Table 2.1 looks at what groups were examined by the various studies. As can be seen, the majority of studies requiring research participants focussed solely on college students. The reasons for this varied, some claiming that this focus group represented a large majority of the user base of the social network (Tufekci, 2008), while others did not explain their decision (Christofides, Muise & Desmarais, 2009) While an exclusive focus on college students probably does not accurately represent all the Facebook users, it does ease the concern that this study's participants will likely mostly consist of this members of this group.

Studies focussing solely on college students	11
Studies focussing solely on Facebook users in general	3
Studies without research participants	5
Others	7

Table 2.1: papers included in literature study by research participants

The amount of research participants was also examined. As can be seen in table 2.2, it appears that the average amount of participants for the studies was 640. However, considering that one single study contributed over 37% of the total amount of participants, this clearly greatly affects the average. As such, another comparison was made that excluded this outlier. This comparison can be seen in table 2.3.

Total amount of participants	13433
Average amount of participants	640
Lowest amount of participants	50
Highest amount of participants	5076

Table 2.2: statistics on research participants in literature study

Of the other studies, the average amount of research participants was 334, a significantly lower amount. Nevertheless, this is still a significant amount of participants for a study. This high average may be explained by the fact that most of these studies could be performed online and required little or no input of the research participant. This enables researchers to more easily gather a large amount of data.

Total amount of participants	8357
Average amount of participants	334

Table 2.3: statistics on research participants in literature study, excluding outlier

2.3 General online disclosure behaviour

Disclosure behaviour has been the subject of many studies, especially within the context of the Internet. This section will look at general studies looking into online disclosure behaviour. The following sections will build on this knowledge and explore different specific aspects.

An earlier study into privacy attitudes and sharing behaviour has shown that they have little to no relation (Acquisti & Gross, 2006). Both of these were measured using an online survey, with the former being measured using questions focussing on level of concern for privacy-related issues (interspersed with other current topics), and the latter being measured using more factual questions about the user's use of Facebook. The results of this study suggest that privacy attitudes were related to Facebook adoption, instead. It should be noted that this paper was published in 2006, when both Facebook and the public awareness of online privacy were very different from what they are today (Stutzman, Gross & Acquisti, 2012). A study performed in 2008 also found no relationship between privacy concerns and information disclosure, noting that participants managed access to the disclosed information using the privacy measures implemented on Facebook and Myspace (Tufekci, 2008). This study also used a more straight-forward approach of measuring the factors, e.g. with respondents being asked "how concerned are you with online privacy?", without interspersing this with other questions to avoid priming the respondents. Apart from this, the approach of measuring the factors used a similar approach, relying solely on survey questions.

In addition, the Acquisti & Gross (2006) found that users did not always report accurately on their own sharing behaviour. When asked whether they shared specific information, such as date of birth or political views, 77.84% of respondents correctly reported their disclosure behaviour. This indicates that using the tool as described in the research approach in chapter 4 would be of great importance, as to avoid the risk of inaccurate data.

Another important study concerns the changes in user disclosure behaviour between 2005 and 2011. In their paper, appropriately titled Silent Listeners: The Evolution of Privacy and Disclosure on Facebook, the authors examine the trends in disclosure behaviour over a long period of time (Stutzman, Gross & Acquisti, 2013). This was done by conducting a yearly examination of the publically available information of Facebook users at the Carnegie Mellon University. Their findings show that, over time, users have shown an increasing desire for privacy. However, they found the amount of personal information users shared privately actually increased over the same period of time. This strongly suggests that, indeed, the implementation of privacy measures encourages users to share more data about themselves. The authors point out that this increased disclosure also means that the "silent listeners" referred to in the title of their paper (which include Facebook, but also third parties) have more data at their disposal. Nevertheless, these trends can also be related to other changes in the use of Facebook. After all, correlation does not necessarily mean causation.

One important point concerning disclosure is the question whether people are more inclined to disclose personal information in online settings than they are offline. This has been put to the test in many studies, of which a systematic overview was created in 2012. The authors found 15 studies with a total of 24 comparisons. There was an equal divide in the number of comparisons claiming greater online disclosure, those claiming greater offline disclosure, and those claiming no difference at all. (Nguyen, Bin & Campbell, 2012). As such, this question remains unanswered.

The result of this paper could indicate several things. First of all, it could point to that disclosure simply is not related to whether it is online or offline. However, it could also be the result of a changing sense of how to handle online disclosure. The studies included by Nguyen et al. were published in a large range of years. Looking at a recent study (Emanuel et al., 2014), the results show that participants disclosed significantly less information in an online setting than in an offline setting. This trend towards more caution when disclosing information online could well be the result of the attention it has gotten over the past years, especially due to social networks such as Facebook.

Looking specifically at Facebook, several studies have examined what users are likely to disclose about themselves. In 2009, Christofides, Muise & Desmarais found most that Facebook users (though their sample consisted solely of students) disclosed personal information, such as birthday and e-mail address, but were also likely to post pictures of various types (profile pictures and pictures of events, some including alcohol). This was measured using an online survey, which asked respondents how likely they were to post specific kinds of information or posts. The authors also examined the possible relationship between disclosure and privacy measure use (called “information control”, in their paper). Unfortunately, they were unable to find a meaningful relationship between the two, and provide very little information about the details of how this was measured.

Another study from 2009 looked into what disclosed information was publically available, and attempted to sort it into different groups according to how sensitive the information was conducted by Nosko, Wood & Molema. The authors found that a significant amount of information was publically available on the randomly sampled profiles. A majority of the profiles found allowed anyone to see their wall, photos they had posted, photos they were tagged in, education history and even sexual orientation. Moreover, they found that disclosure of gender, relationship status and age could be used to predict the amount of sensitive information disclosed by the user.

2.4 Privacy awareness

Numerous studies have tackled the topic of the user’s privacy awareness in relation to social media. One such study was performed in 2005, in the early years of Facebook’s existence (Govani & Pashley, 2005). At this time, Facebook use was limited to college students, giving a somewhat different perspective to privacy on this network. The study was conducted using a survey asking the students what kinds of information they were willing to disclose, and what their motivation was for doing so. This survey was further assisted by copies of the respondent's Facebook profile that were made prior to them taking the survey. The authors conclude that, while users were generally well aware of the consequences of disclosing their information, they were comfortable with it. The authors also note that many users reported not limiting access to their personal information, despite their knowledge of this option.

Later studies indicate that privacy awareness has changed since this early study. For example, a study by Tuuainen, Pitkänen and Hovi (2009), in which 210 Facebook users were surveyed, concludes that while the participants disclosed a substantial amount of personal information, they were generally not well aware of who could see this information, nor had they read or understood Facebook's privacy policy or terms of use. The first two findings of this paper are of big interest, as they show a lack of (proper) privacy awareness. However, this is not necessarily the case with the small amount of users having read the privacy policy and terms of use. After all, these documents can be quite lengthy and full of legal terms. Users may get a better understanding of the content of these documents by reading analyses written in more shorter, more understandable fashion. Nevertheless, the findings of this paper show that users generally have a limited awareness of the possible privacy concerns, while still disclosing a substantial amount of information. Unfortunately, the authors did not describe extensively how they measured the various concepts using the online survey, making it hard to further examine their method.

2.5 Motivators and consequences of disclosure

In order to understand disclosure, many studies have looked into what motivates people to volunteer their personal information. Another common theme among these papers is analysing events that may cause a user to significantly change their sharing behaviour, usually due to a negative experience as a result of disclosure. This section will focus on these two themes.

2.5.1 Motivations for disclosure

In a study comparing the Facebook disclosure behaviour of adolescents and adults, several motivating factors were researched (Christofides, Muise & Desmarais, 2011). The researchers found that for both groups, self-esteem, trust, need for popularity, and awareness of consequences were found to be related to disclosure. It should be noted that the last of these variables is not a motivational factor. As would be expected, this factor was found to be negatively related to disclosure, while the others were positively related.

Looking to a more general case of online disclosure, a study by Hui, Tan & Goh (2006) researched what factors could persuade customers of Internet businesses to volunteer private information. The seven factors found were, in order of strongest overall preference: time saving, novelty, monetary saving, pleasure, social adjustment, self-enhancement and altruism. Obviously, not all of these factors are relevant for social media. For instance, monetary saving is very rarely applicable to these platforms. Nevertheless, this does add to the overall picture of what motivates people to disclose personal information.

One study focussed specifically on encouraging disclosure of personal information in exchange for commercial benefits (Heirman, Walrave & Ponnet, 2013).In their study among 1,042 adolescents, they found that the user's privacy concern and trust propensity were important factors in whether users were willing to disclose personal information for commercial ends. Interestingly enough, this attitude differs from the results from the study by Tufekci (2008), which found that privacy concerns were not related to disclosure on the social network, but rather of the use of privacy measures. This could suggest 2 things. First of all, it could point to a sense of control over the disclosed information on Facebook, giving the user more trust in the system. Alternatively, it could suggest that the purpose for which the disclosed information will be used plays an important role in the decision. For instance, disclosing your place of residence on Facebook could serve to allow friends to find you more easily, while disclosing it on a marketing survey serves only the company's commercial goals.

In section 2.6, several papers will be discussed that have modelled factors related to online disclosure and privacy behaviour. Some of these models also include variables describing the motivation for disclosure, including perceived social benefits (Wilson, Proudfoot & Valacich, 2014) and the relevance of the disclosed information to the functioning of the system (Zimmer et al., 2010). The relationships with other factors will be discussed in the later section.

2.5.2 Consequences of disclosure

There are a number of negative consequences that online disclosure can have. These can range from identity theft to unwanted exposure of personal information. The former is a risk made even stronger by widespread use of personal questions to establish a person's identity. Because of Facebook and other social media, these questions have become increasingly easy to answer for others, decreasing any system relying on these kinds of questions. In 2008, a study by Rabkin surveyed several banks for their set of fall-back questions to verify the user's identity. He found that the questions used were "surprisingly weak", remarking that even without the aid of social networks, the knowledge of answer frequency could allow for a significant amount of false positives in these tests. When using the information available on social networks, this risk obviously increases.

In 2009, Debatin et al. performed a study that focussed on the attitudes towards negative effects of online personal information disclosure. They found that users were generally well aware of the possible consequences of their disclosure behaviour, but often did not act (significantly) on this knowledge. However, users who had experienced an invasion of their privacy as a result of this were shown to be more likely to change their privacy settings after the incident. It should be noted that these incidents also include privacy invasions by third parties, such as advertisers, acting on Facebook. Thus, the authors conclude that merely changing the privacy settings on the social network is inadequate, as it does not protect against these actions.

Another paper focussing on the effect of negative experiences, appropriately titled “Risky Disclosures on Facebook”, was published in 2012 and looked into the effect of these experiences on the knowledge of and use of Facebook’s privacy measures among adolescents. (Christofides, Muise & Desmarais, 2012) It should be noted that these experiences in this study were not limited to unintentional disclosure, but also included “bullying/meanness, unwanted contact, [..] and misunderstandings”. Like the earlier study, they found that those who reported having a bad experience with Facebook were likely to know more about Facebook’s privacy settings and to use them to restrict access to their profile.

2.6 Models describing online disclosure and privacy behaviour

Other studies have studied factors that influence online disclosure and privacy behaviour. This section will discuss several papers that have constructed models describing these factors, and have verified them using quantitative studies. Since some of these models are quite complex, an accompanying figure was constructed to clarify the description in the text.

In 2011, Christofides, Muise & Desmarais researched online disclosure behaviour of both adolescents and adults. Based on an online survey, they found that time spent on Facebook, duration of membership, gender, general trust, self-esteem and awareness of disclosure consequences could all be used to predict use of privacy measure usage. The strongest relationship for both age groups was found with the awareness of consequences. (Christofides, Muise & Desmarais, 2011). The factors can be seen together in Figure 2.4.

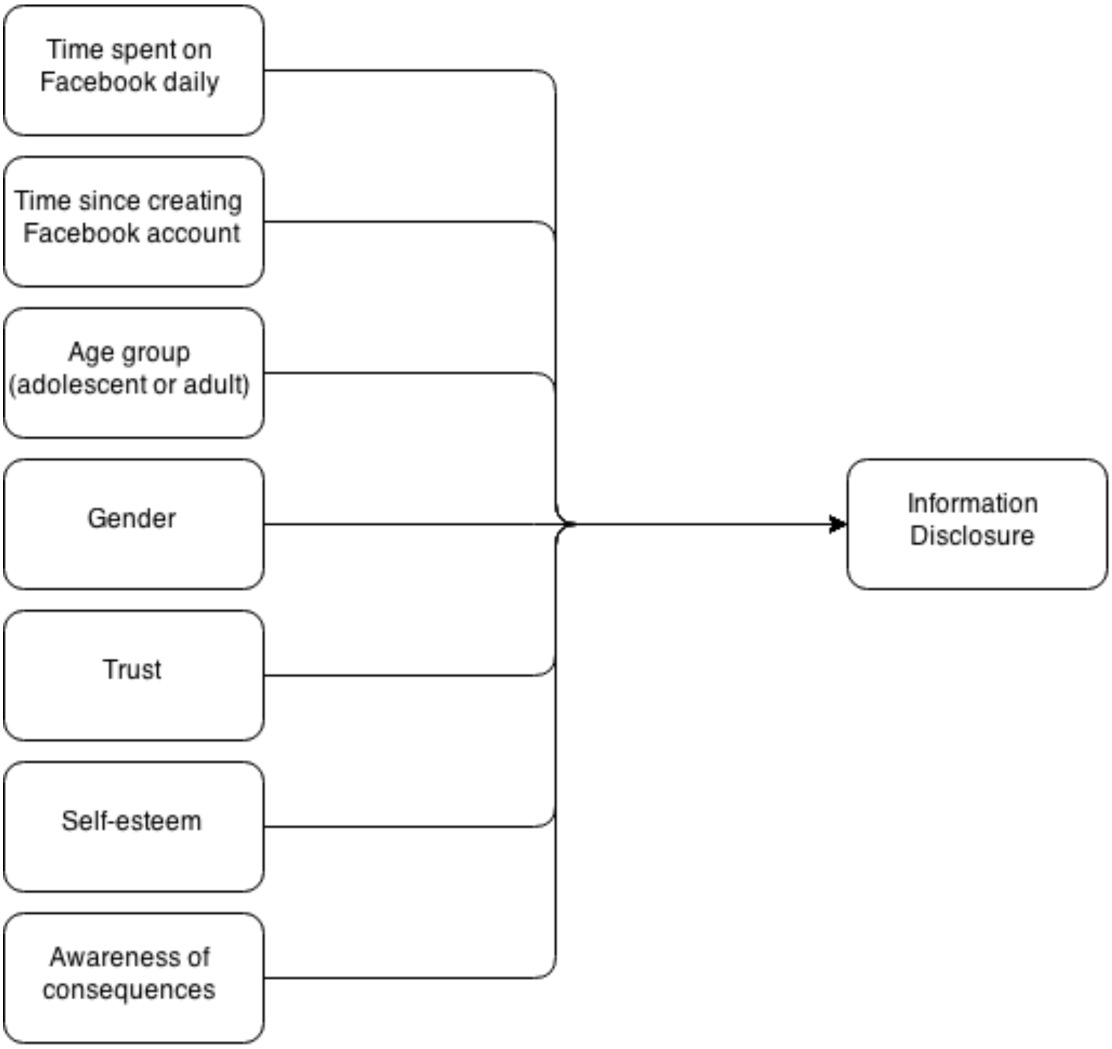

Time spent on Facebook daily
Time since creating Facebook account
Age group (adolescent or adult)
Gender
Trust
Self-esteem
Awareness of consequences
Information Disclosure

Figure 2.4: Factors that, together, can be used to predict information disclosure of a Facebook User, according to Christofides, Muise & Desmarais (2011)

A similar study conducted the year after explored the effects of several factors have on privacy measure use, but in this case all mediated by the variable of information privacy concern (Mohamed & Ahmad, 2012). They found a significant relationship between this mediating variable and privacy measure use. Furthermore, they found that self efficacy(*sic*), perceived severity of the privacy threat, perceived vulnerability and gender were all significantly related to information privacy concern. They also researched the possible relationship between response efficacy (the ability to protect oneself from the perceived threat) and reward for disclosure, but these hypotheses were not supported by the results. The results image of this study can be seen in Figure 2.5.

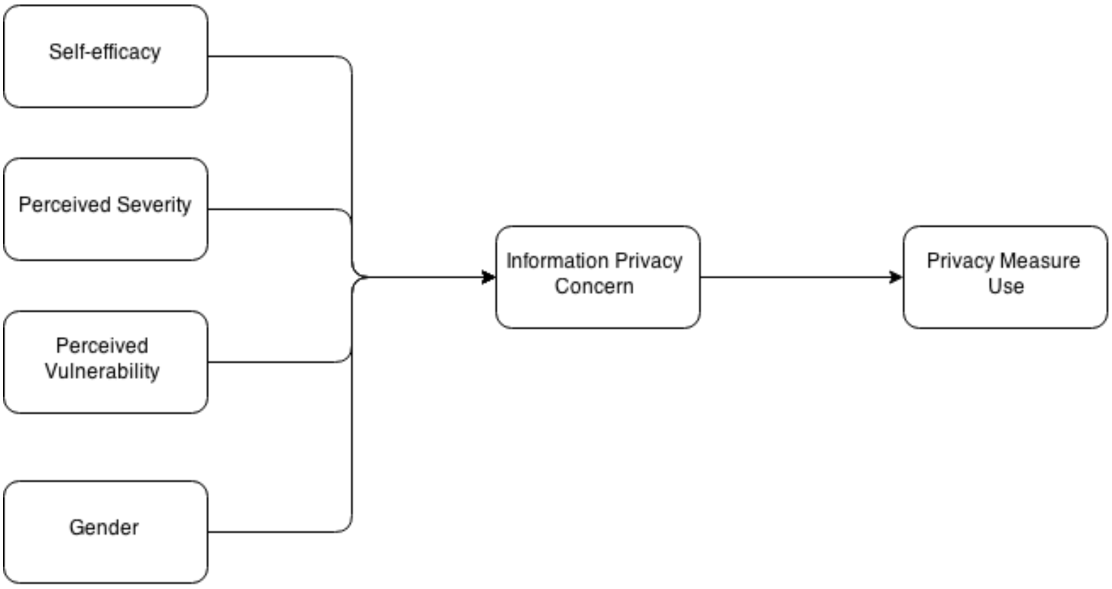


Figure 2.5: Relationships shown by Mohammed & Ahmad (2012)

Another study with similar findings was done by Joinson et al.(2006). The variables used in this study, however, are formulated differently and likely have some overlap with several of those mentioned in the above model. The authors found that in online settings, privacy concern and perceived privacy (mediated by trust) are related to non-disclosure. In a later study, these same researchers found that, while disclosure depends on both trust and perception of privacy, a high score in one of these variables can compensate for a low one in the other (Joinson et al., 2010).

The concept of trust has proven to be central in the consideration of online disclosure. This was clearly shown in a study performed by Taddei & Contena (2013). Their paper showed that while privacy concerns do not directly influence online disclosure (as was also discussed in section 2.3), trust and a sense of control over the disclosed information do play an important role.

In a study by Stutzman, Capra & Thompson (2011) a number of other factors (including trust) influencing online disclosure are researched. They found that people who are more concerned about privacy are less likely to disclose information, and are more likely to have read the social network's privacy policy. It was also found that people, who have read more of the platform's privacy policy are less likely to disclose information. Aside from these findings, they have also looked into some relationships concerning privacy settings. The authors distinguish between two actions for the privacy settings, namely privacy personalisation ("changing the default privacy settings") and customisation ("customizing which individual friends have access to content"). For both these variables, participants were asked if they had ever done them, resulting in either a "yes" or a "no". Results showed privacy customisation was positively related with both privacy attitude and disclosure. Unfortunately, the results concerning privacy personalisation were inconclusive, possibly due to the vast majority of participants having answered the same to this question.

To conclude, an overview of the relationships found by the various papers discussed in this subsection was made. Table 2.4 presents all relationships with the two main concepts under study, being disclosure behaviour and privacy measure use. In this table, indirect effects are noted between brackets following the factor they first influence, and mediating factors are noted in italics following the factor of which they affect the relationship.

Study	**Factor influencing privacy measure use**	**Factor influencing disclosure behaviour**
Christofides, Muise & Desmarais (2011)	Time spent on Facebook, duration of membership, age, gender, general trust, self-esteem, awareness of disclosure consequences	-
Mohamed & Ahmad (2012)	Information Privacy Concern (Self-efficacy, perceived severity, perceived vulnerability, gender)	-
Joinson et al. (2006)	-	Online settings, privacy concern, perceived privacy (*trust)*
Joinson et al. (2010)	-	Trust, perception of privacy
Taddei & Contena (2013)	-	Trust, sense of control over disclosed information
Stutzman, Capra & Thompson (2011)	-	General privacy concern, knowledge of privacy policy, privacy measure customization.
Zimmer et al. (2010)	-	Attitude (user's trust in the social network, perceived risk of loss of control over disclosed information, relevance of disclosed information to the function of the website)

Table 2.4: Relationships with privacy measure use and disclosure behaviour in papers discussed in this subsection

2.7 Conclusions for this study

As was discussed in the introduction to this chapter, there are two important goals that this collection of related works strives for, namely that of providing a context of what is already known, as well as finding out what factors may play a role in conducting this study.

Earlier studies have already looked into a possible correlation between disclosure behaviour and privacy measure use, but have proven to be inconclusive. As was discussed in section 2.3, this relationship was most clearly researched by Christofides, Muise & Desmarais (2009). Unfortunately, their study proved to be inconclusive, as the relationship found was not statistically significant. However, it should be noted that the information gathered for this study was done using a standard online survey, rather than using the actual data available on Facebook. This may have influenced the results somewhat, and have limited the accuracy of the analysis. As Acquist & Gross (2006) have shown, Facebook users are often unable to correctly state what information they have disclosed on the network. By gathering extensive information about the user’s disclosure behaviour and use of privacy settings, using collection tools as much as possible, this study will attempt to create a deeper understanding of the relationship between these two factors. This will allow for a more conclusive result as to whether they are related or not.

In terms of what factors may affect the outcome of the study at hand, this should be clear from the models presented in section 2.6. While there are several factors that have been found to affect use of privacy measures, the most relevant factors for our study are age and gender (Mohammed & Ahmad, 2012) (Wilson, Proudfoot & Valacich, 2014). If the distribution within the sample does not match that of the population, with regard to these factors, this could be an important threat to validity. It should also be noted that the level of education of the research participants might also bias the study. However, considering the majority of studies found for this literature study were conducted with college students, this bias would be present throughout the existing literature.

3. Research model

In this chapter, we will give a brief overview of the focus and aims of this study, including research questions, the concepts involved, and how they are suspected to be related. This will form the basis for the research method described in chapter 4.

3.1 Model

Based on the literature reviewed in chapter 2, a model was constructed that proposes the exact relationships described by the research questions. This model can be seen in Figure 3.1.

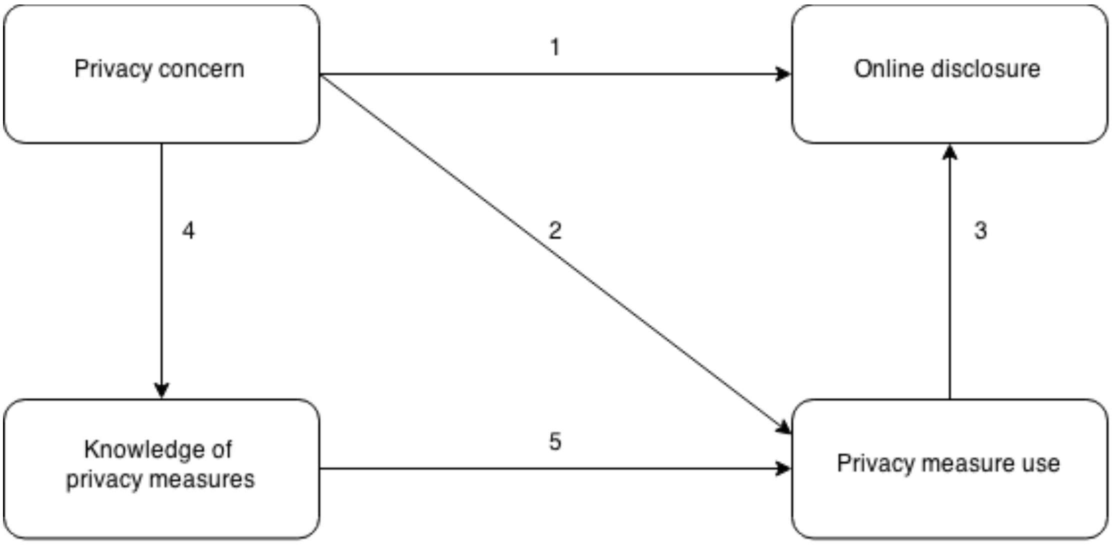


Figure 3.1: Model of concepts

The first relationship is based on the works of Joinson et al. (2006) as well as Stutzman, Capra & Thompson (2011). Both were able to confirm this relationship. However, since it is an integral part of the model, it will also be part of this study. It should be logical that people with more concerns about their privacy are likely to disclose less information. With the exception of a small number of required profile items, users are free to choose what information they share. If users are concerned about their privacy, this should play an important role in this decision. As such, we formulate the following hypothesis for this relationship:

> H1: People who are more concerned with privacy are less likely to disclose personal information online.

The second relationship has also been confirmed by earlier research, in this case by Mohamed & Ahmad (2012). As with the previous relationship, the importance of this connection to the model requires that we research this relationship, as well. Like with the previous relationship, it is to be expected that people with stronger privacy concerns will act on this. Stricter use of privacy measures allow users to limit the likelihood of unintended access to the disclosed information. Based on this, we formulate the following hypothesis.

> H2: People, who are more concerned with privacy, are more likely to use privacy measures more restrictively.

The third relationship relationship is the main object of the study. An earlier study by Stutzman, Capra & Thompson already researched this relationship, but was partially inconclusive. This was also discussed in section 2.6. While privacy measure use may be an effect modifier on relationship 1 (as it should allow users to partially mitigate concerns over disclosing information), we suspect this factor to be directly related to online disclosure. If, for instance, online disclosure is not compensated by privacy measure use, this could be because of personality or attitude towards the social media platform. This leads to the hypothesis below.

> H3: People with more restrictive use of privacy measures are less likely to disclose personal information online.

Looking at relationship 4 in the context of the literature study, we see that while it was not studied directly, a similar study was included. Research by Stutzman, Capra & Thompson (2011) showed a relationship between the user's knowledge of Facebook's privacy policy and his/her disclosure behaviour. Also, they found that the user's knowledge of this privacy policy was influenced by his/her privacy concerns. With the relationships indicated as 4 and 5, a similar relationship is proposed for the user's knowledge of privacy measures. This would be logical, as concerned users would be more likely to investigate the exact workings of the privacy-related aspects of the platform than less concerned users.

> H4: People, who are more concerned with privacy, have more knowledge of Facebook's privacy measures.

Building on this reasoning, relationship 5 proposes a possible relationship between the user's knowledge of the privacy options available on Facebook and his/her use of these measures. The knowledge of how the platform handles information, as well as how users can protect their personal information, should result in an increased, stricter use of these privacy measures. Based on this reason, we formulate the hypothesis below.

> H5: People, who know more about the platform's privacy measures, are more likely to use use privacy measures more restrictively.

3.2 Control variables

Earlier studies have found that different variables influence aspects included in the model. As such, 4 different control variables are included in this study. These are age, gender, level of education and nationality. These will be taken into account for each of the dependant variables in the model.

Age has been shown to affect the user's privacy measure use (Christofides, Muise & Desmarais, 2011). However, it should be noted that in this study, participants were divided in two groups, namely adolescent and adult. Consequently, the groups were linked to privacy measure use. As such, small differences in age may not play a large role in this model.

As can be seen in section 2.6 different relationships with the factor "gender" were found in earlier studies. Mohamed & Ahmad (2012) related it first to Privacy Concern, and then related that factor to privacy measure use. However, Christofides, Muise & Desmarais (2011) researched a direct relationship between gender and privacy measure use. In both studies, the relationship was supported by the data. Mohamed & Ahmad did note that, while women were generally more concerned about their privacy, they were less likely to act on these concerns and increase their use of the available privacy settings. Furthermore, the relationship found by Christofides, Muise & Desmarais explained a larger part of the variation when compared to the other study.

As discussed in chapter 2, earlier studies often used exclusively college students to gather information on Facebook users. Since this study will not be exclusive to students, respondents will be asked their highest level of education they have completed, as this may factor into the results.

As with level of education, it should be noted that earlier studies always focussed on research participants in a single location. Since this study will be spread on using a snowball method, location as well as cultural background may vary greatly among respondents. It was decided to ask respondents for their nationality, as this will likely most closely reflect their cultural background, which may influence their sharing behaviour.

This brings us to a total of 4 control variables, namely age, gender, level of education and nationality. These will be recorded for each of the respondents and used in the final analysis. Chapter 4 will elaborate on how these concepts were measured.

4. Research method

The study at hand will be quantitative in nature, in which hypotheses based on the research questions formulated in chapter 3, will be tested based on the data collected. It was decided to collect the data needed using an online survey. This survey was to be assisted by a tool that gathers data from the respondent's Facebook profile, if allowed. This chapter begin by explaining, for each of the concepts in the study, how they were recorded and what the resulting data type was. After this, it will discuss the survey's distribution and obstacles expected with this approach. To conclude, it will briefly discuss the statistical analysis that was planned for the data collected.

4.1 Concepts

The main relationships in this study concern knowledge about and use of Facebook's privacy measures on the one hand, and disclosed information on the other. However, several other concepts will also be examined in order to create a more complete model of the relationships. As such, the following 5 factors will be examined for each research participant:

- General demographics
- General privacy attitude
- Knowledge about Facebook's privacy measures
- Use of Facebook's privacy measures
- Disclosed information

In the following sections, each of these concepts will be elaborated upon, focussing specifically on how they will be measured. An overview of all questions asked can be found in Appendix C.

4.1.1 General demographics

The questions focussing on general demographics look at age, gender, level of education, and nationality. These factors, while not directly linked to the research question, may factor into the results. As was discussed in section 2.7, some of these characteristics have been shown to affect some of the other factors named here. Below, a short description is given for each of these factors, including a note of their possible relevance.

Factor	Level of measurement	Description
Age	Ratio	Research participants were asked to enter their age as a number. As was described in chapter 2, this factor may play a role in the data collected.
Gender	Nominal	Research participants were also asked to select their gender. As with age, this factor may also influence the outcome of this study.
Level of education	Ordinal	Research participants were asked to select their highest level of education they had completed. The options given were, in order: None, primary school, secondary/high school, trade/vocational training, college (Bachelor), college (master), Doctorate degree.
Nationality	Nominal	As can be seen in section 2.2, previous studies generally took place using participants from a single country. However, since this survey was to be spread over Facebook, there was a good chance that people from different countries will complete the survey. As such, the research respondents were asked to fill in their nationality, as this will most closely reflect their cultural affiliation. While there has not been any formal indication that this may factor into the final results, the data could prove to be useful during the analysis.

Table 4.1: General demographics factors measured

4.1.2 General privacy attitude

The general privacy attitude of the participants will be tested using the method developed by Buchanon et al. (2007). This method was used in many studies, including the study by Joinson et al. (2010), which was included in the literature study. Moreover, the authors of the paper introducing this method were able to correlate the results found with other widely used privacy concern scoring systems, such as Westin. This method was chosen, because the amount of questions and possibility of answering allows for more variation in the data, when compared to other systems, including the widely used method of Westin.

The method by Buchanon et al. (2007) uses a set of 16 questions to measure the respondent's privacy attitude. Each of these questions asks respondents to indicate how concerned they are about a specific issue. Answers are given on a 5-point scale, ranging from "not at all" to "very much". Since these questions are indicators for a reflective construct, it was possible to leave out some of these questions. A total of 8 questions were selected in order to reduce the length of the survey. The selection was made based on the factor loading, as described in the original study. The following questions were used:

- In general, how concerned are you about your privacy while you are using the internet?
- Are you concerned about online organisations not being who they claim they are?
- Are you concerned about online identity theft?
- Are you concerned about people online not being who they say they are?
- Are you concerned about people you do not know obtaining personal information about you from your online activities?
- Are you concerned that if you use your credit card to buy something on the internet your credit card number will obtained/intercepted by someone else?
- Are you concerned that if you use your credit card to buy something on the internet your card will be mischarged?
- Are you concerned that an email you send someone may be inappropriately forwarded to others?

The answers to each of these questions were given a score on a scale between 1 and 5, and were consequently used to calculate a test statistic.

In contrast with the study by Acquist & Gross (2006), this survey will not include other any additional questions. This will be further discussed in section 4.2.2.

4.1.3 Knowledge about Facebook's privacy measures

The participant's knowledge about Facebook's privacy measures will be established based on a number of factual questions concerning privacy options and settings. The value of the variable will be the total score of questions correctly answered. Before answering these questions, participants will be specifically asked not to research the correct answers, as this would affect the results.

The questions used can be seen in table 4.4. The questions were chosen so they truly test the depth of the knowledge of Facebook's privacy measures, rather than only superficial, better-known options. For each of the 8 questions used, the response options are indicated, as well as the correct answer and the source with which this answer can be verified. Based on the answers given, respondents will be given a score between 0 and 8.

In order to ensure respondents do not guess during this part of the survey, a “I don’t know”-option was included for each of the questions. If this option was chosen, it counts as a wrong answer, as it indicates that the respondent is unaware of the correct answer.

Another step that was taken was to include the correct answers to the questions at the end of the survey. Respondents were informed of this before they answered the questions. This was done to ensure that curiosity on the part of the respondent would cause them to look up the correct answer during the survey.

Question	**Options (correct underlined)**	**Source (Websites retrieved on 31 March 2015)**
On Facebook, I can set who can add me to groups.	True, False	https://www.facebook.com/help/162550990475119
On Facebook, I can change the visibility of older posts.	True, False	https://www.facebook.com/help/236898969688346
On Facebook, I can set the visibility of each of my schools/universities I have attended separately.	True, False	Appendix B
On Facebook, I can set the default group to which my future status updates are visible	True, False	https://www.facebook.com/help/325807937506242/
On Facebook, I can set who can tag me in pictures.	True, False	https://www.facebook.com/help/226296694047060
On Facebook, I can make a profile without needing to enter my date of birth.	True, False	https://www.facebook.com/help/188157731232424
On Facebook, I can change settings so only members of my family can see what music I like.	True, False	https://www.facebook.com/help/100522066706974
I can allow Facebook to detect and share my current location when posting a status update.	True, False	https://www.facebook.com/help/115298751894487

Table 4.4: Questions used to test the research participant’s knowledge of Facebook’s privacy measures.

4.1.4 Use of Facebook's privacy measures

After having answered the questions testing the participant's knowledge of Facebook's privacy measures, they will be asked to disclose their own use of them. At this stage, they may need to look up their settings on Facebook. To prevent them correcting the earlier fact-based questions, these should be entered separately and in this order.

All settings relating to privacy on Facebook are quite numerous and can be found at different places. In appendix B, all the settings were collected that were available at time of writing, together with where they can be found and the options available to the user.

As can be seen, there are several instances where the user may enter multiple entries, each with a separate setting. For some of these, the user may be asked for a specific instance, as well as their setting for others. For instance, the user may perceive showing one's current occupation differently than showing one's full employment history. For options where the user may have entered different settings and not one single entry is requested, the respondent will be asked for the setting used that allows for the largest amount of people to see the information entered.

As this section would contain 52 questions if all options were to be included, some steps were taken in order to reduce the length of this part of the survey. These were:

- Respondents were not asked for their settings for each type of like-page separately, reducing the total amount of questions by 15.
- Respondents were not asked for their settings used for previous work and education, reducing the amount of questions by 2.
- Respondents were not asked to fill in the fields on the page "Details about you", as the information entered on this page can vary greatly with respect to the amount of personal information contained. This reduced the amount of questions by 3.

Considering this large amount of questions, and the possibility that the respondent may want to manually enter what information they have disclosed, the remainder of the questions needed to be kept at a minimum to ensure respondents did not quit the survey prematurely.

The data recorded is considered to be ordinal data, considering the size of the group that can see the data. In the table of appendix B, the groups are mentioned in order from largest to smallest, with the exception of the option "custom groups". While this generally appears as the last option on Facebook, these groups are a subsection of the user's friends. Thus, this option is considered to be between "Friends" and "Only me" in terms of group size.

Our approach differs from those used in earlier studies, in that this method looks at the actual settings, as they are present on Facebook. A study by Christofides, Muise & Desmarais (2011) also researched privacy measure use, but asked respondents questions about how likely they were to change certain privacy settings, using a 7-point Likert scale. The hypothetical nature of these questions, as well as their reliance on the assumption that the current settings used did not suffice may have affected their results. Another approach is that used by Mohamed & Ahmad (2012), who asked respondents to indicate whether they had used a certain privacy measure or not, creating a binary value for each. We believe that our method is the closest representation of the actual use of privacy measures by respondents and, as such, most accurately portrays reality.

4.1.5 Disclosed information

The last factor that will be measured is the disclosed information. All information that can be disclosed on Facebook is a subsection of the settings that can be seen in appendix B. Specifically, all the settings that can be found at "User Profile, About" are options for voluntary disclosure. It should be noted that the like-pages associated with each of these groups can be hard to find for the user. For this reason, we decided to only ask respondents to fill in whether they had liked a page for the categories that are displayed on the user's profile by default. These were: films/movies, TV programmes, music, books, sports teams, athletes/sportspeople, people, restaurants, and apps and games.

The part of the data for this factor will be collected using a tool designed for this purpose. Respondents will be asked whether they want to use this tool, or would rather enter the data manually. This is done to limit the privacy concerns respondents may have about this survey. This will be discussed further in section 4.2.1.

Since these items have a large amount of overlap with those recorded for the previous factor, and considering that they are found at the same pages on Facebook, respondents should be asked whether they would like to use the tool to collect the data concerning their disclosed information before starting with the questions about their use of Facebook's privacy measures. If not, the questions should be shown together, in order to make the data collection more efficient.

The resulting list of possible disclosure items can be seen below.

- Work (current)
- Work (past)*
- Professional Skills
- University (Most recent)
- University (Older)*
- High School
- Current City
- Home Town
- Other Places Lived
- E-mail
- Phone number
- Date of Birth
- Year of Birth
- Gender
- Languages
- Interested in
- Religious Views
- Political Views

- Relationship status
- Family Members
- About You*
- Favourite Quotes*
- Favourite films
- Favourite TV Programmes
- Favourite Music (artists)
- Favourite Books
- Favourite Sports teams
- Favourite Athletes and/or Sportspeople
- Favourite people (celebrities)
- Favourite Restaurants
- Favourite Apps and Games

Items marked with an asterisk will not be included in the survey, for the same reasons as those discussed for these items in section 4.1.4. Moreover, it would be nearly impossible to judge whether the information contained in the "About You"-section was personal information, or some other text.

4.2 Survey method

The information outlined in the previous section was collected using an online survey. This survey was accompanied by a Facebook information collection tool, that recorded whether respondents had shared their last name, e-mail address and whether the account had been verified. This information was, then, used to check the information provided by respondents to ensure it was correct. The code for this tool can be found in Appendix D.

Afterwards, the relationship between the various factors was examined using statistical tests. This section will elaborate on the data collection method, as well as the obstacles that were expected beforehand, based on the approach taken.

4.2.1 Data collection

The survey was spread on Facebook. This snowball approach to data collection was chosen, since it is cheap and relatively effective. Moreover, earlier studies have made overviews of its advantages, which include the types of input available, lower costs, and potential for a higher response rate (Baltar & Brunet, 2012). While this does not ensure an entirely representative sample, it suffices for the purposes of this thesis, as limited resources are available. Moreover, considering that earlier studies have researched behaviour on Facebook by looking solely at college students, this approach can be considered to be appropriate.

One important factor that needs to be taken into account is that of selection bias. To compensate for this bias, which is inherent to the snowball method, multiple points of entry should be used. Moreover, it has been argued that an increased amount of respondents will help compensate for this kind of bias (Atkinson & Flint, 2001). Nevertheless, this issue should be kept in mind when looking into the results of the study.

The factors discussed in section 4.1 will each be measured using the method discussed. The first 4 factors will be collected in the first phase of the survey, which will be a standard format online survey.

The second phase of data collection concerns the amount and kinds of data disclosed by the participants. Participants were asked to go through their Facebook profile and indicate what information they had shared. This was combined with the part focussing on the privacy settings, as Facebook does not allow users to set an audience for items that have not been disclosed.

Initially, a tool was developed that allowed respondents to log into Facebook, after which it would record what information they had shared on their profile. Regrettably, Facebook repeatedly rejected the tool when it applied for the appropriate permissions, stating that statistical research was not allowed using their API. We strongly believe such a tool would benefit the study, as it allows for both more accurate data, as well as a reduced time needed for the survey. For this study, the tool was changed so it only focussed on whether the last name and e-mail address were disclosed, and whether the profile is verified or not (which requires the user's phone number), as these items do not require extended permissions. This information could be used to verify the information provided by the respondents.

Another issue that was taken into account was that participants may have privacy concerns about using the tool (even though this limited version was less invasive), keeping them from responding to the survey. As such, this was made optional. Before continuing with this option, however, participants were informed that, since studies have shown that a significant amount of Facebook users are unable to correctly identify what information they have shared, use of the tool is encouraged. Furthermore, respondents were asked to keep Facebook open in a separate window and manually copy the information requested, when filling out the survey.

4.2.2 Expected obstacles and limitations

Apart from the privacy concerns that research participants may have (as explained in the previous subsection), there are several other possible pitfalls for this approach. The first was that the sample may not be representative for the overall population. Considering the method of distribution, the most likely 2 factors in this regard are age and level of education. While this may be hard to mitigate, it should be kept in mind when analysing the results as it may play a role in the generalisation of the conclusions.

Another possible obstacle is obtaining only a very limited amount of responses. However, as will be discussed in section 2.2, studies that used online surveys spread through Facebook generally got a large amount of responses. Nevertheless, if this is still a problem for any reason, the survey could be spread through other means as well.

In the study by Acquist & Gross (2006), other questions about recent news topics were mixed into the survey to avoid priming the respondents when responding to questions concerning their privacy attitude. In this study, this will not be done, as the topic of the study will be clearly indicated during the distribution of the survey, in order to attract interest. It should be noted that the often-cited studies by Westin also did not include additional questions, so the risk of priming respondents is considered to be minimal.

In the same study, the authors indicated that respondents were often unable to correctly indicate what information they had disclosed on Facebook. If respondents choose to manually enter their information, this may be an issue. As such, the research participant will be warned of this fact and asked to keep their Facebook profile open in a separate window so they can check whether they have entered their information correctly. Considering that the survey will be spread using Facebook, this should not prove too problematic.

One last possible threat to the study is that, when users are asked to enter their settings, they may change these settings as they feel they are no longer appropriate. For this study, this is considered to be a problem of instrumentation, rather than a (possibly welcome) chance for users to review whether they still agree with their privacy settings. As such, users will be asked to enter their original setting, rather than the new one.

As was mentioned before, a limitation of this study is the fact that the sample resulting from the snowball data collection method may not accurately represent the population. This should be taken into account when analysing the results.

4.3. Statistical analysis

Based on the model shown in section 3.1, linear regression analyses will be performed on the various concepts. The factors will, then, be evaluated on their predictive capability. All assumptions made for performing such an analysis will also be tested in order for us to be able to draw conclusions from the statistical regression analysis. This can be done based on the data set acquired in this study and will give a clear indication of how strong the relationship is between the various factors. During this evaluation, other factors such as age and gender may be accounted for, as these have been shown to be related to some of these factors (Christofides, Muise & Desmarais, 2011) (Mohamed & Ahmad, 2012).

5. Results

In this chapter, we will discuss the results collected and the results of the statistical analysis, based on the model presented in chapter 3. Firstly, a brief summary of the data collection process will be given to compliment the method described in chapter 4, along with an examination of several responses that were excluded for various reasons. Secondly, a description of how the concepts were calculated from the questions asked will be given. Thirdly, the sample will be described in general terms, examining several aspects and demographics of respondents. Finally, the results of the statistical tests conducted will be presented and examined. This will be limited to the interpretation of the data collected. Further analysis about the relationships and their implications will be discussed in chapter 6.

5.1 Data collection

The survey was distributed through various channels. Since respondents needed to be Facebook users, posts on the social network were the primary method of distribution. The communities in which the survey was shared predominantly consisted of students and former students, as these were easiest to reach. Nevertheless, the survey was also spread over the AISWorld mailing list, to allow for a more varied sample. Using various starting points for the snowball data collection method is one of the ways by which a possible selection bias was to be avoided, or at least minimized. As was discussed in section 4.2.2, this can be a shortcoming of the chosen method of distribution.

5.2 Excluded entries

Before the data collected was analysed, responses received needed to be checked on validity. This was done based on several aspects, which will be discussed in this section.

The first check conducted was whether there were any duplicate entries in the database of responses. Two completely identical entries were found and, even though they had been submitted almost 24 hours apart, it was decided to remove one of these responses. The reason for the delayed second response remains unclear.

The second check looked at whether the information collected by the tool (if the respondent had chosen to do so) was not disputed by information provided in the survey. This was done to ensure respondents had actually logged in on Facebook and copied the items requested, as was explained in section 4.2. The primary item used for this analysis was the user’s e-mail address. Based on this, 8 responses were left out of the analysis, as they the information provided and that found by the tool did not concur.

The last 2 items that were used to check the validity of the entries were the “last name” and the “verified” fields. For the “last name”-field, we expected all respondents to have entered this field on Facebook, with the exception of profiles from Indonesia. This is the only country, where entering one’s last name is not required when creating a Facebook profile (Katwal, 2014). The tool showed that all respondents had entered their last name on Facebook. For this reason, none of the entries gathered were excluded from the analysis as a result of this check.

The field showing whether the user profile had been verified was an issue, however. The tool concluded that all user-profiles found had been verified. However, this is unlikely, as verification requires users to add their phone number to their Facebook profile. While it is possible that all respondents had, at some point, verified their account in this way, this seems implausible. A more likely explanation of this result is that users may have entered their phone number in Facebook’s settings, without allowing it to be on their profile. Alternatively, they may have entered their phone number at an earlier stage, but removed it later on. Another possible explanation is that the tool developed did not function with Facebook as it did during the testing phase. Since we cannot be certain about the exact circumstances, we have decided not to exclude entries based on this criterion. The use of this value was already a compromise for not being allowed to access the actual field showing the user’s phone number, but appears to have been a poor substitute.

One final item that may be relevant when considering responses to exclude is that of outliers. This will be discussed in section 5.6.

5.3 Concepts

In chapter 4, we discussed the questions used in the survey to measure the various concepts. In this section, we build on this by discussing how the answers were used to calculate variables to represent the concepts, as well as checking important assumptions made for these variables.

5.3.1 Transformation into variables

Before the answers given to the questions in the survey could be used for analysis, they needed to be transformed into single variables. These variables, in turn, would be used for the linear regression tests, which will be discussed in section 5.5. The following paragraphs will explain how the answers given were used to calculate the variables. The questions posed can be found in chapter 4.

The concept of **privacy concern** was measured using 8 questions on a 5-point scale. Each of these answers was given a numerical value, ranging from 1 to 5. Next, factor reduction was used to calculate the value of this latent variable. Before this was done, however, Cronbach's Alpha was calculated to assess the internal consistency. The value found was 0.853, showing an appropriate level of consistency. It should be noted, however, that some respondents remarked that one of the questions asked concerned a situation involving a credit card, while they themselves did not have one. This could threaten the consistency of the answers, since they might respond that they had no privacy concerns about this, at all. Nevertheless, the value for Cronbach's Alpha shows that the answers combined show a proper level of scale reliability (Field, 2009). For this reason, it appears the variable can be calculated without the need to remove the answers to this question, so this threat appears to be very limited.

The variable representing **knowledge of privacy measures** was calculated based on factual questions asked. For each question, the respondent got 1 point if it was answered correctly, and 0 points if the answer was incorrect. Questions where the respondent indicated they did not know the answer counted as incorrectly answered questions, as the concept measures knowledge of Facebook's privacy measures, not knowledge of the respondent's own knowledge. The total amount of points scored with these questions was used as the value for knowledge of privacy measures. Since there were 8 questions, the score could vary between 0 and 8.

In order to calculate a score for **privacy measure use**, each of the settings submitted in the survey were converted to scores using 4-point scale. This scale was based on the most frequent options appearing in Facebook's privacy settings, namely Public, Friends of friends, Friends, and Only me. These options were given a score ranging from 1 to 4. The options as found on Facebook can be seen in Figure 5.1. Each of the settings respondents filled in on the survey was scored based on this scale. Settings that did not have all these options available still used the same mapping as those, which did. For all settings that did not refer to this list of publics, the strictest settings were given a score of 1, and the least restrictive setting was given a score of 4. Another important point to note is that in cases that respondents had opted for a custom audience, this was not included in the overall score. The reason for this is that the custom audience can have almost any size, making it impossible to judge how strict the actual setting was.

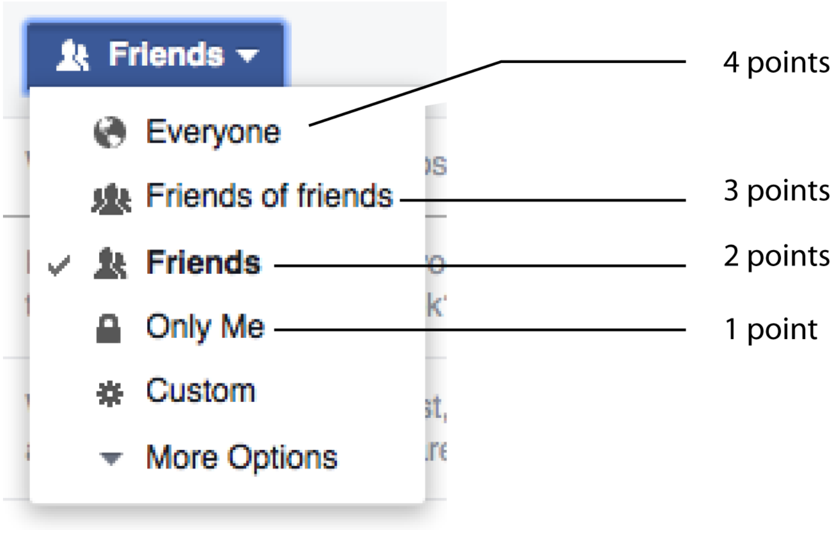


Figure 5.1: Example of options available for privacy settings on Facebook and corresponding scores for privacy measure use.

The maximum possible total amount of settings submitted for this factor was 27, depending on how much the respondent had actually disclosed. Once all the values were converted to the 4-point scale, an average was calculated. This average was used as the variable for privacy measure use.

After all settings had received their settings, the variable representing use of privacy measures was calculated based as the average setting the respondent had. This was done because Facebook does not allow users to set a disclosure audience for items that had not been filled in. As such, a score that would look at, for example, the total of all the values assigned would result in people with very few items disclosed to have very low scores, regardless of the actual use of privacy measures. The items that were not disclosed would all add a score of 0 to the total. Another consideration here was that the average score of the scale, being between that for Friends of friends and Friends, was representative for the default privacy settings for Facebook. For a long time, Public or Friends of friends was the default audience for items on a user's profile, but this was changed in 2013 and 2014 to be Friends (Yeung, 2013; Magid, 2014). This further justifies using the 4-point scale to calculate an average score.

The last concept remaining is **online disclosure**. This score was calculated by first determining the number of items the respondents could share. For instance, if the respondent was a student, who had not had a job, he/she cannot disclose anything what kind of work they do. Respondents could indicate that this was the case by choosing the "does not apply"-option in the survey. After the total number of disclosure items was calculated, the score was calculated based on the number of items the respondent had disclosed as a percentage of the number of items the respondent could disclose.

An overview of the resulting data set, including levels of measurement and the range of values, can be seen in Table 5.1. This table also mentions that for 2 of the factors, the ^{10}log score was used. This was done due to the fact that the data collected was not normally distributed, which will be further explained in section 5.3.2.

Factor	**Level of measurement**	**Original range of values**	**Remarks**
General Demographics	Various	-	Age, gender, level of education and country of residence.
General Privacy attitude	Interval	0-5	
Knowledge about Facebook's privacy measures	Interval	0-8	
Use of Facebook's privacy measures	Interval	1-4	^{10}log score was used.
Disclosed Information	Ratio	0%-100%	^{10}log score was used.

Table 5.1: Concepts in data set after transformations

5.3.2 Assumptions of the variables

For the analysis of the various relationships, a number of assumptions are made. One assumption relates specifically to the variables separately, rather than taken together as part of the statistical analysis, namely that they are normally distributed (Field, 2009). This assumption was tested for each of the variables using the Kolmogorov-Smirnov and the Shapiro-Wilk tests.

Unfortunately, both use of privacy measures and online disclosure were not (sufficiently) normally distributed. To remedy this, the ^{10}log score was calculated for each of the variables of these concepts (Field, 2009). The test was repeated for the new scores, revealing that the transformation had resolved the issue for privacy measure use. Disclosure of personal information was determined to be normally distributed according to the Shapiro-Wilk test (with a significance of 0.020), but not by the Kolmogorov-Smirnov test. We will need to keep this in mind for the remainder of the analysis.

In figures 5.2 up until 5.5, the distribution of privacy measure use and online disclosure can be seen both before and after the additional transformation.

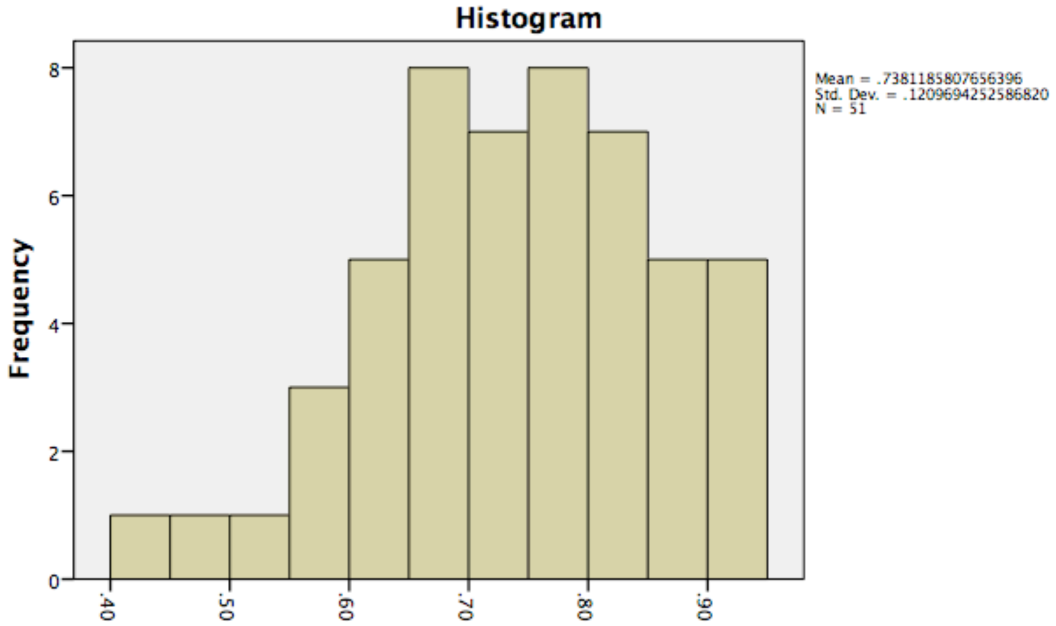


Figure 5.2: Distribution of online disclosure values before ^{10}log transformation

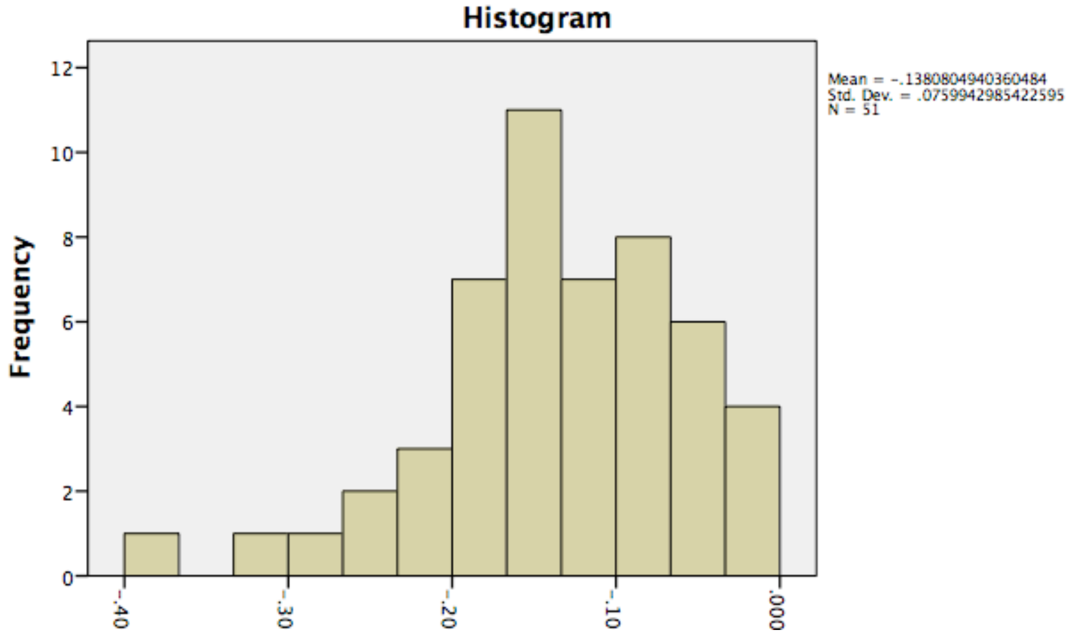


Figure 5.3: Distribution of online disclosure values after ^{10}log transformation

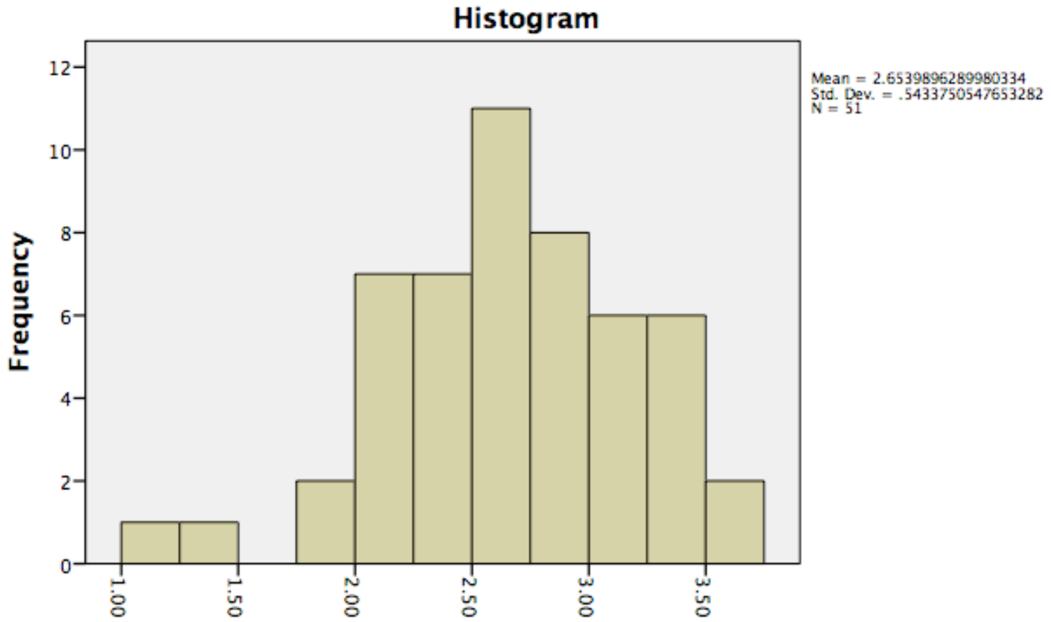


Figure 5.4: Distribution of privacy measure use values before ^{10}log transformation

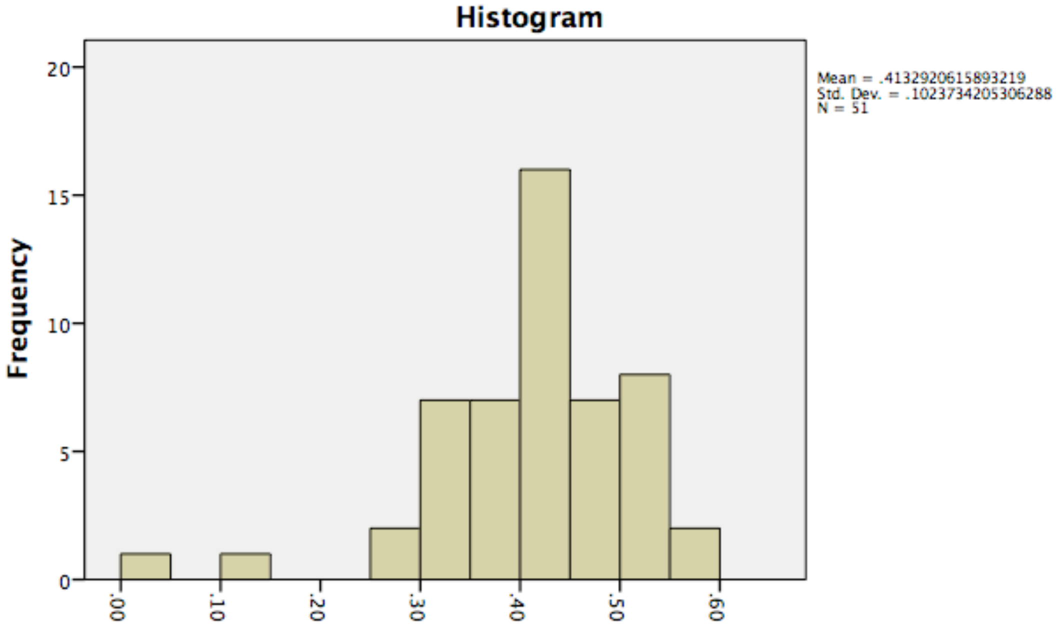


Figure 5.5: Distribution of privacy measure use values after ^{10}log transformation

5.4 General information on the sample collected

In this section, we will look at general demographics of the sample collected, as well as some other aspects that may be relevant for building a proper understanding of the data. In total, 60 responses were collected, of which 9 were removed from the sample, as explained in the previous section.

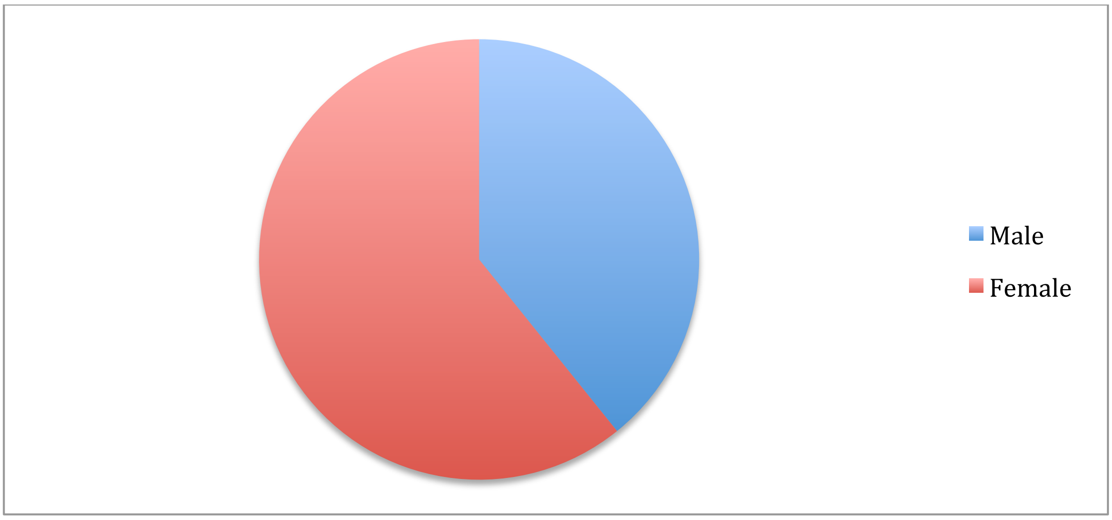


Figure 5.6: Gender division in the sample

As can be seen in Figure 5.6, the majority of respondents were female. Nevertheless, there is still a large group of male Facebook users who took part in the study, allowing us to use gender as a control variable as proposed in section 4.1.1.

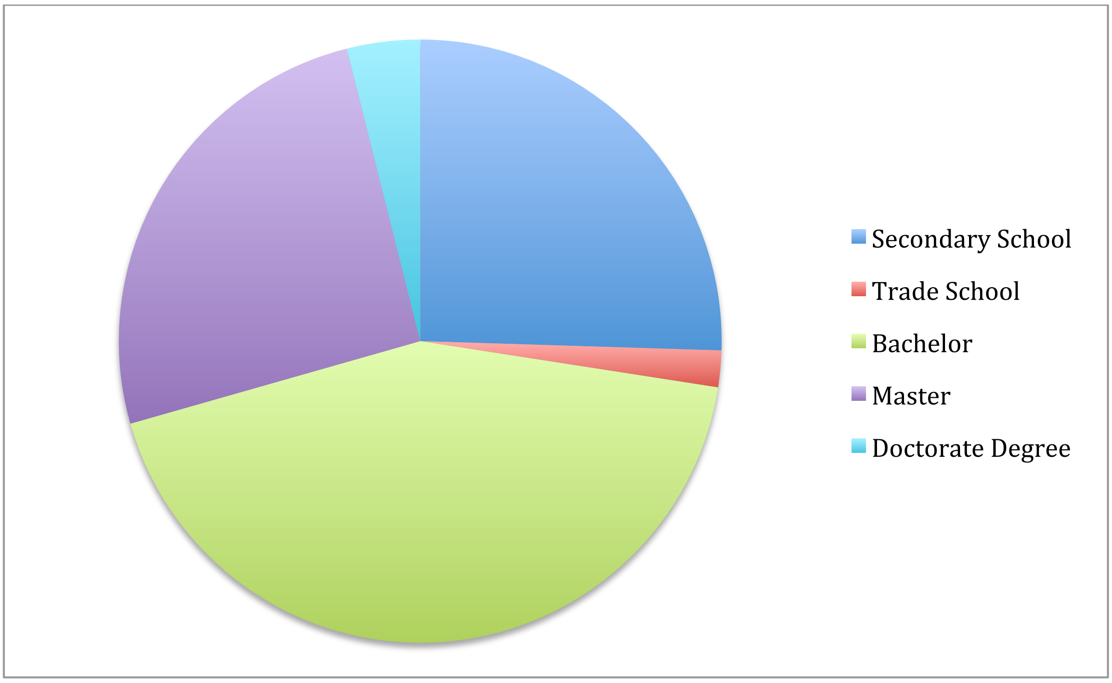


Figure 5.7: Highest level of education completed among respondents

In Figure 5.7, it easily becomes clear that sample collected was generally well educated, with the largest group of respondents having completed their Bachelor degree at university. Moreover, the majority of respondents had received at least their Bachelor degree. Obviously, this is not representative of the general population. As such, this should be taken into account when analysing the results. However, it should be noted that many earlier studies focussed exclusively on university students and may have had a similar bias, as was discussed in section 2.2.

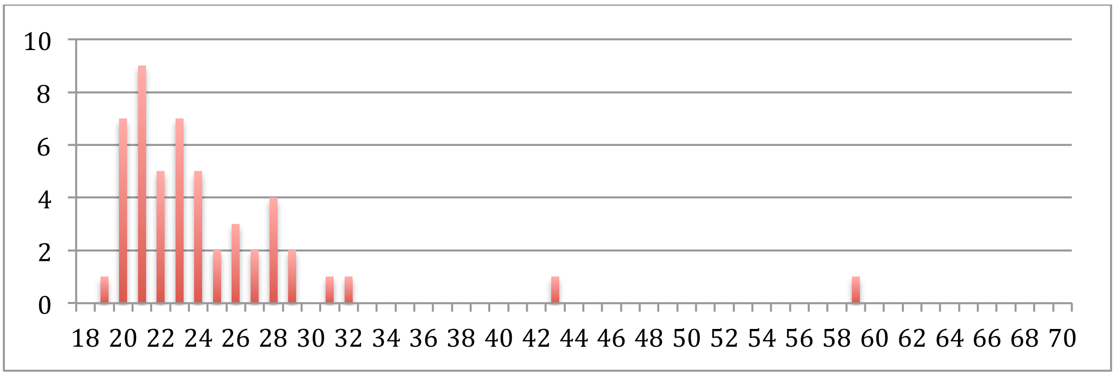


Figure 5.8: Histogram of the age of respondents

Figure 5.8 shows the distribution of ages among respondents. As was to be expected, most of the people who participated in the study are quite young. This is likely the result of both the distribution method used, as many of the communities used as starting point for the data collection consisted mostly of students.

As can be seen in the graph, there are some outliers in terms of age. However, the analyses that will be discussed in section 5.6, dealing with outliers in the various concepts, showed that these entries were not outliers in any other respect. For this reason, they were not removed from the sample.

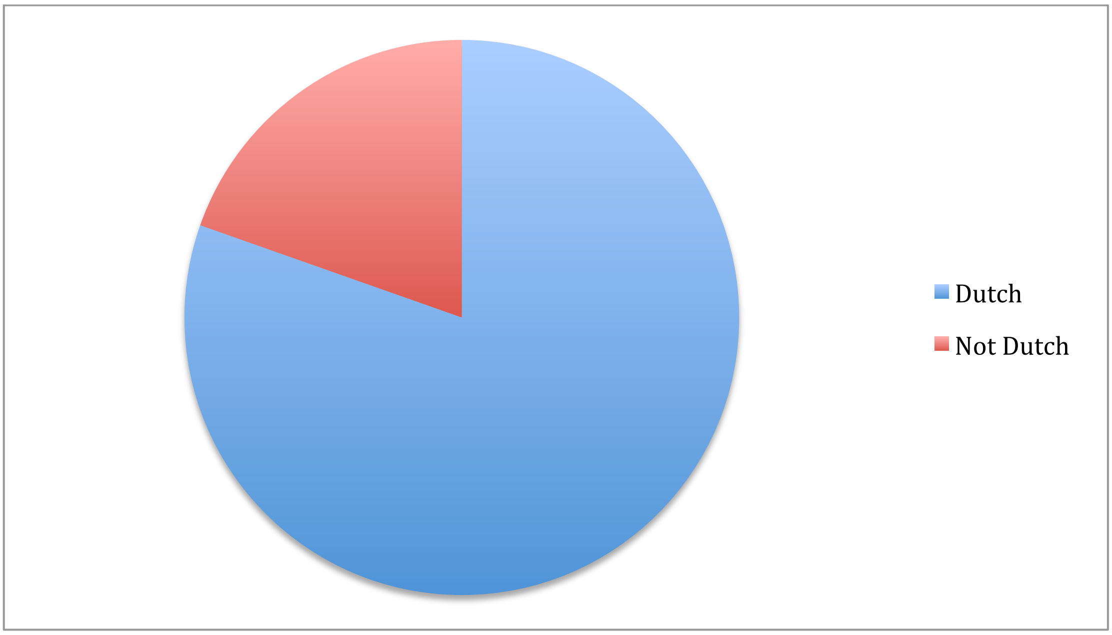


Figure 5.9: Nationality of respondents

The vast majority of respondents had the Dutch nationality, as can be seen in Figure 5.9. In fact, of all other nationalities included in the final sample, there was no other nationality with more than a single entry. For this reason, any analysis that indicated that a certain nationality (except for Dutch) affected the scores was not considered to be a relevant outcome, as this would give a large influence of these separate entries over the overall dataset.

Since nationality was measured as nominal values, they needed to be coded into so-called dummy variables in order for them to be used in the analysis (Field, 2009). This means that, for each nationality in the sample, a variable was added. The value of this variable was 1 in the respondent belonged to this nationality, and 0 if he/she did not. Considering the distribution of respondents in the final sample, only the variable "Dutch" will be considered to be relevant. For the sake of completeness, all nationalities will be included in the initial analysis below, as they may hint at a possible relationship. However, no regression models including a nationality variable (other than the one indicating whether a person is Dutch) will be accepted. If necessary, the analsysis will need to be done a second time, without including these variables. This was the case with relationship 2 (see section 5.5.2).

Aside from Dutch, nationalities that were included in the sample were Albanian, American, French, German, Greek, Polish, Romanian, Russian, Spanish, and Turkish.

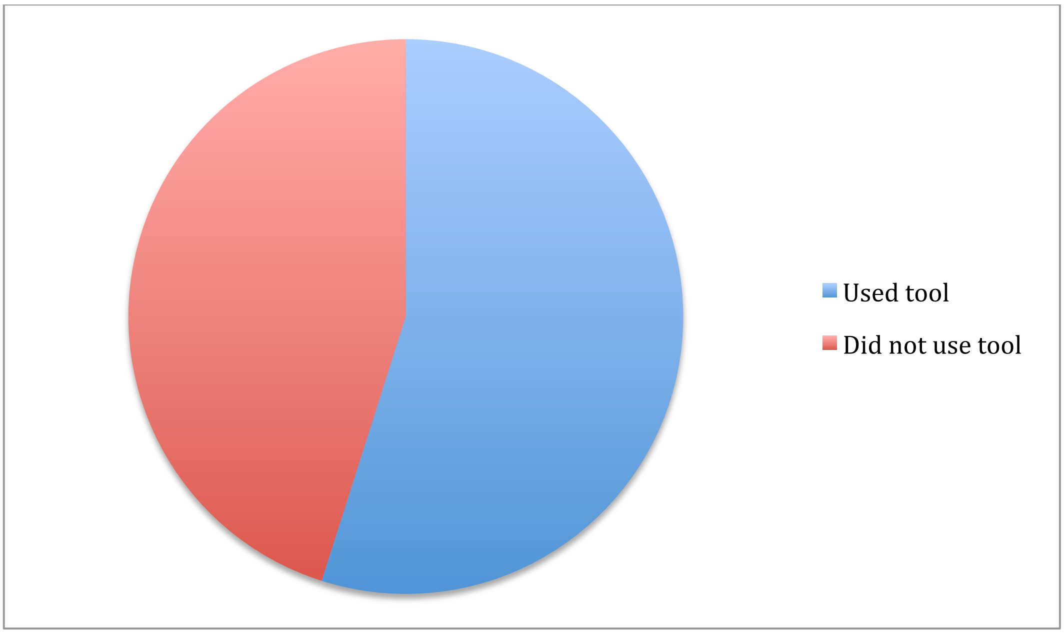


Figure 5.10: Division in tool usage

Since there were two versions of the survey available to respondents, examining the difference between the two groups created by this division may also yield some interesting findings. Before starting the survey, respondents were given the choice whether they wished to use the data collection tool, which would allow us to verify part of the information submitted. As can be seen in Figure 5.10, the majority of respondents did use the tool-assisted version. Nevertheless, a substantial amount of respondents chose not to allow access to their Facebook profile. It should be noted that this is the division after 8 of the tool-assisted entries were removed.

Apart from demonstrating a certain desire for more privacy when responding to the survey, the two groups created by this choice may also hold some other insights. This analysis will be done in section 5.7.

5.5 Relationships

In this section, we will examine each of the relationships proposed in the model presented in chapter 3. These relationships will be tested using linear regression. Furthermore, the assumptions made in this analysis will also be checked using various tests.

According to Andy Field (2009), there are 9 assumptions that are made for linear regression analysis. A number of these are valid for all analyses; others need to be addressed with each test.

The first of the group of general assumptions mentioned is that of variable types. As was described earlier in section 5.3.1, not all variables were in the required format. As such, the methods described in section 5.3 were used to ensure all variables met this requirement.

The second assumption is that of independence. Andy Field (2009) describes this assumption as “each value of the outcome variable comes from a separate entity”. With the exception of the duplicate entry that was found and removed from the data set, we can be fairly certain that each set of values collected originated from a different correspondent. However, since the survey was spread online, rather than taken in a controlled environment, we cannot be absolutely sure.

The last 2 assumptions that apply to all relationships are those of (log)linearity and non-correlation with external variables. Unfortunately, these cannot be tested and must simply be taken into consideration when looking at the results. Especially with regard to the correlation with external variables, we cannot exclude the possibility of other variables influencing behaviour. This is, however, true for all studies. Nevertheless, relationships found in this analysis should prove a good starting point for future research into the possibility of other factors that influence behaviour.

The remaining 5 assumptions can be seen in Table 5.2, along with the test used to assess the assumption and, if applicable, the desired numerical outcome of these tests. For each of the relationships examined, these tests will be conducted.

In the next sections, we will discuss the findings for each of the relationships in the model. Each of these relationships will be shown as it was tested, including the control variables that were taken into account. In the corresponding table of test results, all factors that were found to be significant in the regression analysis can be seen. All factors shown in the model, but not mentioned in the test output were included in the test, but were found to be not significantly related ($p > 0.05$).

Assumption	Test used	Desired numerical value
Non-zero variance	Regression (B coefficient)	Not 0
No perfect multicollinearity	Collinearity statistics (VIF and tolerance)	VIF: larger or equal to 1 Tolerance: larger than 0.2
Independent errors	Durbin-Watson	Close to 2
Homoscedasticity	Plot of residuals	-
Normally distributed errors	Histogram of residuals	-

Table 5.2: Assumptions to be tested for each relationship found.

5.5.1 Relationship 1: Privacy concern and online disclosure

The first relationship is that between privacy concern and online disclosure and can be seen in Figure 5.11. In Figure 5.12, part of the SPSS output of the linear regression test can be seen. The results show 3 possible models, the first of which uses only privacy concern to explain variation in online disclosure, while the 2 and 3 add the factors German and American. Since the sample included only a single entry from both the German and the American nationalities, model 1 is considered to be most acceptable as a result. This model shows a significance of 0.001 for the factor of privacy concern. Therefore, this relationship is considered to be confirmed.

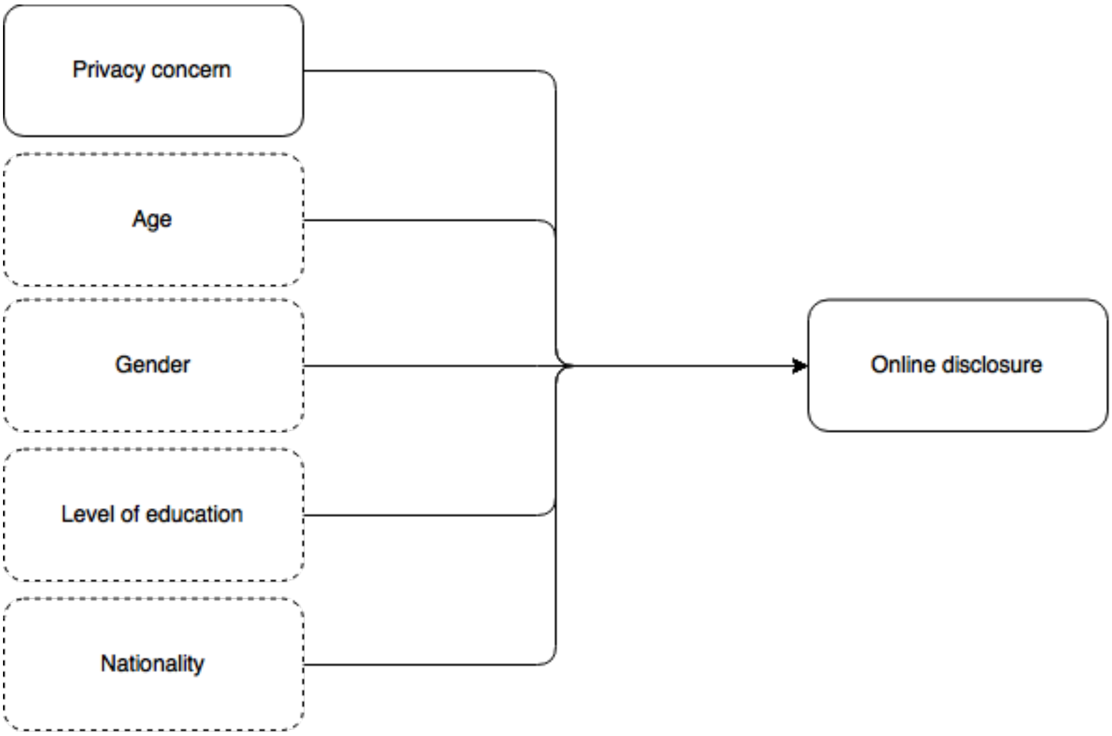
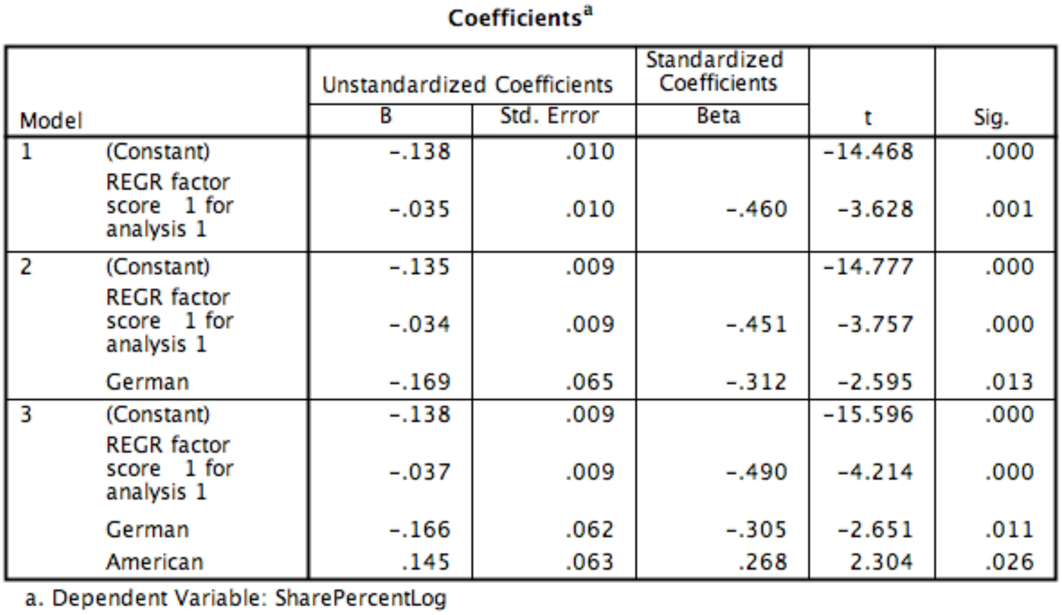


Figure 5.11: Concepts involved in relationship 1.



Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	-.138	.010		-14.468	.000
	REGR factor score 1 for analysis 1	-.035	.010	-.460	-3.628	.001
2	(Constant)	-.135	.009		-14.777	.000
	REGR factor score 1 for analysis 1	-.034	.009	-.451	-3.757	.000
	German	-.169	.065	-.312	-2.595	.013
3	(Constant)	-.138	.009		-15.596	.000
	REGR factor score 1 for analysis 1	-.037	.009	-.490	-4.214	.000
	German	-.166	.062	-.305	-2.651	.011
	American	.145	.063	.268	2.304	.026

a. Dependent Variable: SharePercentLog

Figure 5.12: SPSS output of coefficients of the linear regression analysis.

The R squared for model 1 was calculated to be 0.212. Therefore, privacy concern explains 21.2% of the variation in online disclosure. The relationship found shows that people who are more concerned with their online privacy generally disclose less information.

The result of the first 3 assumption tests can be seen in Table 5.3. As is clear from the comparison to the values mentioned in Table 5.1, all of these test statistics reflect that the assumptions were met.

Assumption	**Test used**	**Value for relationship 1**
Non-zero variance	Regression (B coefficient)	-0.035
No perfect multicollinearity	Collinearity statistics (VIF and tolerance)	VIF: 1.000 Tolerance: 1.000
Independent errors	Durbin-Watson	2.102

Table 5.3: Results of assumptions tested for relationship 1.

The last 2 relationships that were tested were homoscedasticity and the normal distribution of errors. The plots used for these tests can be seen in Figures 5.13 and 5.14, respectively. Both of these plots indicate that the assumption was not violated. Therefore, all assumptions made related to relationship one have been checked and found to be correct.

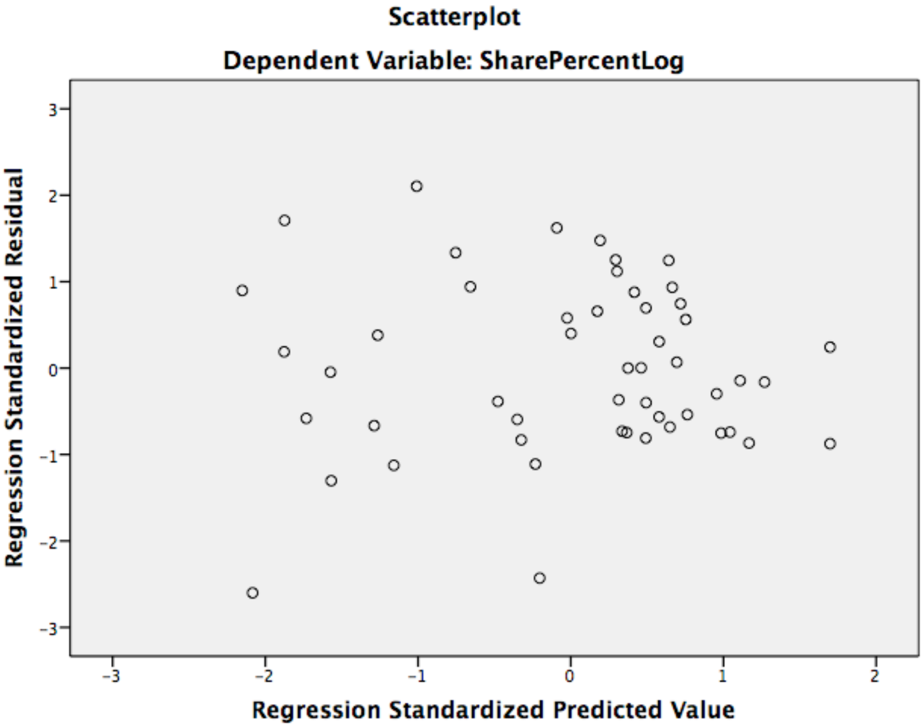


Figure 5.13: Plot of predicted and actual residual values for relationship 1.

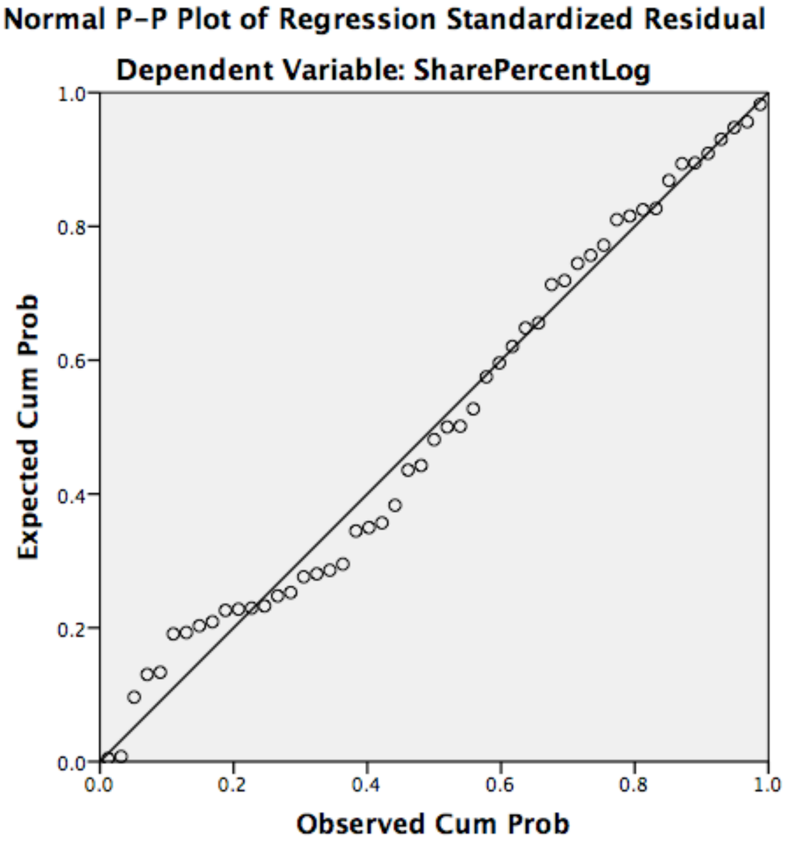


Figure 5.14: Normal P-P plot of predicted and actual residual values for relationship 1.

5.5.2 Relationship 2: Privacy concern and use of privacy measures

The second relationship in the model links privacy concern with the use of privacy measures. The initial SPSS analysis can be seen in Figure 5.16. As with the last relationship, we prefer models that do not include nationalities with only one entry. This is the case with French. As such, another analysis was performed that excluded these nationalities, so that a model could be constructed that was acceptable according to this criterion. The output of coefficients for this analysis can be seen in Figure 5.13.

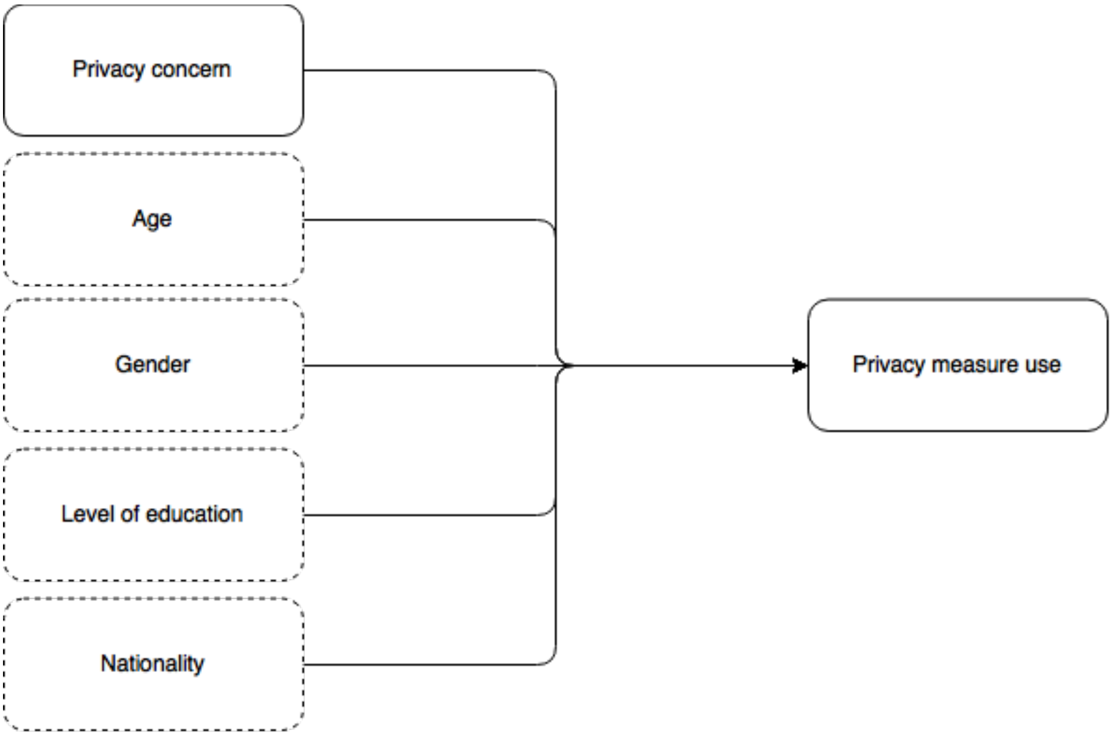
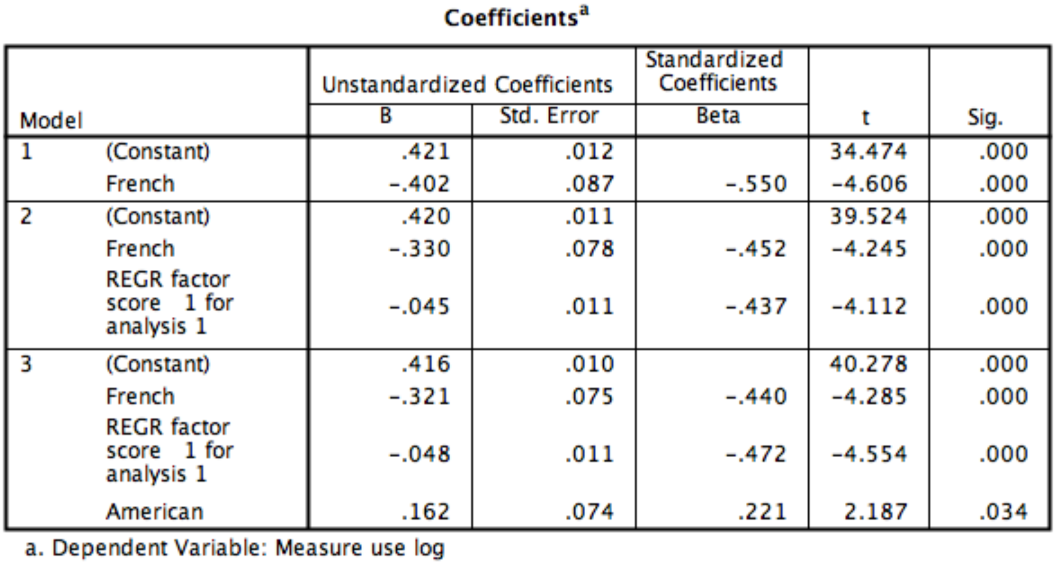


Figure 5.15: Concepts involved in relationship 2.

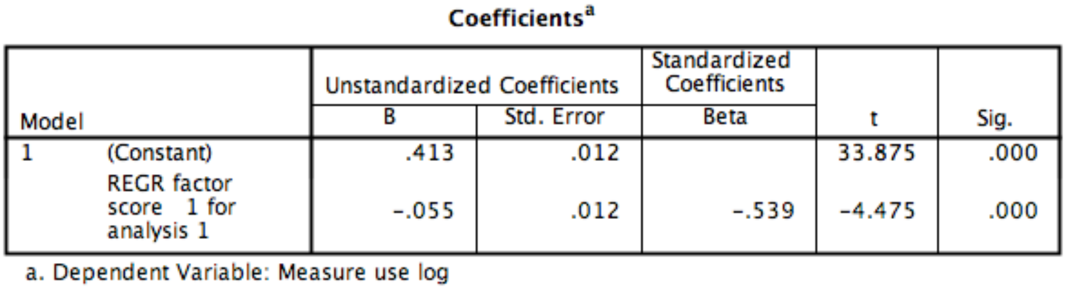


Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.421	.012		34.474	.000
	French	-.402	.087	-.550	-4.606	.000
2	(Constant)	.420	.011		39.524	.000
	French	-.330	.078	-.452	-4.245	.000
	REGR factor score 1 for analysis 1	-.045	.011	-.437	-4.112	.000
3	(Constant)	.416	.010		40.278	.000
	French	-.321	.075	-.440	-4.285	.000
	REGR factor score 1 for analysis 1	-.048	.011	-.472	-4.554	.000
	American	.162	.074	.221	2.187	.034

a. Dependent Variable: Measure use log

Figure 5.16: SPSS output of coefficients of the initial linear regression analysis.



Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.413	.012		33.875	.000
	REGR factor score 1 for analysis 1	-.055	.012	-.539	-4.475	.000

a. Dependent Variable: Measure use log

Figure 5.17: SPSS output of coefficients of the second linear regression analysis.

As can be seen in Figure 5.17, the relationship was found to be significant with a confidence level of 99.9%. The calculated R squared for this model was 0.290, meaning that privacy concern explains 29% of the variation in privacy measure use. The relationship found states that people who are more concerned with their online privacy are more likely to have strict privacy settings.

Like with the previous relationship, the result of the first 3 assumptions of assumptions can be seen in Table 5.4. All of these values are in line with what was expected.

Assumption	**Test used**	**Value for relationship 2**
Non-zero variance	Regression (B coefficient)	-0.055
No perfect multicollinearity	Collinearity statistics (VIF and tolerance)	VIF: 1.000 Tolerance: 1.000
Independent errors	Durbin-Watson	2.044

Table 5.4: Results of assumptions tested for relationship 2.

The plots for the last 2 tests can be seen in Figures 5.18 and 5.19. The plots appear to be in order, although the deviations from the line in the second figure are somewhat striking. Nevertheless, it is does not deviate enough to cause concern. For this relationship, we also conclude that it conforms to all assumptions made.

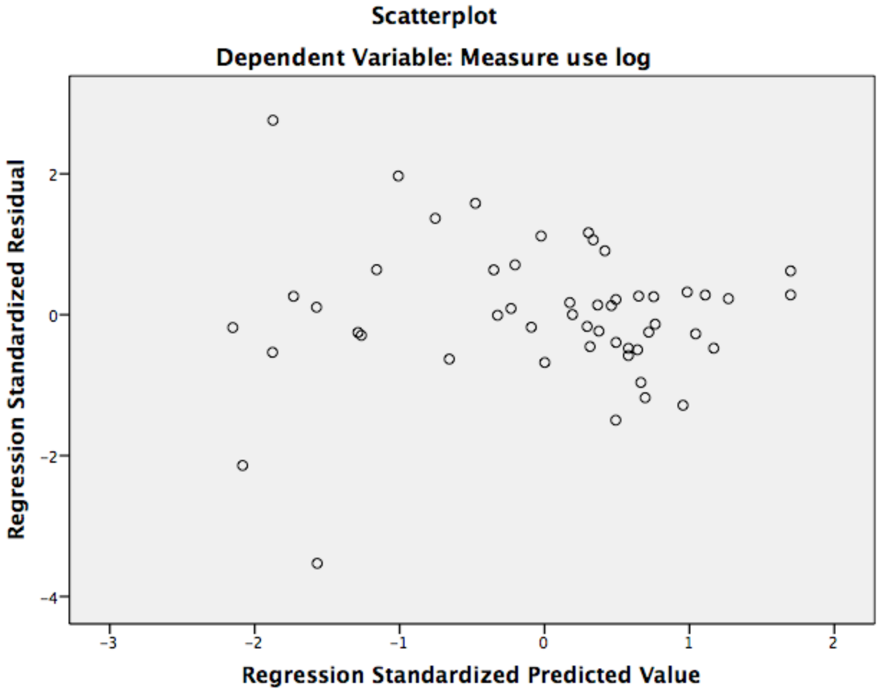


Figure 5.18: Plot of predicted and actual residual values for relationship 2.

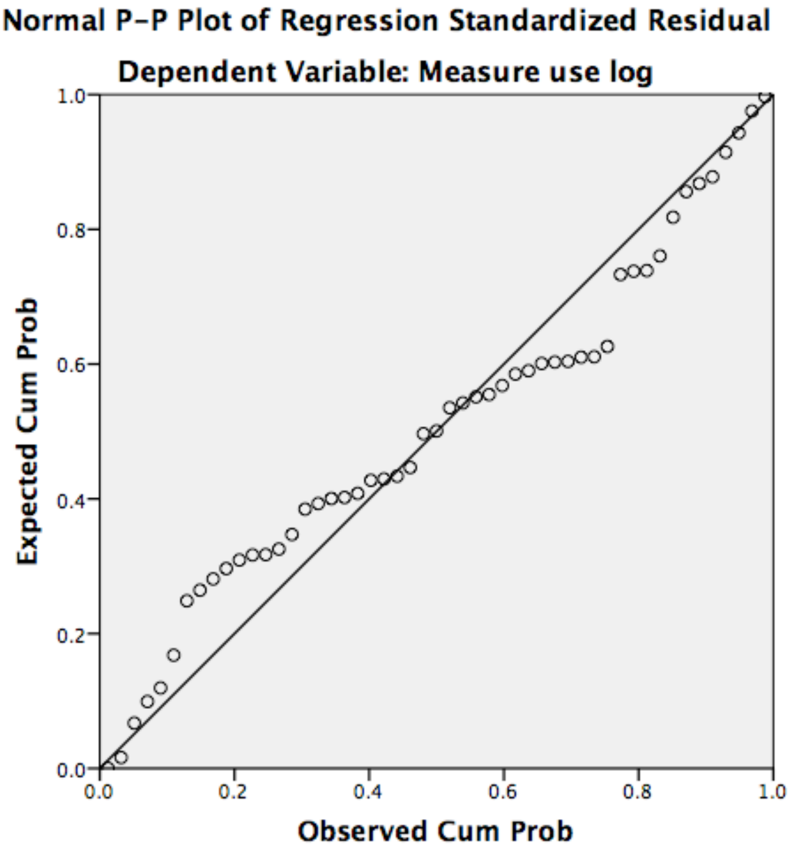


Figure 5.19: Normal P-P plot of predicted and actual residual values for relationship 2.

5.5.3 Relationship 3: Use of privacy measures and online disclosure

The third relationship, which was the primary focus of this study, concerns the relationship between the use of privacy measures and online disclosure. As can be seen in Figure 5.21, this relationship was found to be significant, with a confidence level of 99.9%. For the same reasons as the previous 2 relationships, model 1 was chosen over model 2. The R squared for this model was 0.272. Thus, privacy measure use explains 27.2% of the variation in online disclosure. Looking at the value of the B-coefficient, the relationship found states that people with stricter use of privacy settings are likely to disclose less information online.

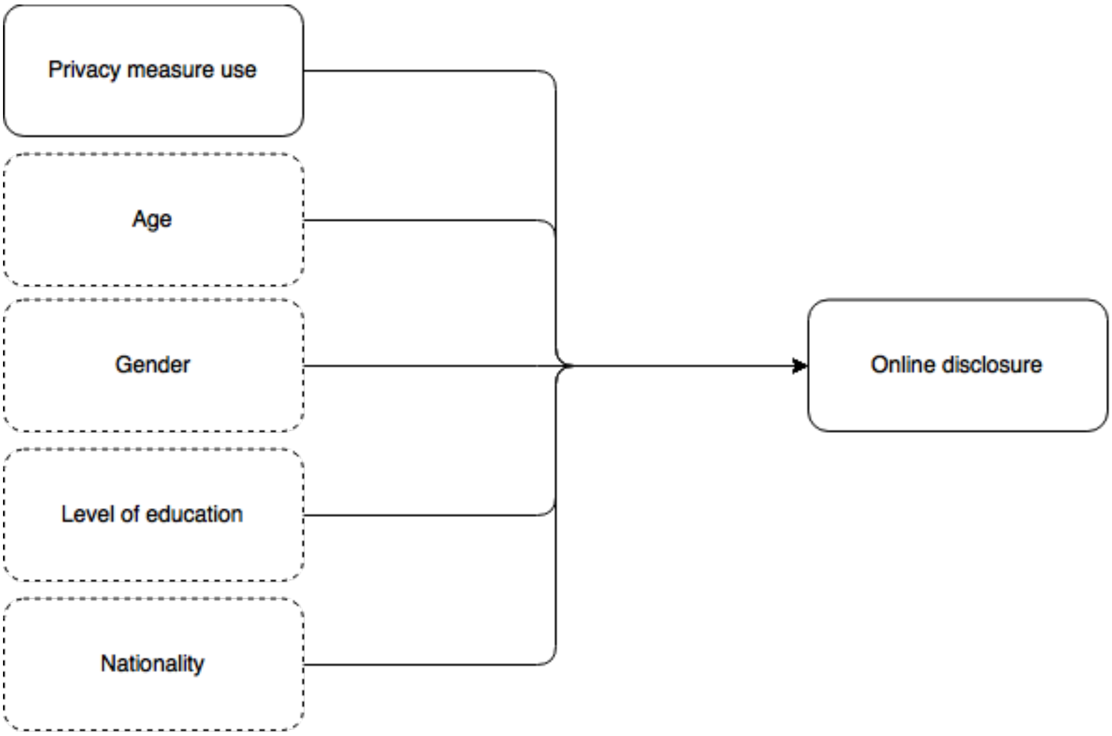
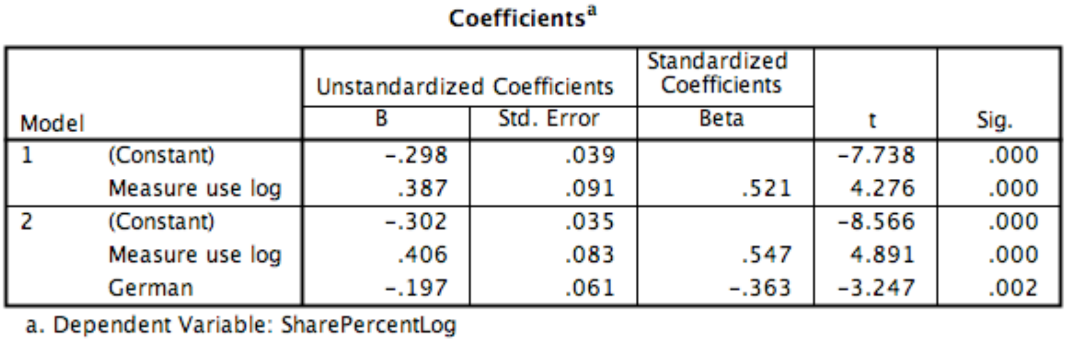


Figure 5.20: Concepts involved in relationship 3.



Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-.298	.039		-7.738	.000
	Measure use log	.387	.091	.521	4.276	.000
2	(Constant)	-.302	.035		-8.566	.000
	Measure use log	.406	.083	.547	4.891	.000
	German	-.197	.061	-.363	-3.247	.002

a. Dependent Variable: SharePercentLog

Figure 5.21: SPSS output of coefficients of the linear regression analysis.

The assumptions concerning non-zero variance, multicollinearity and independent errors were tested, the results of which can be seen in Table 5.5. As should be clear, all of these assumptions were met.

Assumption	**Test used**	**Value for relationship 3**
Non-zero variance	Regression (B coefficient)	0.387
No perfect multicollinearity	Collinearity statistics (VIF and tolerance)	VIF: 1.000 Tolerance: 1.000
Independent errors	Durbin-Watson	2.118

Table 5.5: Results of assumptions tested for relationship 3.

Also the last 2 assumptions were found not to have been violated. This can be seen from the plots in Figures 5.22 and 5.23. As a result, we can conclude that this relationship did not violate any of the assumptions.

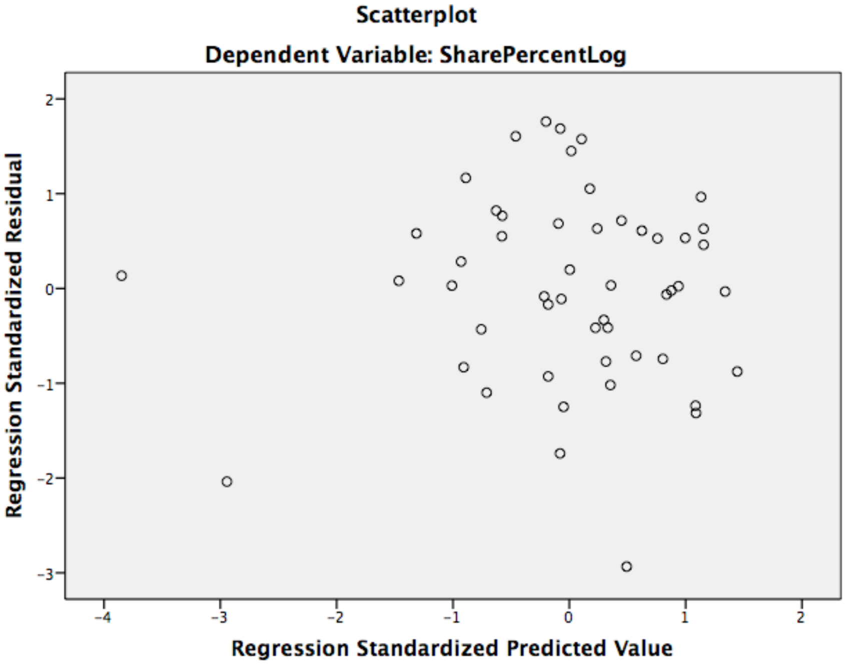


Figure 5.22: Plot of predicted and actual residual values for relationship 3.

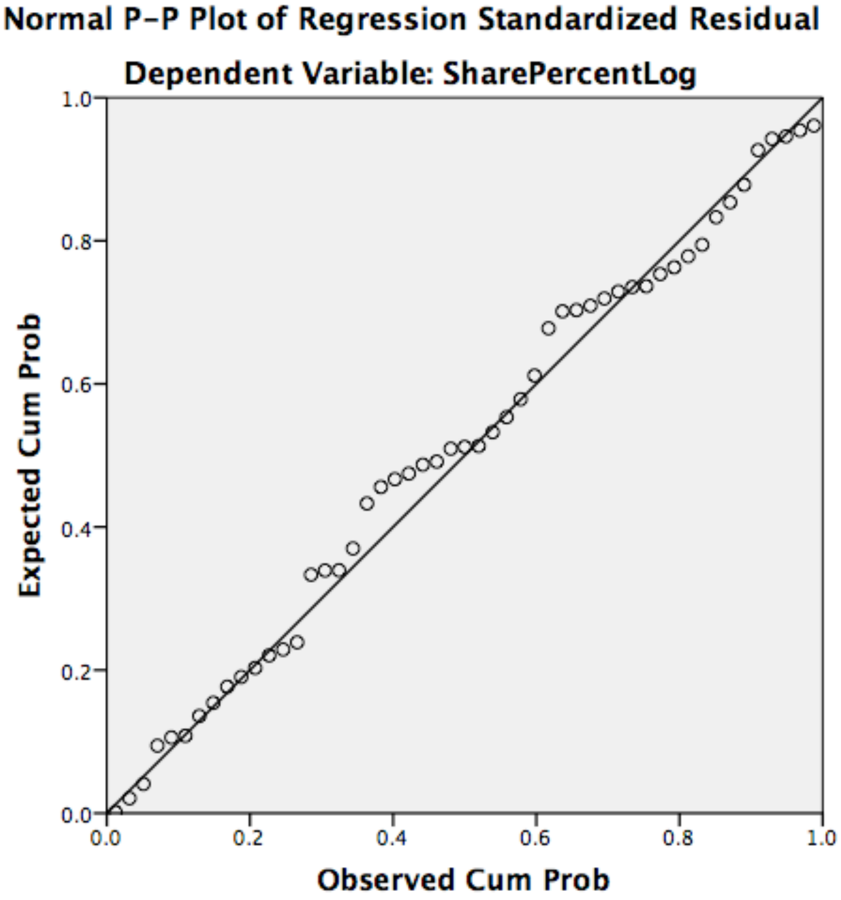


Figure 5.23: Plot of predicted and actual residual values for relationship 3.

5.5.4 Relationship 4: Privacy concern and knowledge of privacy measures

The fourth relationship tested was that between privacy concern and the respondent's knowledge of Facebook's privacy measures. As can be seen in Figure 5.25, this relationship proved to have insufficient statistical support, with a significance score of 0.315. As such, the null hypothesis cannot be rejected.

Since the relationship was not proven, the assumptions made do not need to be checked. If they were found to have been violated, this would not make the possible relationship more acceptable.

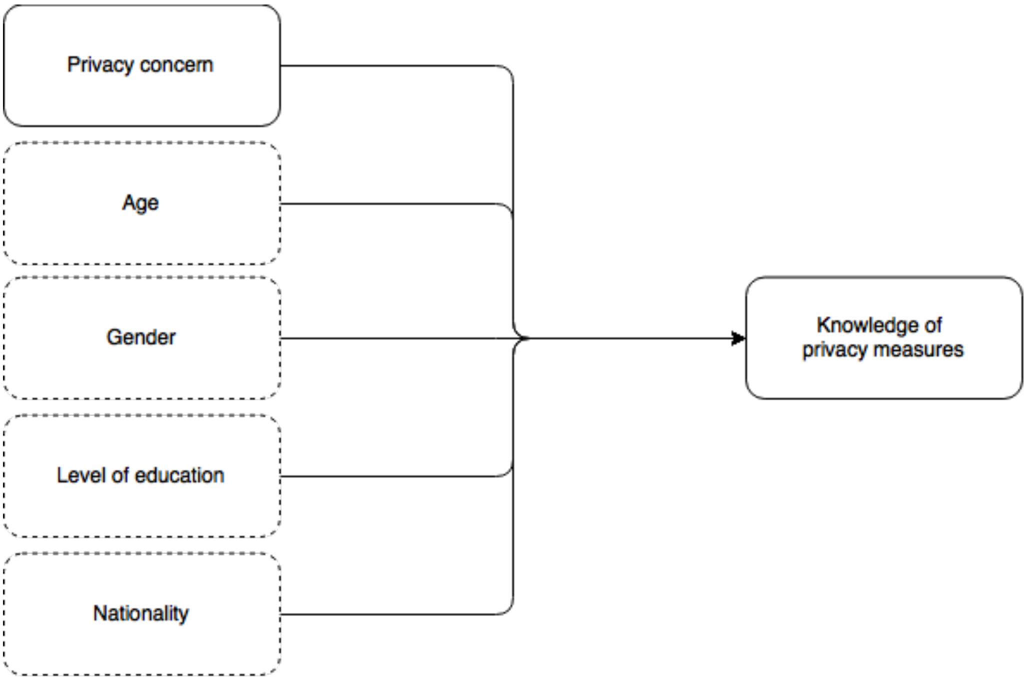
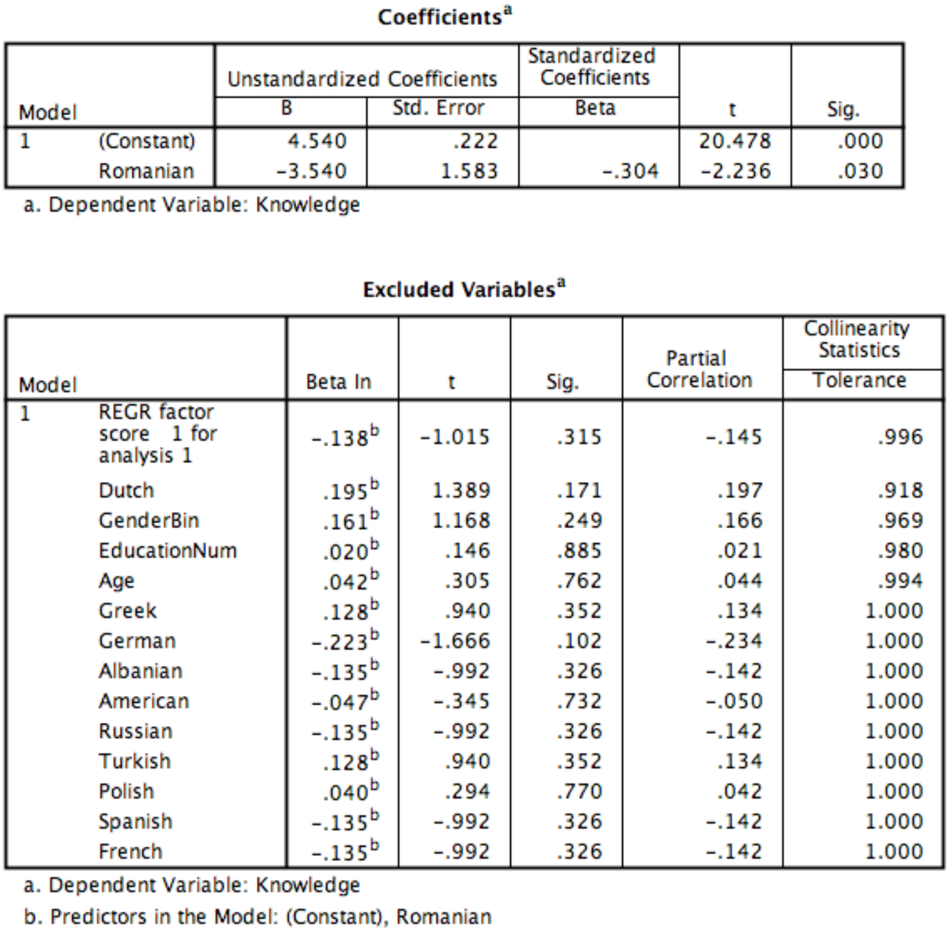


Figure 5.24: Concepts involved in relationship 4.



Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	4.540	.222		20.478	.000
	Romanian	-3.540	1.583	-.304	-2.236	.030

a. Dependent Variable: Knowledge

Excluded Variables[a]

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics
						Tolerance
1	REGR factor score 1 for analysis 1	-.138[b]	-1.015	.315	-.145	.996
	Dutch	.195[b]	1.389	.171	.197	.918
	GenderBin	.161[b]	1.168	.249	.166	.969
	EducationNum	.020[b]	.146	.885	.021	.980
	Age	.042[b]	.305	.762	.044	.994
	Greek	.128[b]	.940	.352	.134	1.000
	German	-.223[b]	-1.666	.102	-.234	1.000
	Albanian	-.135[b]	-.992	.326	-.142	1.000
	American	-.047[b]	-.345	.732	-.050	1.000
	Russian	-.135[b]	-.992	.326	-.142	1.000
	Turkish	.128[b]	.940	.352	.134	1.000
	Polish	.040[b]	.294	.770	.042	1.000
	Spanish	-.135[b]	-.992	.326	-.142	1.000
	French	-.135[b]	-.992	.326	-.142	1.000

a. Dependent Variable: Knowledge

b. Predictors in the Model: (Constant), Romanian

Figure 5.25: SPSS output of coefficients and excluded variables of the linear regression analysis.

5.5.5 Relationship 5: Knowledge and use of privacy measures

The final relationship that was researches was between the user's knowledge about and use of Facebook's privacy measures. Like the last relationship, these two concepts are not sufficiently connected to one another, showing a significance score of 0.834. The complete overview of coefficients can be seen in Figure 5.27. As with the last relationship, assumptions will not be checked, as they are no longer of interest.

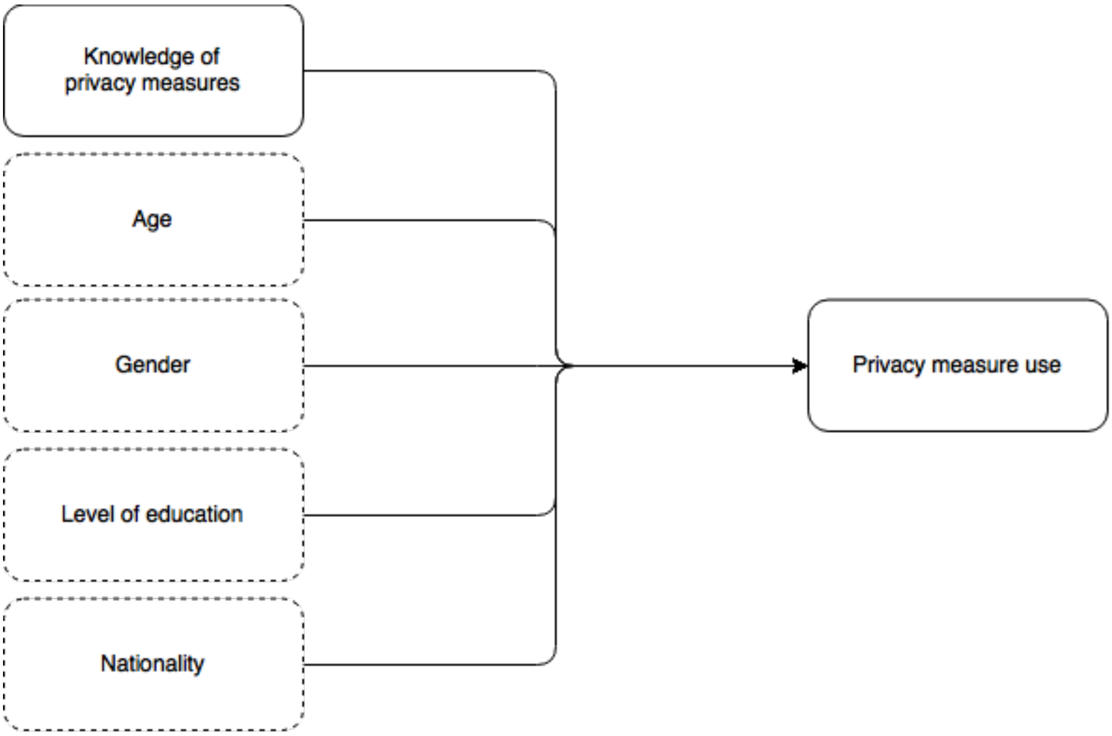
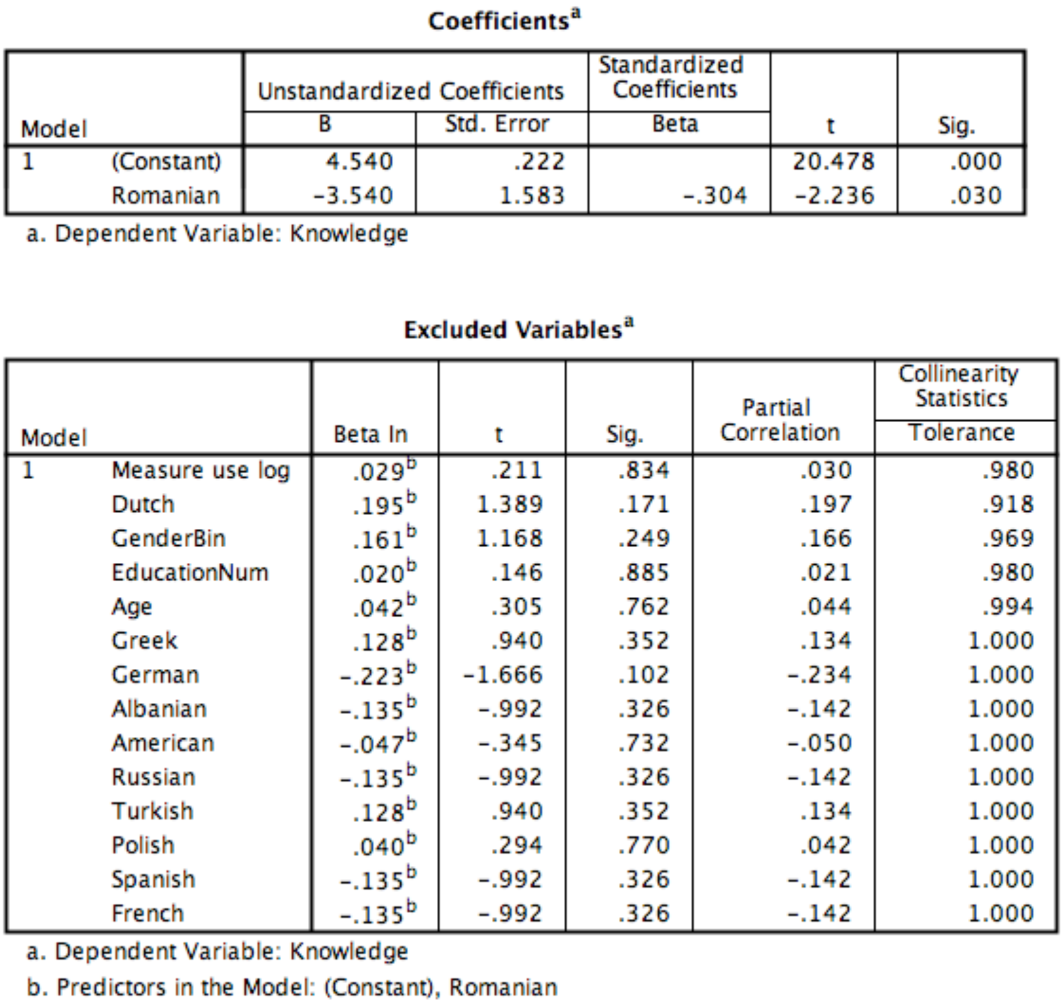


Figure 5.26: Concepts involved in relationship 5.



Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	4.540	.222		20.478	.000
	Romanian	-3.540	1.583	-.304	-2.236	.030

a. Dependent Variable: Knowledge

Excluded Variables[a]

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics
						Tolerance
1	Measure use log	.029[b]	.211	.834	.030	.980
	Dutch	.195[b]	1.389	.171	.197	.918
	GenderBin	.161[b]	1.168	.249	.166	.969
	EducationNum	.020[b]	.146	.885	.021	.980
	Age	.042[b]	.305	.762	.044	.994
	Greek	.128[b]	.940	.352	.134	1.000
	German	-.223[b]	-1.666	.102	-.234	1.000
	Albanian	-.135[b]	-.992	.326	-.142	1.000
	American	-.047[b]	-.345	.732	-.050	1.000
	Russian	-.135[b]	-.992	.326	-.142	1.000
	Turkish	.128[b]	.940	.352	.134	1.000
	Polish	.040[b]	.294	.770	.042	1.000
	Spanish	-.135[b]	-.992	.326	-.142	1.000
	French	-.135[b]	-.992	.326	-.142	1.000

a. Dependent Variable: Knowledge

b. Predictors in the Model: (Constant), Romanian

Figure 5.27: SPSS output of coefficients and excluded variables of the linear regression analysis.

5.6 Statistical Outliers

One important aspect to look at when conducting linear regression analysis is the role of statistical outliers, as these can greatly affect the result. To find potential outliers, SPSS descriptive analyses were used, which resulted in 2 potential outliers. This can be seen most clearly in the graph shown in Figure 5.28. These 2 entries concerned respondents with very strict use of privacy measures, and with very little information disclosed. It should be noted that these entries can be rationalised easily. After all, users with a strong distrust of sites like Facebook, or with very strong concerns about disclosing personal information, would very likely act this way. Moreover, this group of users should generally be harder to reach with a survey that was primarily spread over Facebook, which also explains the small amount of respondents in this group.

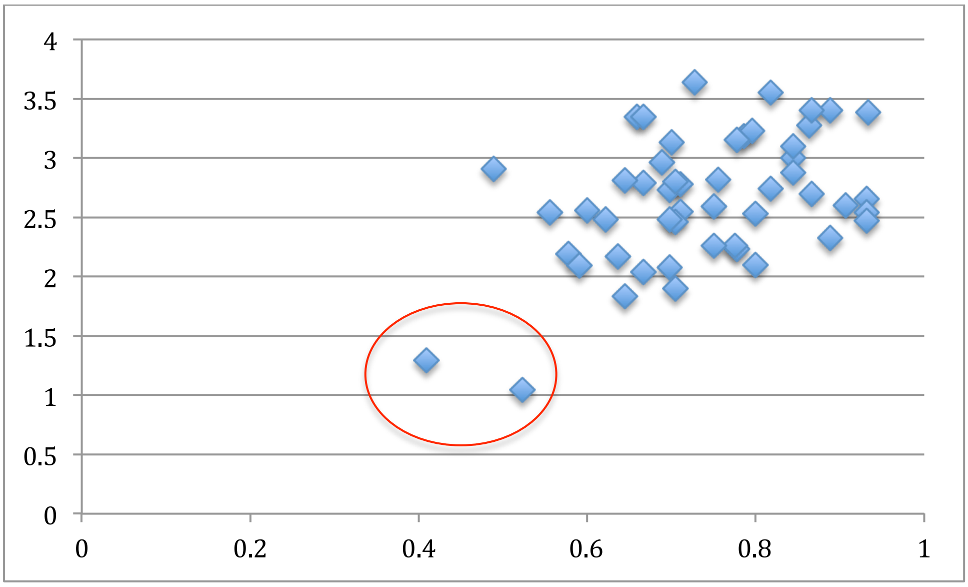


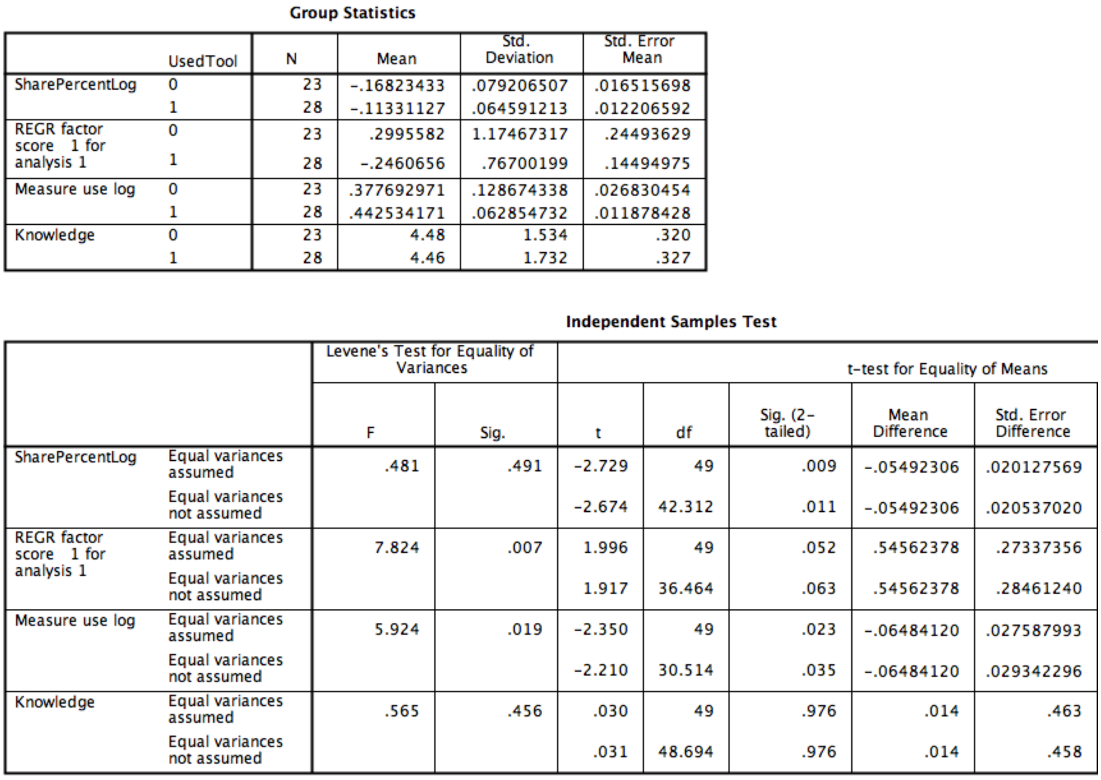
Figure 5.28: Scatterplot of online disclosure (X-axis) versus privacy measure use (Y-axis)

5.7 Groups based on tool usage

Since respondents could choose whether they would like to fill in the survey with or without the use of the data collection tool, they were split into 2 groups. As can be seen in section 5.4, the majority of respondents chose to use the tool-assisted version, but a substantial amount chose the alternative. Since the use of such a tool is quite new, some insight into these 2 groups may add to our understanding of the effect of these groups. If there is a considerable difference between the groups, this would be of interest, as it would show that having the option to not use the tool can influence what kind of respondents choose to take part in the study.As such, independent two-sample t-tests were performed on all the concepts collected. The hypotheses for this test were:

> H_1: Respondents who used the tool-assisted version of the survey differ significantly from those who did not, with respect to their privacy concerns/online disclosure/use of privacy measures/knowledge of privacy measures.
>
> H_0: Respondents who used the tool-assisted version of the survey do not differ significantly from those who did not, with respect to their privacy concerns/online disclosure/use of privacy measures/knowledge of privacy measures.

The results of the tests conducted can be seen in Figure 5.29.



Group Statistics

	UsedTool	N	Mean	Std. Deviation	Std. Error Mean
SharePercentLog	0	23	-.16823433	.079206507	.016515698
	1	28	-.11331127	.064591213	.012206592
REGR factor score 1 for analysis 1	0	23	.2995582	1.17467317	.24493629
	1	28	-.2460656	.76700199	.14494975
Measure use log	0	23	.377692971	.128674338	.026830454
	1	28	.442534171	.062854732	.011878428
Knowledge	0	23	4.48	1.534	.320
	1	28	4.46	1.732	.327

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means				
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference
SharePercentLog	Equal variances assumed	.481	.491	-2.729	49	.009	-.05492306	.020127569
	Equal variances not assumed			-2.674	42.312	.011	-.05492306	.020537020
REGR factor score 1 for analysis 1	Equal variances assumed	7.824	.007	1.996	49	.052	.54562378	.27337356
	Equal variances not assumed			1.917	36.464	.063	.54562378	.28461240
Measure use log	Equal variances assumed	5.924	.019	-2.350	49	.023	-.06484120	.027587993
	Equal variances not assumed			-2.210	30.514	.035	-.06484120	.029342296
Knowledge	Equal variances assumed	.565	.456	.030	49	.976	.014	.463
	Equal variances not assumed			.031	48.694	.976	.014	.458

Figure 5.29: Test statistics of Independent Samples test between groups based on tool usage (cropped for legibility reasons)

The tests show the 2 groups varied significantly in 2 of the 4 concepts measured. For online disclosure and privacy measure use, the relationships found were significant at 0.011 and 0.023, respectively. This allows us to reject H_0 for these concepts. It should also be mentioned that the concept of privacy concern shows a nearly significant relationship at 0.052. The only variable that was found to be unrelated to the groups was knowledge. Based on these findings, we can conclude that respondents who used the tool generally disclosed more about themselves on Facebook and had stricter privacy settings. Moreover, we cannot exclude the possibility that respondents who used the tool were generally more concerned with their privacy than those who did not.

There are 4 assumptions that are made for a two-sample t-test. As explained by Andy Field (2009), these are:

- The sample values are normally distributed
- The data is measured at, at least, the interval level
- The values are independent
- The variance in the 2 groups is (close to) equal

The first of these assumptions had already been tested in section 5.3.2, where we saw that all variables were likely normally distributed. The assumption concerning level of measurement is also true for these variables, thanks to the transformations described in section 5.3.1. This brings us to the assumption of independence. For this, the same is true as what was described in the introduction to section 5.5, namely that, while we cannot be completely certain that the data collected is completely independent, we can be fairly sure that it was. Moreover, there was only a single duplicate entry in the data set, and it was removed.

The last assumption concerns the variance of the 2 groups. As can be seen in Figure 5.19, this assumption was tested for each of the concepts. For online disclosure, and knowledge, this assumption was violated when using $p < 0.05$. Based on whether the assumption was found to be true or not, the relevant level of significance can be found in the table.

6. Discussion

Based on the results presented in the previous chapter, we will discuss the conclusions and considerations that resulted from the study. First, we will discuss the final version of the model, which can also be seen in Figure 6.1. Next, we will elaborate on the role of the data validation tool. To close off, threats to the validity of this research will be discussed, and we will examine opportunities for future research.

6.1 Confirmed relationships

Of the 3 confirmed relationships, 2 had already been confirmed by earlier studies. In this section, we will discuss the relationships found, how this relates to earlier findings, and what the implications of the final model are.

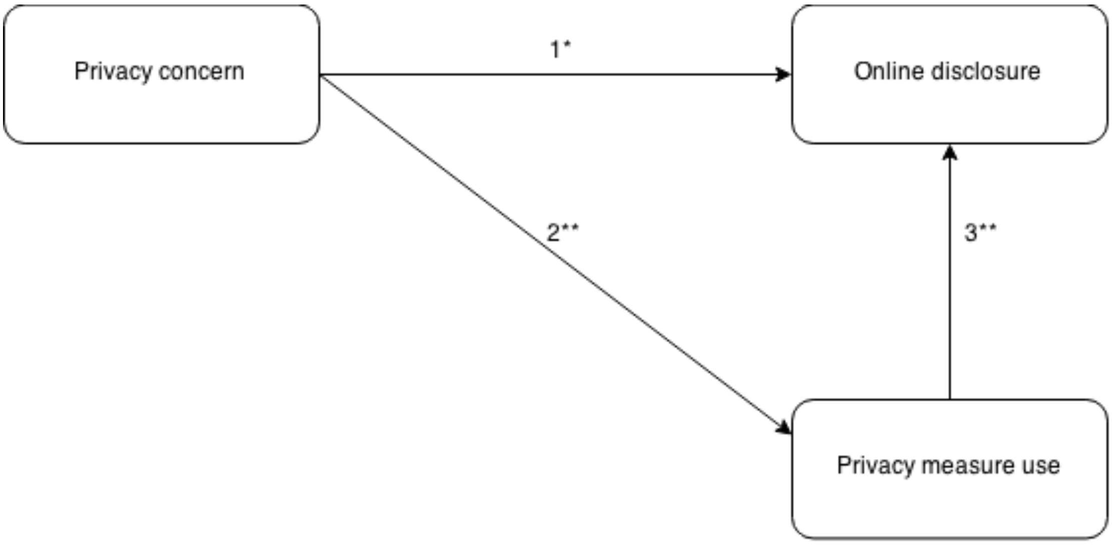


Figure 6.1: Final model of relationships found. Relationships indicated with * are significant at a level of p < 0.01, while those indicated with ** are significant at a level of p < 0.001.

Relationship 1 showed that people, who were more concerned with their privacy, generally disclose less information on Facebook. While this is logical, this finding has not been consistent throughout literature. As was discussed in chapter 2, studies in 2006 and 2012 also found a relationship between these 2 concepts, but a study in 2008 did not. The first 2 papers used different methods in order to research this relationship. Joinson et al. (2006) used a survey with privacy concern questions based on the work of Westin, while measuring disclosure with hypothetical "would you share"-questions. This survey was filled in by 759 students. Based on the data collected, they found that both privacy concern and trust are related to disclosure behaviour. Moreover, they found that perceived privacy affected trust. The most important limitation of this study is the hypothetical nature of the questions measuring online disclosure. The answers given to these questions may not accurately reflect the sharing behaviour of people on social media. The fact that the sample likely does not reflect the actual Facebook population, as there are many people on the social media platform, who are not a student, is also an important shortcoming.

The second study able to confirm the relationship was performed by Stutzman, Gross & Acquisti (2012). Their study used annual snapshots of a database consisting of publicly available Facebook profiles to gather data on how sharing behaviour changed over time. A total of 5076 Facebook users within the Carnegie Mellon University network were included in this study, which took place over the course of 7 years. They base their finding that the increased awareness of the consequences of online disclosure on an increasing tendency to seek privacy on social media, by limiting access to certain kinds of information, or restricting access to it. Here, we see that the assumption is implicitly made that privacy concerns can be measured by one's sharing behaviour. While this had, indeed, been confirmed by earlier studies, it remains an indirect measure for this factor. Considering the overall aim of the study, which was to observe trends on sharing behaviour using the database snapshots, it may not have been its priority to accurately measure the participant's privacy concerns.

The last study, which was unable to confirm this relationship, was Tufekci (2008). Like the study by Joinson et al. (2006), a survey was used to collect all information required. This survey was filled in by 704 university students. For measuring privacy concern, respondents were simply asked how concerned they were about their online privacy, with answers based on a 4 point scale. Online disclosure, on the other hand, was measured by asking whether the students had disclosed their real names on Facebook and/or Myspace. Based on this data, the author was unable to establish a relationship between online privacy concerns and disclosure. For this study, the most limiting factor seems to be the fact that online disclosure was only measured based on a single option for disclosure, namely the use of one's real name. There is a good chance that this does not reflect the amount of personal information being shared. Moreover, the use of a nickname may also be considered as disclosure of personal information, if this is the name the respondent uses often. Clearly, the measure used for this factor in particular fell short, which may explain why the relationship could not be established.

Looking at these shortcomings and how they were handled in our study, we attempted to include both students and non-students by not restricting participation. Moreover, we tried to measure the factors involved in the most direct way possible: by asking respondents to copy them from their source. This allowed our study to get a more complete and more correct idea of what information had actually been disclosed. Considering the significance of the relationship as well as the effect size, we consider the evidence for this relationship to be quite substantial, especially considering earlier findings.

Moving on to relationship 2, we found that people who are more concerned with their privacy generally use privacy measures more strictly. This relationship had also been confirmed in an earlier study, namely that by Mohamed & Ahmad (2012). In this study, a survey filled in by 345 students was used to measure the various factors. For privacy concern, was split into 6 separate components, namely self efficacy, perceived severity, perceived vulnerability, response efficacy, reward and gender. Each of these components were measured using between 2 and 5 questions, all of which used a Liktert scale. Unfortunately, the paper does not describe how privacy measure use was determined. For this reason, it is hard to judge whether the findings are accurate. This relationship now has been properly established in our study with a clearly defined method, providing a valuable addition to existing literature.

The third relationship, like the previous 2, had been the object of study for earlier research as well. However, as was discussed in chapter 2, this study proved to be partially inconclusive. Now, however, this relationship has been confirmed, showing that people with stricter use of privacy measures generally disclose less information.

An interesting aspect of the findings is that, unlike the intended purpose of privacy measures, we could not confirm that they performed any mediating function in relationship 1. In fact, when the regression test was performed with online disclosure as dependent variable, and both privacy concern and privacy measure use (along with all the control variables) as potential factors, privacy concern was not included in the regression model. Privacy measure use, on the other hand, was included. This shows that relationship 3, indeed, is stronger than relationship 1, even to the point that relationship 1 does not significantly add to the regression model once relationship 3 has been entered. Of course, there is much overlap between the variation in these variables, as was shown with relationship 2.

There is one aspect of the results that deserves some more clarification, namely the unstandardized B-coefficients for the relationships found. As can be seen, some of these appear to be quite low. However, this is not an indication of low statistical power. In section 6.4.4, we will demonstrate that almost all of the confirmed relationships were found to have a large effect.

Looking at the assumptions made for the relationships, we can clearly see that all those that are measurable have not been violated. Moreover, we have good reason to believe that the remaining assumptions also should not be a problem for our study, as was explained in section 5.5. This means that the relationships found have no important statistical side notes. All other limitations and threats to validity will be discussed in section 6.4.

This brings us to the question of what the implications are of the final model. Most interestingly, our findings show that people generally do not use privacy measures to enable themselves to share more information with a more restricted group of people. While this goes against the theory that these measures are meant to help compensate for possible privacy concerns in this matter, they may play another role. If Facebook had not implemented privacy measures such as these, people with strong privacy concerns may have chosen not to use the platform anymore. Indeed, the choice of whether to use a social media platform has been left out of this model, as it would fall outside our defined scope, which was limited to Facebook users. This was primarily done to allow a focus on what information people disclosed to whom, and use this focus in our data collection strategy. Nevertheless, future studies into the effect of privacy measures may want to give the option to respondents to say that they would not use the platform in certain situations.

It is absolutely not the case that privacy measures do not achieve what they are designed to do. Seeing that the usage of these measures is clearly and strongly linked to the user’s privacy concern does show there is a certain need for options like these. However, one should have the right expectations when implementing privacy measures. Future studies may reveal the impact of different implementations of privacy measures, aiding in choosing one that is right for the system at hand.

6.2 Rejected relationships

The model contained 2 proposed relationships that could not be confirmed by the data collected. This part of the model can be seen in Figure 6.2. in this section, we will examine several aspects of these relationships, including the measurement of knowledge of privacy measures and possible reasons why these relationships could not be confirmed. These insights may prove valuable for future research aimed at finding relationships that include the user’s knowledge of privacy measures.

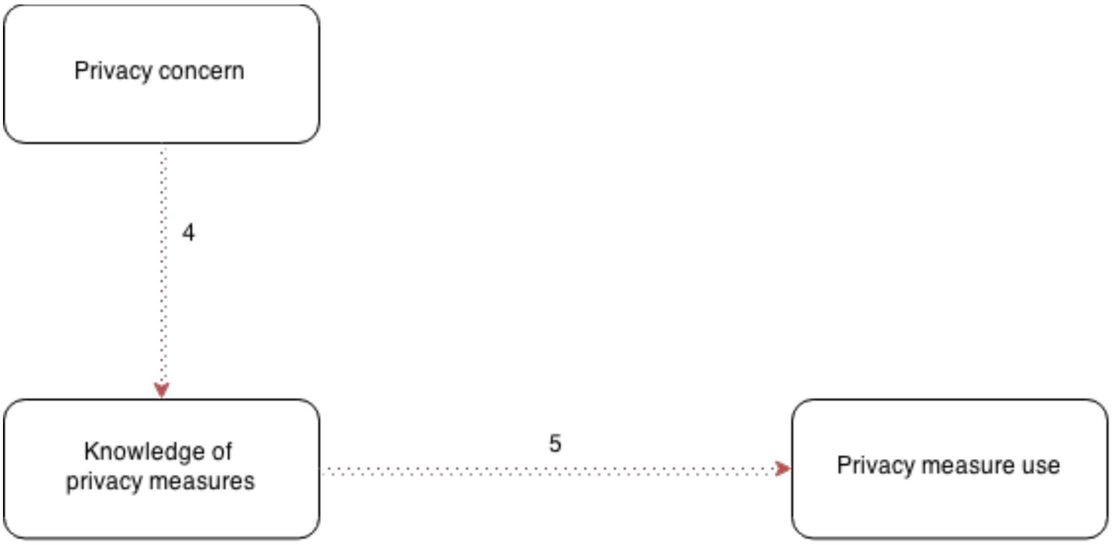


Figure 6.2: Rejected relationships in the model.

As both of these relationships include the concept “Knowledge of privacy measures”, it is only logical that we examine this variable of the model first. As was pointed out in chapter 4, no earlier methods of measuring this concept were found. For this reason, we proposed our own method, using 8 factual questions to calculate a score that reflected the number of correct answers. When we look at the outcome of this section of the survey, we can see that this method resulted in data that was well spread out. A histogram of this variable can be seen in Figure 6.3.

The results of the normality tests (as conducted in 5.3.2) was also promising, showing significance for the Kolmogorov-Smirnov test of 0.003, and of 0.035 for the Shapiro-Wilk test. As such, we conclude that the method employed was correct for our purposes and resulted in the data that was desired. Of course, the actual questions asked may be switched to improve the measurement even more. In doing this, it is important to keep in mind the difficulty of every question, to ensure the balance is kept.

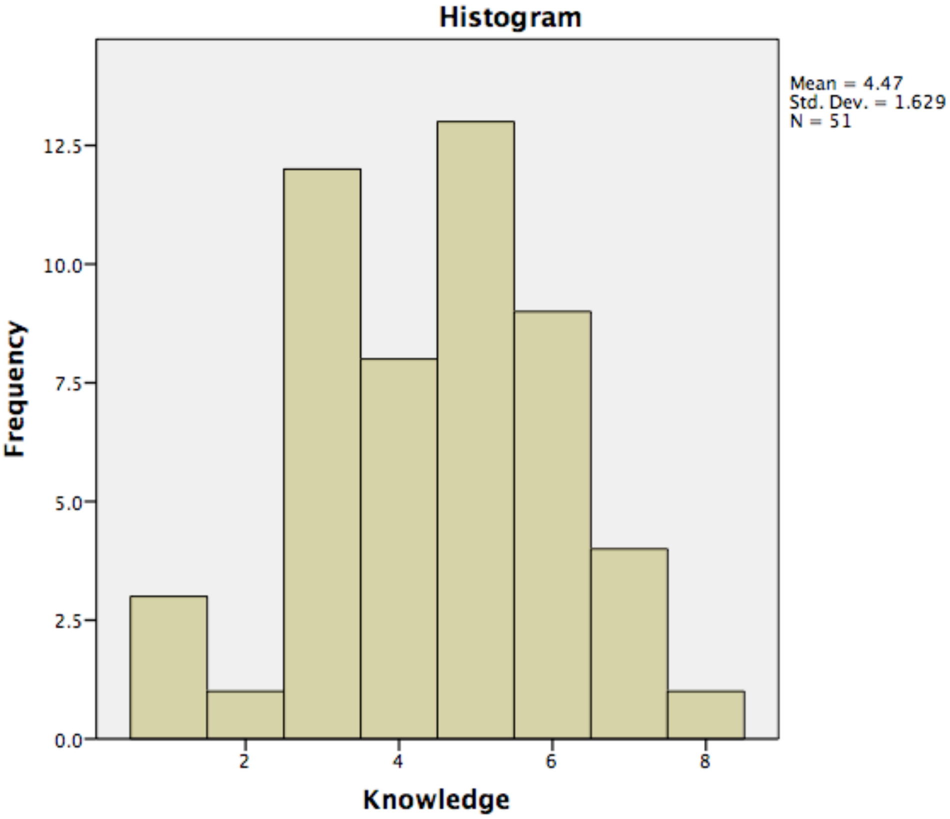


Figure 6.3: Histogram of the concept Knowledge of privacy measures.

This brings us to possible reasons why the relationships were not as expected. As was shown in the model, the expectation was that people with more concerns about their online privacy would know more about Facebook’s privacy measures, and that those who know more about the platform’s privacy measures would use them more strictly. Being unable to find supporting data for these propositions means several things. Firstly, it reflects that Facebook users do not need to be worried about their privacy to know about how privacy is handled on the platform. Secondly, it shows that users do not need to know exactly how the privacy-related aspects of the platform work to be able to reflect their privacy concerns in their own privacy settings. After all, use of privacy measures was found to be related to privacy concerns, but not to knowledge of these privacy measures. These findings should be good news for Facebook. Privacy is one of the major issues at play on the social media platform, and these findings show a significant effort on their part to make the platform and its settings accessible to all.

There are, however, other possible candidates for concepts that may be related to the knowledge of privacy measures. Since these privacy measures are all part of the platform's rules and functionalities, it may very well be that knowledge of these measures is more closely related to the amount of time users spend on Facebook. Alternatively, or possibly additionally, it may also be related to the use of Facebook's features. After all, the more a person uses the various parts of Facebook, the more exposure he/she gets to the platform and the way it functions. These may form interesting topics for future research.

6.3 Use of the tool

One aspect in which this study was exceptional was the use of a data collection tool. There are many different conclusions and considerations for future studies that resulted from this application, which will be the focus of this section. We will discuss the role of the tool as part of this study, the application of a data collection tool instead of a data validation tool, as well as the availability of a version of the survey that did not use the tool provided.

In our study, the tool was used to compare the information found with what the respondent had submitted in the survey. This was a big compromise, considering the original intention of the tool. Fortunately, the majority of respondents chose to use the tool-assisted version of the survey, providing us with a large amount of valuable and verifiable results. As was discussed in chapter 5, of the 3 profile items the tool looked at, 2 were used to determine the correctness of the information provided. The excluded variable, namely whether the profile had been verified, showed very different results than were to be expected. Moreover, it was discovered that users could enter their phone number in Facebook's settings menu, without it appearing on their profile. This is in contrast to the user's e-mail address, which will always appear in both. Based on these considerations, it was decided that the "verified" item on user profiles could not be used to accurately determine whether the respondent had submitted incorrect information.

The remaining 2 items the tool collected, namely whether the respondent had disclosed his/her e-mail address and last name, did work as expected. Considering that these items are, in many cases, mandatory, there was little variation in the data for these items. Nevertheless, they could be use very well to determine whether the information that had been submitted was, indeed, correct. In this, we consider the use of the tool to be a success, and a valuable addition to our study.

It is very unfortunate that the tool could not be implemented as it was originally intended. The tool's purpose was to allow respondents to log in with their Facebook account, after which it would collect information on whether each of the items included in the disclosure aspect of the study were present on the profile. Obviously, this would require an extensive array of permissions from the respondent. To protect against applications that ask too much of their users, Facebook forces each application that uses more than a small set of pre-approved items from a user's profile to undergo an approval process. In this process, the application is examined to determine whether each of the permissions requested is relevant for the stated purpose, as well as several other aspects.

The original tool was submitted several times to be examined by Facebook for this purpose, and was repeatedly rejected. The reasons varied somewhat, but primarily consisted of the desire to avoid back-end analytics. We tried to compensate for the objections, using different arguments for why the application serves a user-focussed purpose as well, but also using different versions of the applications that would take away some of the stated concerns. Nevertheless, all of the versions of the tool were rejected, leaving as only option to use the pre-approved set of data.

Since the tool could only get access to a small number of items, it was used to validate the information submitted by respondents. Should the original, full version of the tool have been used, the complete set of disclosure items could have been collected, assuring that the information collected would be completely correct. A data collection tool, instead of a data validation tool, could prove very valuable for future research into disclosure behaviour. In order to gain access to the permissions required for this, a change in the approval process of Facebook is definitely required. As it stands now, the requirements do not allow for any substantial scientific application to make use of the information available on Facebook, regardless of whether the user will allow access to it. While this may also be to protect their position as a social media platform, it may very well stand in the way of a deeper understanding of online behaviour, even beyond that of disclosure of personal information. Perhaps future research may be conducted with the agreement of Facebook, but as it stands now, the platform is shutting its doors for research, without any avenue for discussion or compromise.

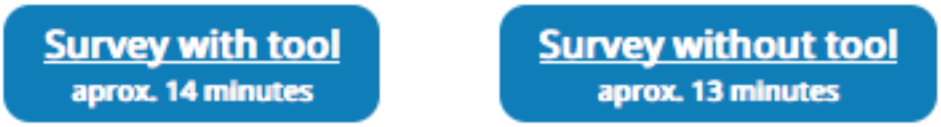


Figure 6.4: choice between the versions of the survey, as presented to respondents

Another aspect of interest is the availability of a version of the survey without the use of the data collection tool. The choice respondents were presented can be seen in Figure 6.4. Not only did this allow for more respondents to take part in the survey (as some might object to the use of such a tool), but also provided interesting information on the 2 groups created by this choice. In section 5.7, we examined the differences in terms of the concepts measured between these 2 groups. The results showed that in at least 2 of the 4 concepts, the groups differed significantly, with a third being close to statistical significance as well.

The difference between the 2 groups supports the decision of allowing respondents to choose whether to use the data validation tool. Since self-selection can create a bias in studies conducted in this way, forcing respondents to use the tool could have strongly shifted the data, as respondents who had previously used the tool may choose not to participate. On the other hand, allowing for a version of the survey leaves us with the risk that some of the submissions may not be completely valid. Depending on the purpose, future studies should consider having both options for their respondents, as it appears to combine the best of the 2 options: validated responses for one part, while allowing people with concerns to take part in the study.

6.4 Threats to validity

As with any study, there are various aspects that need to be kept in mind when looking at the results. In this section, we will discuss what issues may be a threat to the validity of our findings. In this, we distinguish between 4 different kinds of threats, as categorized by Cook, Campbell & Day (1979), namely internal, external, construct and conclusion validity. These will form the basis of the structure of this section. It should be noted that this approach was chosen after comparing it to several alternatives. Wiersma (2012) suggests using only internal and external validity to assess the validity of survey-based studies, while Guion (1980) suggests that all validity assessments should be based on the trinity of content, criterion and construct validity. The 4 categories proposed by Cook, Campbell & Day (1979) will allow us to delve deeper into the statistical side of the study, while simultaneously keeping a clear distinction between threats to internal and external validity.

For the remainder of this section, we will use the threats named and explained by Wohlin et al. (2012) as a primary basis for our analysis. Please note that any omission of threats named in their work implies merely that we do not believe this to have been applicable to our study. This is only logical, considering the list of threats named should cover a wide range of different studies. Moreover, any other relevant threats not named by Wohlin et al. (2012) will still be discussed in the appropriate section. We used their work to structure the possible concerns, and to give as complete an overview as possible.

6.4.1 Internal validity

Threats to internal validity refer to the ability to draw conclusions from the study performed. In this section, we will explore several issues that may be cause for concern and should be kept in mind when looking at the results.

The first threat we will look at is that of selection. As the study was conducted using a snowball method of data collection, selection may have been an issue in several respects. First of all, the starting points of the snowballing method may have caused a specific group to be targeted. Secondly, by allowing people to decide for themselves whether they will take part in the study, a degree of self-selection bias may also have occurred. When we look at the demographics of the sample, as presented in section 5.4, we can see that the sample population is likely not representative of the general population of Facebook users. This is mostly evident when looking at the highest level of education completed. As should be expected, with the survey having been primarily distributed among Facebook communities of students, the level of education is quite high. This may have affected the results. However, it should be noted that the majority of studies found in our literature study focussed solely on students. Moreover, self-selection was likely reduced by allowing respondents to fill in the survey without using the data validation tool, as was explained in section 6.3.

Another issue that may have played a role concerning the sample is the sample size. As was indicated in section 4.2, the selection bias of the snowball sampling method can be overcome by having a larger sample. While the sample size is quite satisfactory (60 or 51, depending on how it is counted), it still falls on the lower part of the spectrum when compared to the samples used in the studies included from literature.

Moving on to the issue of instrumentation. This threat refers to the possibility that the way the information was collected may have affected the results. This aspect may have played a role in two ways. First of all, respondents may have been conditioned to be more privacy conscious by the decision not to intersperse unrelated questions in the segment of the survey testing privacy concern. This was done in an earlier study, but it was decided not to do so here, as this would make the already long survey even longer.

The other way in which instrumentation may have played a role is in with respect to online disclosure and privacy measure use. In this part of the survey, respondents were asked to log in on their Facebook profile and answer the questions based on what was already set. However, several respondents reported to have changed their settings as a result of this survey. We anticipated this, and asked respondents to submit their original settings and disclosed items, rather than what they would be after these changes. This was done because we wanted to minimize the role of the survey in measuring these aspects, even the new settings may arguably be give a better representation of the respondent's intended disclosure and use of privacy measures.

Another issue with instrumentation concerns the verification of the information provided. As was discussed at length in section 4.2.1, we allowed respondents to choose whether they wished to use the tool. If they did, we were able to verify part of the information provided, and decide to exclude responses based on this. This does leave us with a number of responses that were not verified, and may include invalid values. On the other hand, allowing for 2 versions of the survey likely reduced the self-selection bias for our study. We recognise both possible threats, and believe that the balanced approach used was the right one.

The next issue to tackle is that of respondent mortality. Indeed, it may have occurred that respondents started the survey, only to leave it unfinished. Unfortunately, it is hard for us to assess to what degree this was an issue, and how it affected the outcome. This is because the survey was not done in a controlled environment, but by respondents on their computers, wherever they might be. Nevertheless, we should mention this potential limitation.

The last issue that may be relevant for our study is that of ambiguity about direction of causal influence. In the model constructed, the relationships were proposed based on earlier studies and further argumentation. Nevertheless, we cannot be certain that the relationships are the proposed direction. This should be an interesting topic for future studies, which will be discussed further in section 6.5.

6.4.2 External validity

External validity looks at threats that may limit the generalizability of our results. The issues discussed below, then, should be taken into account when trying to apply what we have learned in our study to other situations.

The first threat mentioned is that of interaction of selection and treatment. This refers to the issue that the population of the sample may not be representative of the population we would like to generalize to. As was discussed in the previous section, there are some issues with the demographics of the sample that should be kept in mind in this respect. Should this prove to be an issue, however, this also casts a doubt on many of the earlier studies conducted in this field.

Another issue in this respect is broader question of what people were reached with the survey, and who chose to respond to it. As discussed previously, the snowball sampling method will likely not give the most accurate distribution within a population. This limitation can also be seen in the demographics, but may have played a role in other ways that were not measured by the survey.

Another possible threat in this respect is the possible interaction of history and treatment. With this threat, we consider the possibility that the timing of the test may have somehow affected its results. While we cannot exclude the possibility that anything relevant occurred to respondents on an individual level, no major changes occurred to the user-side of Facebook during or shortly prior to the study. The only aspect that was changed on the platform is what API may be used by apps. Therefore, we consider the effects of this threat to be limited in scope.

6.4.3 Construct validity

When we look at the construct validity, we consider the degree in which the measurements can be generalised to the concepts we attempt to measure. As such, this section looks at the measurements set out in chapter 4, as well as the transformations performed in section 5.3.1 to assess the degree in which these values represent what they are intended them to.

The first threat stated by Wohlin et al. (2012) is that inadequate preoperational explication of constructs. As we explained earlier, the concepts knowledge of privacy measures, use of privacy measures and online disclosure have had little or no measurement with well-documented operationalization to base our concepts on. For this reason, we have proposed various methods to measure these concepts, and to transform these measurements into the desired scales. A problem with this is that the methods used may be lacking, in some respect, and may misrepresent the concepts. In this document, we have thoroughly explained our methods and reasoning in this respect. For some, the reasoning is quite straightforward, as was the case with the measurement of online disclosure. Others, such as the transformation of privacy measure use to numerical values, may prove to be more subjective. We acknowledge this, but we do believe that the approach used is correct for our purposes.

The role of confounding constructs is another that should be addressed. As we already discussed in section 6.2, there may be other concepts that play a role in this model. For knowledge of privacy measures, we proposed concepts such as time spent on Facebook, or use of Facebook's features as possible candidates. This may be an interesting topic for future research, which could add to our model.

One last item of (possible) concern is the Interaction of testing and treatment. As we discussed when looking at internal validity, the interaction between way we measured and what we measured may have played a role in several instances. While we took some steps to limit this possibility, it may still have played a role, most specifically for the concept of privacy concern. In this case, we did not use other irrelevant questions to avoid the respondent growing more concerned about their privacy with the questions, as was done in some earlier studies. However, many other studies with an exclusive focus on privacy (such as Westin's studies as examined by Kumaraguru & Cranor in 2005) did not do this either. This somewhat eases the concerns for this threat.

6.4.4 Conclusion validity

The last category of threats to validity we will examine concern the validity of the conclusion. These threats look at whether the correct conclusions can be drawn from the research conducted.

The first of the possible threats in this case is the low statistical power of the relationships found. For this, we look primarily at the R squared statistics, as these show exactly how much of the variance in the data was explained by the relationships found. The percentage of variation explained for each of the 3 confirmed relationships can be seen in Table 6.1. Using the work of Cohen (1988, 1992) to assess the size of the effect, we can categorize relationships 2 and 3 as having a large effect, while relationship 1 had a medium to large effect. For this reason, we do not consider this threat to be a large concern.

Relationship	1	2	3
Variation explained	21.2%	29.0%	27.2%

Table 6.1: Variation in the data explained by regression for each relationship found.

Next, we need to look at violated assumptions of statistical tests. This has already been discussed at length in the Results chapter, showing that all of the measurable assumptions appeared not to have been violated for the confirmed relationships. Nevertheless, there may have been some assumptions that have been violated, which we cannot be sure of. For example, an external variable (i.e. one that was not present in the model) may have played a role. Some examples of these were suggested in section 6.2, but another factor that may have played a role is IT literacy. Future studies may be able to add to the model by researching such relationships.

Conclusion validity is also strongly dependent on the reliability of the measures used, as well as the treatment implementation. We have already looked at the effect of way the concepts were measured and how they were operationalized when looking at internal and construct validity. The concerns discussed there should also be taken into account in this respect.

Wohlin et al. (2012) also mention random irrelevancies in experimental setting as a possible cause for concern. Since the survey was not taken in a controlled environment, we cannot exclude the possibility of other events interrupting or influencing respondents during the study. Future studies may choose to use a controlled environment for this study, though we expect the effect of this to be limited.

Random heterogeneity of subjects is another threat for the validity of conclusions. This refers to random variations within the sample population perhaps being stronger than the variation that needs to be measured. At the same time, however, too little variation in the sample may not be reflective of the overall population either. This may have also played a role in our study, as we did not have a pre-selected, representative sample for our population.

One final issue is that of fishing for results, which refers to the possibility researchers were looking for specific kinds of people for their sample to distort the relationships found. We have tried to have a diverse group using the tools available to us for this study. Admittedly, our sample may not be completely representative of the entire population, but we did strive to get a group that was as good a fit as was possible. Any random variation in the sample population should be purely coincidental, and we hope this was kept to a minimum.

6.5 Future research

Looking to the future, there are many different related aspects that still need to be researched. Moreover, the use of a data collection tool (or alternatively, a data validation tool) will provide new and improved methods of collecting data and assuring that it is correct. In this section, we will look at a number of potentially interesting topics for future studies.

Since we were unable to relate the concept of knowledge to other concepts in our study, a lot of potential for interesting findings remains. As was discussed in section 6.2, we suspect that knowledge is likely to be related to time spent on Facebook, or alternatively to the use of Facebook's features. Results of these studies, combined with our research, could potentially reveal a lot about knowledge of the privacy functionality of a social media platform. For instance, should it be related to time spent on Facebook, this would mean that while Facebook has made the privacy settings accessible to those, who do not look into the exact working of the platform, only those who use the platform for a significant amount of time actually understand how the privacy-related aspects of it function. This could indicate an area of improvement for the social media platform. For these studies, the method used to measure knowledge could be taken from ours. While our results were inconclusive, the method used to measure this concept did produce a well-spread normal distribution.

Looking more at the effect of privacy measures, a study that divides respondents into different groups with different privacy measures, and asks whether they would disclose their personal information to the system with these measures could give a lot of insights into what effect privacy measures actually have. Factors like perceived benefit and decision to join the network may also play a big role in this study. The results of such a study could be of great value not only for scientific purposes, but for business as well. Social media platforms like Facebook could see what the effect of having a wider range of privacy settings has on the user's disclosure and decision to use the platform. Moreover, future studies with where respondents may have privacy-based objections could benefit from this study, as it will allow them to find a better balance between data collection and privacy measures. Such a study would have to find a way to ensure that the hypothetical nature of the questions, as proposed, does not interfere with the validity of the answers.

Considering the large amount of factors that have been shown to affect the different aspects in our model, we can only assume that there will be many more variables that can be added. For the concept of knowledge, we already suggested 2 variables, which may be at play. These may, in turn also be related to the concept of online disclosure, as increased use of the platform could make disclosing personal information more (or, arguably, less) appealing. These are mere examples of variables that could be added, however. We welcome any and all additions to the model found.

The biggest problem for the generalisation of the results of our study is that the sample collected may not have been reflective of the entire Facebook user base. Considering the limited resources available for our project, we are confident our findings will provide a solid basis for future studies to confirm our findings with a more representative sample. If such a study would be allowed to use a data collection tool with access to all information on the user's profile required, this would add even more to the validity of the result. In this respect, we believe that this study provides a significant amount of fertile ground for future studies.

Looking at the use of a data collection tool more broadly, we strongly recommend that future studies implement a similar system. Depending on the situation and the privacy concerns at hand, researchers should also consider allowing for a version of the survey without the use of the tool, as this may affect what people respond to the survey. While it has now been implemented in the context of online disclosure, one can imagine other settings where such a tool might be useful for the collection of correct data. For instance, one can imagine such a tool being used for research into how people use a certain website or online platform. It should be noted that, as these new approaches develop, so should their standard methods of operationalization. With the methods we described in section 5.3.1, we aimed to get the best representation of the data as possible, which was fit for analysis. While we are confident that we have succeeded in this respect, this is a highly situational approach.

Building on this, a more detailed study looking into the willingness of respondents to take part in a study that employs a data collection tool may be of interest. In our study, we concluded there was a significant difference between the groups in several aspects. While it is reasonable to assume that this would affect self-selection if use of the tool would be mandatory, we cannot conclusively say what the effect of this would be. A study that looks into this could greatly improve our understanding of the effect of such a tool.

One last potentially interesting area of study is how privacy measures are related to offline disclosure. Obviously the privacy measures that are applicable in these instances are very different from those in an online setting, but the relationship may still be very interesting. Findings of this study could benefit studies that take place in offline environments and could potentially help more privacy-sensitive studies in these contexts to attract more respondents. Since the discussion about online disclosure versus offline disclosure is one that has lasted long and has been inconclusive, the findings of our study could function as a basis for a hypothesis for this study.

7. Conclusion

In our study, we managed to confirm the suspected relationship between use of privacy measures and online disclosure, showing that people who have strict privacy settings are likely to disclose less personal information. Moreover, we showed that people who are more concerned about their privacy are also likely to disclose less personal information, and are furthermore likely to have stricter privacy settings. All these relationships showed a high level of significance and explained large amount of the variation in the data. These findings seem to confirm that, while privacy measures are meant to ensure people are willing to share more information on Facebook, they are used in stead by people concerned with their privacy to restrict access to their personal information to those they see fit.

The message to take away from these conclusions is that awareness of the consequences of online disclosure are paramount in order to properly enable users to assess what and how they wish to disclose their personal information. When a user is aware of the effects of disclosing personal information, they will be able to translate their concerns into their use of privacy measures, as well as their decision to share certain kinds of information. Features like the ability to see one's profile as they would be shown to a friend or to a stranger, as is already implemented on Facebook, may very well help in this process if it gets the attention it deserves.

Unfortunately, we were unable to incorporate knowledge of the platform in our model. While the results do say something about the way Facebook has implemented its privacy measures, it does leave room for future studies to link this variable to other factors. We have suggested some, which may also lead to it being connected to our model in the future.

One important aspect of our study was the use of a data collection tool. It allowed us to validate part of the information in order to ensure the correctness of our data. We are convinced that the use of such a tool can prove invaluable to many studies, and can be used in a wide range of topics. However, there are some considerations that need to be made when implementing such a tool in order to ensure other threats to validity do not play a role.

Our findings are important for any project relying on the collection of personal data through voluntary online disclosure. As we have seen, implementing privacy measures will not encourage concerned participants to share more of their information, and compensating through stricter use of these measures. This does not mean that these measures are not important, since they may well play a role in the initial decision to take part in the project, be it scientific, commercial, or of any other nature. Our findings provide a solid basis for future studies to see how this decision is affected by the availability of various privacy measures. The fact that privacy measures do not compensate for stronger privacy concerns only shows how strong these concerns are, and how critical people have become when asked to share their personal information. Thus, with respect to awareness of our own behaviour, many people already appear to act according to their concerns, rather than what a platform encourages them to do.

With this study, we have managed to confirm several important relationships, while also providing a good basis for many interesting future works. Furthermore, we hope to have set a precedent by confirming our data using a specifically designed tool. In this way, we have contributed significantly to the understanding of how the various aspects involved in online disclosure are related, filling an existing gap in scientific literature.

References

Acoca, B. (2008). Scoping paper on online identity theft (Ministerial Background Report DSTI/CP, 2007). Retrieved from http://www.oecd.org/sti/40644196.pdf on 17 June 2015

Acquisti, A., & Gross, R. (2006). Imagined communities: Awareness, information sharing, and privacy on the Facebook. In *Privacy enhancing technologies* (pp. 36-58). Springer Berlin Heidelberg.

Atkinson, R. & Flint, J. (2001). Accessing Hidden and Hard-to-Rech Populations: Snowball Research Strategies. Social Research Update 33.

Baltar, F., & Brunet, I. (2012). Social research 2.0: virtual snowball sampling method using Facebook. *Internet Research* 22(1). pp. 57 - 74

Buchanon, T., Paine, C., Joinson, A.N., & Reips, U.D. (2007). Development of Measures of Online Privacy Concern and Protection for Use on the Internet. *Journal of the American society for information science and technology* 58(2), pp.157-165

Christofides, E., Muise, A., & Desmarais, S. (2012). Risky Disclosures on Facebook: The Effect of Having a Bad Experience on Online Disclosure. *Journal of Adolescent Research*. doi: 10.1177/0743558411432635

Christofides, E., Muise, A., & Desmarais, S. (2009). Information Disclosure and Control on Facebook: Are They Two Sides of the Same Coin or Two Different Processes? *CyberPsychology & Behavior*, 12(3), doi: 10.1089/cpb.2008.0226

Cook, T. D., Campbell, D. T., & Day, A. (1979). Quasi-experimentation: Design & analysis issues for field settings (Vol. 351). Boston: Houghton Mifflin.

Cohen, J. (1992). A power primer. *Psychological bulletin*, *112*(1), pp. 155.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Academic press.

Debatin, B., Lovejoy, J.P., Horn, A.K. & Hughes, B.N. (2009) Facebook and Online Privacy: Attitudes, Behaviors, and Unintended Consequences. Journal of Computer-Mediated Communication 15. doi:10.1111/j.1083-6101.2009.01494.x

Emanuel, L., Neil, G.J., Bevan, C., Stanton Fraser, D., Stevenage, S.V., Whitty, M.T., & Jamison-Powell, S. (2014) Who am I? Representing the self offline and in different online contexts. *Computers in Human Behavior*, 41(1), pp. 146-152

Field, A. (2013). Discovering statistics using IBM SPSS statistics. Sage.

Free, C., Phillips, G., Galli, L., Watson, L., Felix, L., Edwards, P., Patel, V., & Haines, A. (2013). The effectiveness of mobile-health technology-based health behaviour change or disease management interventions for health care consumers: a systematic review. *PLoS Med 10(1*): e1001362. doi:10.1371/journal.pmed.1001362

Govani, T., & Pashley, H. (2005). *Student awareness of the privacy implications when using Facebook.* Unpublished paper presented at the "Privacy Poster Fair" at the Carnegie Mellon University School of Library and Information Science, Pittburgh, Pennsylvania, USA.

Guion, R.M. (1980) On Trinitarian Doctrines of Validity. *Professional Psychology 11 (June):* 385-98

Heirman, W., Walrave, M., & Ponnet, K. (2013) Predicting Adolescents' Disclosure of Peronal Information in Exchange for Commercial Incentives: An Application of an Extended Theory of Planned Behavior. *Cyberpsychology, Behavior, and Social Networking*, 16(2). doi: 10.1089/cyber.2012.0041

Hui, K. L., Tan, B. C., & Goh, C. Y. (2006). Online information disclosure: Motivators and measurements. *ACM Transactions on Internet Technology*, *6*(4), pp. 415-441.

Joinson, A. N., Paine, C., Reips, U. D., & Buchanan, T. (2006). Privacy and Trust: The role of situational and dispositional variables in online disclosure. In *Workshop on Privacy, Trust and Identity Issues for Ambient Intelligence*.

Joinson, A. N., Reips, U. D., Buchanan, T., & Schofield, C. B. P. (2010). Privacy, trust, and self-disclosure online. *Human–Computer Interaction*, *25*(1), pp. 1-24.

Katwal, S. (2014, 29 August). Removing Your Last Name From Facebook [Updated]. Retreived from http://www.slideshare.net/suraj2kc/removing-facebook-last-name on 29 May 2015.

Kumaraguru, P. & Cranor L.F. (2005) *Privacy Indexes: A Survey of Westin's Studies.* Pittburgh, Pennsylvania, USA: School of Computer Science, Carnegie Mellon University.

Magid, L. (2014, 22 May). Facebook Changes New User Default Privacy Setting To Friends Only – Adds Privacy Checkup. Retrieved from http://www.forbes.com/sites/larrymagid/2014/05/22/facebook-changes-default-privacy-setting-for-new-users/ on 29 May 2015.

Mohamed, N., & Ahmad, I. H. (2012). Information privacy concerns, antecedents and privacy measure use in social networking sites: Evidence from Malaysia. *Computers in Human Behavior*, *28*(6), pp. 2366-2375.

Nguyen, M., Bin, Y. S., & Campbell, A. (2012). Comparing online and offline self-disclosure: A systematic review. *Cyberpsychology, Behavior, and Social Networking*, *15*(2), pp. 103-111.

Nosko, A., Wood, E. & Molema, S. (2010). All about me: Disclosure in online social networking profiles: The case of FACEBOOK. *Computers in Human Behavior*, *26*(1), pp. 406-418.

Rabkin, A. (2008, July) Personal knowledge questions for fallback authentication: Security questions in the era of Facebook. Paper presented at the meeting of the Symposium on Usable Privacy and Security, Pittsburgh, PA, USA.

Sharda, R., Aronson, J. E., & King, D. N. (2008). *Business intelligence: A managerial approach*. Upper Saddle River: Pearson Prentice Hall.

Stutzman, F., Capra, R., & Thompson, J. (2011). Factors mediating disclosure in social network sites. *Computers in Human Behavior*, *27*(1), pp. 590-598.

Stutzman, F., Gross, R., & Acquisti, A. (2013). Silent listeners: The evolution of privacy and disclosure on facebook. *Journal of Privacy and Confidentiality*, *4*(2).

Taddei, S., & Contena, B. (2013). Privacy, trust and control: Which relationships with online self-disclosure?. *Computers in Human Behavior*, *29*(3), pp. 821-826.

Tufekci, Z. (2008). Can you see me now? Audience and disclosure regulation in online social network sites. *Bulletin of Science, Technology & Society*, *28*(1), pp. 20-36.

Tuunainen, V.K., Pitkänen, O. & Hovi, M. (2009) *User's Awareness of Privacy on Online Social Netowrking sites – Case Facebook.* Paper presented at the 22nd Bled eConference, Bled, Slovenia.

Wiersma, W. (2012) The validity of surveys: online and offline. Retreived from http://papers.wybowiersma.net/abstracts/Wiersma,Wybo,The_validity_of_surveys_online_and_offline.pdf on 4 August 2015.

Wilson, D.W., Proudfoot, J. G. & Valacich, J.S. (2014) *Saving Face on Facebook: Privacy Concerns, Social Benefits, and Impression Management.* Paper presented at the 35th International Conference on Information Systems, Auckland, New Zealand.

Wohlin, C., Runeson, P., Höst, M., Ohlsson, M. C., Regnell, B., & Wesslén, A. (2012). Experimentation in software engineering. Springer Science & Business Media.

Yeung, K. (2013, 16 October). Facebook narrows default privacy settings for new teens users from 'friends of friends' to just 'friends'. Retreived from http://thenextweb.com/facebook/2013/10/16/facebook-narrows-default-privacy-settings-new-teens-users-friends-friends-just-friends/ on 29 May 2015.

Zimmer, J. C., Arsal, R. E., Al-Marzouq, M., & Grover, V. (2010). Investigating online information disclosure: Effects of information relevance, trust and risk. *Information & Management*, *47*(2), pp. 115-123.

Appendix A: Overview of the literature collected including search terms and meta analysis

Search queries used

- Online Disclosure
- Privacy awareness
- Facebook
- Facebook disclosure
- Facebook privacy
- Facebook security
- Disclosure security

Meta-analysis

Title	Authors	Year publication	Country	Field (Based on journal)	Participants in study	Participant group	Research Method	Analysis performed
Can You See Me Now? Audience and Disclosure Regulation in Online Social Network Sites	Zeynep Tufekci	2008	USA	Technology, Society	704	College students	Quantitative	Logistic Regression

Imagined Communities: Awareness, Information Sharing, and Privacy on the Facebook	Alessandro Acquisti and Ralph Gross	2006	USA	Computer Science	318	College students	Quantitative	Correlation, Multivariate Regression
Predicting Adolescents' Disclosure of Personal Information in Exchange for Commercial Incentives: An Application of an Extended Theory of Planned Behavior	Wannes Heirman, Michel Walrave, and Koen Ponnet	2013	Belgium	Psychology	1042	Pupils between 12 and 18 years of age	Quantitative	Regression
Hey Mom, What's on Your Facebook? Comparing Facebook Disclosure and Privacy in Adolescents and Adults	Emily Christofides , Amy Muise and Serge Desmarais	2011	Canada	Psychology	573	Youth group: 288 Facebook users between 9 and 18 of age. Adult group: non-students between 19 and 71 years of age	Quantitative	Correlation

Comparing Online and Offline Self-Disclosure: A Systematic Review	Melanie Nguyen, Yu Sun Bin, and Andrew Campbell	2012	-	Psychology	-	-	Qualitative	-
Facebook and Online Privacy: Attitudes, Behaviors, and Unintended Consequences	Bernhard Debatin, Jenette P. Lovejoy, Ann-Kathrin Horn, and Brittany N. Hughes	2009	USA	Communication	119	College students	Qualitative, Quantitative	Correlation
Factors mediating disclosure in social network sites	Fred Stutzman, Robert Capra, and Jamila Thompson	2010	USA	Information Science, Behaviour	122	Facebook users	Quantitative	Logistic Regression, Ordered and Multinomial Logit Regression

Information privacy converns, antecedents and privacy measure use in social networking sites: evidence from Malaysia	Norshidah Mohamed, and Ili Hawa Ahmad	2012	Malaysia	Information Science, Behaviour	340	College students	Quantitative	Regression
Investigating online information disclosure: Effects of information relevance, trust and risk	J. Christopher Zimmer, Riza Ergun Arsal, Mohammad Al-Marzouq, and Varun Grover	2010	USA	Information Science, Management	264	College students	Quantitative	Correlation
Online Information Disclosure: Motivators and Measurements	Kai-Lung Hui, Bernard C.Y. Tan, and Chyan-Yee Goh	2006	Singapore	Computer Science	687	371 college students exploratory research, 316 college students for confirmation survey	Quantitative	Correlation, regression

Personal knowledge questions for fallback authentication: Security questions in the era of Facebook	Ariel Rabkin	2008	USA	Computer Science	-	-	Qualitative	-
Privacy, trust and control: Which relationships with online self-disclosure	Stefano Taddei and Bastianina Contena	2013	Italy	Psychology	718	Online Social Network Users	Quantitative	Regression
Privacy, Trust and Self-Disclosure Online	Adam N. Joinson, Ulf-Dietrich Reips, Tom Buchanan, and Carina B. Paine Schofield	2010	UK	Psychology	759	No specific target demographic	Quantitative	Regression
Risky Disclosures on Facebook: The Effect of Having a Bad Experience on Online Behavior	Emily Christofides, Amy Muise and Serge Desmarais	2012	Canada	Psychology	256	Facebook users between 12 and 18 years of age	Qualitative, Quantitative	Correlation

Saving Face on Facebook: Privacy Concerns, Social Benefits, and Impression Management	David W. Wilson, Jeffrey G. Proudfoot, and Joseph S. Valacich	2014	USA	Information Science	244	College students	Quantitative	Correlation
Silent Listeners: The Evolution of Privacy and Disclosure on Facebook	Fred Stutzman, Ralph Gross, and Alessandro Acquisti	2012	USA	Information Science, Behaviour	5076	College students	Quantitative	Correlation
Student Awareness of the Privacy Implications When Using Facebook	Tabreez Govani and Harriet Pashley	2005	USA	Policy, Law, Technology	50	College students	Quantitative	Descriptive only
The Influence of user affect in online information disclosure	Robin Wakefield	2013	USA	Information Science	301	No specific target demographic	Quantitative	Correlation

Privacy and Trust: The role of situational and dispositional variables in online disclosure	Adam N. Joinson, Carina Paine, Ulf-Dietrich Reips, and Tom Buchanan	2006	USA	Psychology	759	College students	Quantitative	Regression
User's Awareness of Privacy on Online Social Networking sites - Case Facebook	Virpi Kristiina Tuunainen, Olli Pitkänen, and Marjaana Hovi	2009	Finland	Information Science, Psychology	210	Facebook users	Quantitative	Descriptive only
Privacy Indexes: A Survey of Westin's Studies	Ponnurangam Kumaraguru and Lorrie Faith Cranor	2005	USA	Psychology	-	-	Qualitative	-

Who am I? Representing the self offline and in different online contexts	Lia Emanuel, Greg J. Neil, Chris Bevan, Danaë Stanton Fraser, Sarah V. Stevenage, Monica T. Whitty, and Sue Jamison-Powell	2014	UK	Psychology	148	College students	Quantitative	Correlation
Information Disclosure and Control on Facebook: Are They Two Sides of the Same Coin or Two Different Processes?	Emily Christofides , Amy Muise and Serge Desmarais	2009	Canada	Psychology	343	College students	Quantitative	Correlation, Multivariate Regression
All about me: Disclosure in online social networking profiles: The case of FACEBOOK	Amanda Nosko, Eileen Wood and Seija Molema	2009	Canada	Information Science, Behaviour	400	Facebook users	Quantitative	Correlation

Appendix B: Facebook settings related to privacy

Location of setting	**Setting**	**Options**
Settings, Privacy Settings and Tools	Who can see your future posts?	Public, Friends, Only me, Custom group
	Who can send you friend requests?	Everyone, Friends of friends
	Whose messages do I want filtered into my inbox?	Basic Filtering, Strict Filtering
	Who can look you up using the email address you provided?	Everyone, Friends of friends, Friends
	Who can look you up using the phone number you provided?	Everyone, Friends of friends, Friends
	Do you want other search engines to link to your timeline?	Yes, No
Settings, Timeline and Tagging	Who can post on your timeline?	Friends, Only me
	Review posts that friends tag you in before they appear on your timeline?	Enabled, Disabled
	Who can see posts you've been tagged in on your timeline?	Everyone, Friends of friends, Friends, Only Me, Custom group
	Who can see what others post on your timeline?	Everyone, Friends of friends, Friends, Only Me, Custom group
	Review tags people add to your own posts before the tags appear on Facebook?	Enables, Disabled
	When you're tagged in a post, who do you want to add to the audience if they aren't already in it?	Friends, Only me, Custom group
	Who sees tag suggestions when photos that look like you are uploaded?[3]	Unavailable[3]
User Profile, Friends, Manage, Edit privacy	Who can see your friend list?	Public, Friends, Only me, Custom group
	Who can see the people and lists you follow?	Public, Friends, Only me, Custom group
User Profile, About, Work and Education	Work (current)[1]	Public, Friends, Only me, Custom group

Location of setting	**Setting**	**Options**
	Work (past)[2]	Public, Friends, Only me, Custom group
	Professional Skills[2]	Public, Friends, Only me, Custom group
User Profile, About, Work and Education (continued)	University (Most recent)[1]	Public, Friends, Only me, Custom group
	University (Older)[2]	Public, Friends, Only me, Custom group
	High School	Public, Friends, Only me, Custom group
User Profile, About, Places You've Lived	Current City	Public, Friends, Only me, Custom group
	Home Town	Public, Friends, Only me, Custom group
	Other Places Lived[2]	Public, Friends, Only me, Custom group
User Profile, About, Contact and Basic Info	E-mail[2]	Public, Friends, Only me, Custom group
	E-mail[2] (Shown on timeline)	Shown, Hidden
	Date of Birth	Public, Friends, Only me, Custom group
	Year of Birth	Public, Friends, Only me, Custom group
	Gender	Show on my Timeline, Don't Show on my Timeline
	Languages	Public, Friends, Only me, Custom group
	Interested in	Public, Friends, Only me, Custom group
	Religious Views	Public, Friends, Only me, Custom group
	Political Views	Public, Friends, Only me, Custom group
User Profile, About, Family and Relationships	Relationship status	Public, Friends, Only me, Custom group
	Family Members[2]	Public, Friends, Only me, Custom group
User Profile, About, Details About You	About You	Public, Friends, Only me, Custom group
	Favourite Quotes	Public, Friends, Only me, Custom group

Location of setting	Setting	Options
User Profile, About, Likes, Edit Privacy	Movies	Public, Friends, Only me, Custom group
	Television	Public, Friends, Only me, Custom group
User Profile, About, Likes, Edit Privacy (continued) User Profile, About, Likes, Edit Privacy	Music	Public, Friends, Only me, Custom group
	Books	Public, Friends, Only me, Custom group
	Sports teams	Public, Friends, Only me, Custom group
	Athletes	Public, Friends, Only me, Custom group
	Inspirational People	Public, Friends, Only me, Custom group
	Restaurants	Public, Friends, Only me, Custom group
	Games	Public, Friends, Only me, Custom group
	Activities	Public, Friends, Only me, Custom group
	Interests	Public, Friends, Only me, Custom group
	Sports	Public, Friends, Only me, Custom group
	Foods	Public, Friends, Only me, Custom group
	Clothing	Public, Friends, Only me, Custom group
	Websites	Public, Friends, Only me, Custom group
	Other	Public, Friends, Only me, Custom group
User Profile, Friends, Edit Privacy	Who can see your friend list?	Public, Friends, Only me, Custom group
	Who can see the people and lists you follow?	Public, Friends, Only me, Custom group

[1] Multiple options may be entered, this option is asked separately from the rest.

[2] Multiple options may be entered with separate settings for each.

[3] At time of writing, this setting is visible, but not yet available to the user.

Appendix C: Complete list of questions used in the survey

All questions used in the survey can be found in the table below. Please note that, if the respondent chose to use the tool, there was an additional question at the start of the survey, asking for the respondent ID.

Question	Options, if provided
Please enter your age	
Please select your gender	Female, Male
Please select the highest level of education you have completed.	None, Primary School, Secondary/high school, Trade school or vocational training, University (Bachelor), University (Master), Doctorate degree
Please select your nationality.	Afghan, Albanian, Algerian, American, Angolan, Argentinian, Australian, Austrian, Bangladeshi, Belarusian, Belgian, Bolivian, Botswanan, Brazilian, British, Bulgarian, Burmese, Cambodian, Canadian, Chilean, Chinese, Colombian, Congolese, Croatian, Cuban, Cypriot, Czech, Danish, Dutch, Ecuadorian, Egyptian, Estonian, Ethiopian, Fijian, Filipino, Finnish, French, Georgian, German, Ghanaian, Greek, Guatemalan, Hungarian, Icelandic, Indian, Indonesian, Iranian, Iraqi, Irish, Israeli, Italian, Ivorian, Jamaican, Japanese, Jordanian, Kazakhstani, Kenyan, Kuwaiti, Laotian, Latvian, Lebanese, Libyan, Lithuanian, Malagasy, Malaysian, Mexican, Mongolian, Moroccan, Mozambican, Namibian,

	Nepalese, New Zealand, Nigerian, North Korean, Norwegian, Pakistani, Palestinian, Paraguayan, Peruvian, Polish, Portuguese, Romanian, Russian, Saudi Arabian, Serbian, Singaporean, Slovak, Slovenian, Somali, South African, South Korean, Spanish, Sri Lankan, Sudanese, Swedish, Swiss, Syrian, Taiwanese, Tanzanian, Thai, Tunisian, Turkish, Ugandan, Ukrainian, Uruguayan, Venezuelan, Vietnamese, Yemeni, Zambian, Zimbabwean
Privacy concern section	
Below, you will see several statements. For each, please indicate to what degree they apply to you. [In general, how concerned are you about your privacy while you are using the internet?]	Not at all, Slightly, Moderately, Much, Very much
Below, you will see several statements. For each, please indicate to what degree they apply to you. [Are you concerned about online organisations not being who they claim they are?]	Not at all, Slightly, Moderately, Much, Very much
Below, you will see several statements. For each, please indicate to what degree they apply to you. [Are you concerned about online identity theft?]	Not at all, Slightly, Moderately, Much, Very much
Below, you will see several statements. For each, please indicate to what degree they apply to you. [Are you concerned about people online not being who they say they are?]	Not at all, Slightly, Moderately, Much, Very much
Below, you will see several statements. For each, please indicate to what degree they apply to you. [Are you concerned about people you do not know obtaining personal information about you from your online activities?]	Not at all, Slightly, Moderately, Much, Very much
Below, you will see several statements. For each, please indicate to what degree they apply to you. [	Not at all, Slightly, Moderately, Much, Very

Are you concerned that if you use your credit card to buy something on the internet your credit card number will obtained/intercepted by someone else?]	much
Below, you will see several statements. For each, please indicate to what degree they apply to you. [Are you concerned that if you use your credit card to buy something on the internet your card will be mischarged?]	Not at all, Slightly, Moderately, Much, Very much
Below, you will see several statements. For each, please indicate to what degree they apply to you. [Are you concerned that an email you send someone may be inappropriately forwarded to others?]	Not at all, Slightly, Moderately, Much, Very much
Knowledge of privacy measures section	
For each of the 8 questions below, please indicate the correct answer. If you do not know which is correct, please select that option instead. [On Facebook, I can set who can add me to groups.]	True, False, I don't know
For each of the 8 questions below, please indicate the correct answer. If you do not know which is correct, please select that option instead. [On Facebook, I can change the visibility of older posts.]	True, False, I don't know
For each of the 8 questions below, please indicate the correct answer. If you do not know which is correct, please select that option instead. [On Facebook, I can set the visibility of each of my schools/universities I have attended separately.]	True, False, I don't know
For each of the 8 questions below, please indicate the correct answer. If you do not know which is correct, please select that option instead. [On Facebook, I can set the default group to which my future status updates are visible.]	True, False, I don't know
For each of the 8 questions below, please indicate the correct answer. If you do not know which is correct, please select that option instead. [On Facebook, I can set who can tag me in pictures.]	True, False, I don't know
For each of the 8 questions below, please indicate the correct answer. If you do not know which is correct, please select that option instead. [On Facebook, I can make a profile without needing to enter my date of birth.]	True, False, I don't know
For each of the 8 questions below, please indicate the	True, False, I don't know

correct answer. If you do not know which is correct, please select that option instead. [On Facebook, I can change settings so only members of my family can see what music I like.]	
For each of the 8 questions below, please indicate the correct answer. If you do not know which is correct, please select that option instead. [I can allow Facebook to detect and share my current location when posting a status update.]	True, False, I don't know
Online disclosure and use of privacy measures section	
Who can see your future posts?	Public, Friends, Only me, Custom group
Who can send you friend requests?	Everyone, Friends of friends
Whose messages do I want filtered into my inbox?	Basic Filtering, Strict Filtering
Who can look you up using the email address you provided?	Everyone, Friends of friends, Friends
Who can look you up using the phone number you provided?	Everyone, Friends of friends, Friends
Do you want other search engines to link to your timeline?	Yes, No
Who can post on your timeline?	Friends, Only me
Review posts that friends tag you in before they appear on your timeline?	Enabled, Disabled
Who can see posts you've been tagged in on your timeline?	Everyone, Friends of friends, Friends, Only Me, Custom group
Who can see what others post on your timeline?	Everyone, Friends of friends, Friends, Only Me, Custom group
Review tags people add to your own posts before the tags appear on Facebook?	Enables, Disabled
When you're tagged in a post, who do you want to add to the audience if they aren't already in it?	Friends, Only me, Custom group
Who can see your friend list?	Public, Friends, Only me, Custom group
Who can see the people and lists you follow?	Public, Friends, Only me, Custom group
Please select "Work and Education" in the left-hand column and indicate the options selected for the	Public, Friends, Only me, Custom group, Did not fill in

privacy setting of each item. [Work]	this option, Does not apply to me
Please select "Work and Education" in the left-hand column and indicate the options selected for the privacy setting of each item. [Professional skills]	Public, Friends, Only me, Custom group, Did not fill in this option, Does not apply to me
Please select "Work and Education" in the left-hand column and indicate the options selected for the privacy setting of each item. [University]	Public, Friends, Only me, Custom group, Did not fill in this option, Does not apply to me
Please select "Work and Education" in the left-hand column and indicate the options selected for the privacy setting of each item. [High School]	Public, Friends, Only me, Custom group, Did not fill in this option, Does not apply to me
Please select "Places you've lived" in the left-hand column and indicate the options selected for the privacy setting of each item. [Current City]	Public, Friends, Only me, Custom group, Did not fill in this option, Does not apply to me
Please select "Places you've lived" in the left-hand column and indicate the options selected for the privacy setting of each item. [Home Town]	Public, Friends, Only me, Custom group, Did not fill in this option, Does not apply to me
If you have entered any other places you have lived, please click on "options" on the right-hand side of this location, and select "edit". The privacy setting can be seen at the bottom of this window. [Other Places Lived]	Public, Friends, Only me, Custom group, Did not fill in this option, Does not apply to me
Please select "Contact and Basic Info" in the left-hand column and indicate the options selected for the privacy setting of each item. [Mobile Phone]	Public, Friends, Only me, Custom group, Did not fill in this option
Please select "Contact and Basic Info" in the left-hand column and indicate the options selected for the privacy setting of each item. [Other account(s)]	Public, Friends, Only me, Custom group, Did not fill in this option
Please select "Contact and Basic Info" in the left-hand column and indicate the options selected for the privacy setting of each item. [Website]	Public, Friends, Only me, Custom group, Did not fill in this option
Please select "Contact and Basic Info" in the left-hand column and indicate the options selected for the privacy setting of each item. [E-mail]	Public, Friends, Only me, Custom group, Did not fill in this option
Please select "Contact and Basic Info" in the left-hand column and indicate the options selected for the privacy setting of each item. [Date of Birth]	Public, Friends, Only me, Custom group, Did not fill in this option
Please select "Contact and Basic Info" in the left-hand	Public, Friends, Only me,

column and indicate the options selected for the privacy setting of each item. [Year of Birth]	Custom group, Did not fill in this option
Please select "Contact and Basic Info" in the left-hand column and indicate the options selected for the privacy setting of each item. [Interested In]	Public, Friends, Only me, Custom group, Did not fill in this option
Please select "Contact and Basic Info" in the left-hand column and indicate the options selected for the privacy setting of each item. [Languages]	Public, Friends, Only me, Custom group, Did not fill in this option
Please select "Contact and Basic Info" in the left-hand column and indicate the options selected for the privacy setting of each item. [Religious Views]	Public, Friends, Only me, Custom group, Did not fill in this option
Please select "Contact and Basic Info" in the left-hand column and indicate the options selected for the privacy setting of each item. [Political Views]	Public, Friends, Only me, Custom group, Did not fill in this option
Is your email address shown on timeline?	Shown, Hidden
Is your gender shown on timeline?	Yes, No
Please select "Family and Relationships" in the left-hand column and indicate the options selected for the privacy setting of each item. [Relationship Status]	Public, Friends, Only me, Custom group, Did not fill in this option, Does not apply to me
Please select "Family and Relationships" in the left-hand column and indicate the options selected for the privacy setting of each item. [Family members]	Public, Friends, Only me, Custom group, Did not fill in this option, Does not apply to me
On the right-hand side, click the pencil labelled "Manage", and select "Edit privacy". Please fill in the most restrictive privacy option set for your Likes. [Like pages]	Public, Friends, Only me, Custom group, Did not fill in this option, I have not liked any pages
For each of the categories of pages below, please indicate whether you have liked any. You can find this information by switching between the categories at the top of this block. [Films/Movies]	Yes, No
For each of the categories of pages below, please indicate whether you have liked any. You can find this information by switching between the categories at the top of this block. [TV Programmes]	Yes, No
For each of the categories of pages below, please indicate whether you have liked any. You can find this information by switching between the categories at	Yes, No

the top of this block. [Music]	
For each of the categories of pages below, please indicate whether you have liked any. You can find this information by switching between the categories at the top of this block. [Books]	Yes, No
For each of the categories of pages below, please indicate whether you have liked any. You can find this information by switching between the categories at the top of this block. [Sports teams]	Yes, No
For each of the categories of pages below, please indicate whether you have liked any. You can find this information by switching between the categories at the top of this block. [Athletes/Sportspeople]	Yes, No
For each of the categories of pages below, please indicate whether you have liked any. You can find this information by switching between the categories at the top of this block. [People]	Yes, No
For each of the categories of pages below, please indicate whether you have liked any. You can find this information by switching between the categories at the top of this block. [Restaurants]	Yes, No
For each of the categories of pages below, please indicate whether you have liked any. You can find this information by switching between the categories at the top of this block. [Apps and Games]	Yes, No

Appendix D: Code for data collection tool

NB: The code displayed below uses the Facebook PHP API, version 4.0. This code was meant to both collect the data required, as well as to provide the respondent with the login screen and the link to the survey. The code includes comments for each of the steps taken. These comments can be found below in the // and / mark up standard to PHP.*

```
<?php
session_start();

//Set up all elements of Facebook PHP SDK 4.0 to be used.
require_once 'facebook-php-sdk-v4-4.0-dev/autoload.php';
use Facebook\FacebookSession;
use Facebook\FacebookRequest;
use Facebook\GraphUser;
use Facebook\GraphObject;
use Facebook\FacebookRequestException;
use Facebook\FacebookRedirectLoginHelper;

//Indicate application and corresponding secret to use in
communication with Facebook.
FacebookSession::setDefaultApplication(/*application*/,/*
appsecret*/);

//Set up login helper
$helper                          =                          new
FacebookRedirectLoginHelper('http://www.disclosurestudy.c
om/t/index.php');

//Check if user is logged in and has granted permissions.
try {
  $session = $helper->getSessionFromRedirect();
} catch(FacebookRequestException $ex) {
  echo 1; // When Facebook returns an error
} catch(\Exception $ex) {
  echo 2;// When validation fails or other local issues
}

//If logged in and permissions granted
if ($session) {
     try {
       $disclosed = array();

       //Get profile root items
       $me = (new FacebookRequest(
          $session, 'GET', '/me'
       ))->execute()-
>getGraphObject(GraphUser::className());

       //Check if e-mail and last name have been set.
       $pr = $me->getProperty('email');
       if (isset($pr)) {
          //email set
          array_push($disclosed, 'email');
       }
        $pr = $me->getProperty('last_name');
       if (isset($pr)) {
          //last name set
          array_push($disclosed, 'last_name');
       }

       //Check if value of verified field is true
        $pr = $me->getProperty('verified');
       if ($pr) {
          //verified is true
          array_push($disclosed, 'verified');
       }


      //Convert into binary array
       $toFind       =    array('last_name',    'email',
'verified');
       $found = array();
       for ($i=0 ; $i < count($toFind); $i++) {
          if (in_array($toFind[$i], $disclosed)) {
               array_push($found, 1);
          }
          else {
               array_push($found, 0);
          }
       }

       //Get respondent ID
       $current                                       =
file_get_contents('../../private/last.txt');
       file_put_contents('../../private/last.txt',
($current+1));
       //Set following respondent ID
       $id = ($current * 3) + 1;


       //Add respondent ID to beginning of found array
       array_unshift($found, $id);

       //Save binary results
       $fp  =  fopen('../../private/results/'  .  $id  .
'.csv', 'w');
      fputcsv($fp, $found, ";");
```

```
       fclose($fp);

       //Save original results
       $fp  =  fopen('../../private/results/'  .  $id  .  '-
org.csv', 'w');
      fputcsv($fp, $disclosed, ";");
       fclose($fp);


       //Load results page
       include('top.php');
       echo  'Thank  you  for  participanting  in  the  study.
The information was succesfully stored.';
       echo 'For the second part of the survey, you will
need your respondent ID. Please write it down and provide
it in the survey when prompted.';
       echo '<br/><br/>';
       echo 'Your respondent ID is: ' . $id;
       echo '<br/><br/>';
       echo                      '                      <a
href="https://docs.google.com/forms/d/1j3W60yTr0CuM9r4MK0
JL_FrdFGEK8RMzsW-AteO1WCM/viewform?usp=send_form"
target="_blank">Click here to continue to the second part
of the survey.</a>';
       echo '<br/><br/>';
       echo 'In order to verify the information collected
during this research, the tool at the following items of
your profile to see whether the information in question
had been shared:';
       echo '<ul>';
       foreach ($toFind as $d) {
          echo '<li>';
          print($d);
          echo '</li>';
       }
       echo '</ul>';
       include('bottom.php');



       //Catch exceptions when using Graph API
     } catch (FacebookRequestException $e) {
          echo 5;
       // The Graph API returned an error
     } catch (\Exception $e) {
       // Some other error occurred
       echo 6;
     }//}
}

//If user is not logged in or has not granted permissions
else {
     //Set permissions required
     $permissions = array('email');

     //Load login page
     include('top.php');
     echo 'For the first part of the survey, we kindly
ask that you log in on Facebook using the link below.
After this, you will be asked if you want to share the
information requested by the tool. Please be patient
after accepting, as your information may take some time
to process.';
     echo '<br/><br/>';
     echo 'If you see this screen for a second time,
please click on the link below again to continue.';
     echo '<br/><br/>';
     echo '<a href="' . $helper-
>getLoginUrl($permissions) . '">To contintue, please
login with Facebook</a>';
     include('bottom.php');
}

?>
```