

To Know You is to Love You

The Effect of Survey Length and Personally Knowing the Experimenter on Response Rate and Response Quality

Abstract

This research investigated how the length of a survey and personally knowing the experimenter influence the response rate and response quality of voluntary web-based surveys. An experiment was designed in which the respondents were randomly allocated to either a short or a long version of a similar survey. In addition, all respondents were categorised as personally knowing the experimenter or not. By means of Fisher's exact tests, Pearson's chi-squared tests, Mann-Whitney U tests and a logistic regression, the effects of survey length and familiarity were analysed. Results show that, as expected, there are significant differences in response rates as a result of varying survey length and varying familiarity conditions. In addition, dropout rate, a measure of response quality was also found to be significantly affected. A logistic regression of dropout rate has shown that in this sample, participating in a short survey and personally knowing the experimenter make it less likely to drop out of the survey. Furthermore, respondents who started this survey three days after the initial invitation are also less likely to drop out of the survey than respondents who started the survey on the day of the invitation. For the other two measures of response quality, uncertain responses and variability to questions in grids, no significant differences were observed as a result of varying survey length and varying familiarity conditions.

Master Thesis Behavioural Economics
Samare Huls – 367264
Erasmus School of Economics, Erasmus University
Supervisor: Dr. J.T.R. Stoop
10-08-2017

Content

Content	3
1. Introduction.....	5
2. Theoretical Framework.....	7
2.1 Survey Length.....	7
2.1.1 Survey Length and Response Rate	7
2.1.2 Survey Length and Response Quality.....	8
2.2 Familiarity.....	11
2.2.1 Familiarity and Response Rate	12
2.2.2 Familiarity and Response Quality.....	13
3. Data and Methodology	15
3.1 Experimental Design	15
3.2 Variable Description.....	16
3.2.1 Version	16
3.2.1 Familiarity.....	16
3.2.3 Response Rate	17
3.2.4 Response Quality.....	17
3.2.5 Control Variables	18
3.3 Descriptive Statistics	19
3.4 Methodology	22
3.4.1 Response Rate	22
3.4.2 Response Quality.....	22
4. Results.....	24
4.1 Survey Length.....	24
4.1.1 Survey Length and Response Rate	24
4.1.2 Survey Length and Response Quality.....	24
4.2 Familiarity.....	27
4.2.1 Familiarity and Response Rate	27
4.2.2 Familiarity and Response Quality.....	28
4.3 Extra Analysis Dropout Rate including Control Variables.....	30
5. Conclusion and Discussion.....	33
Bibliography.....	37
Appendix A: Survey Design.....	40
A.1 Email invitation	40
A.2 Introductory Page	40

A.3 Long Survey	41
A.3.1 Page 1: Egoism Scale.....	41
A.3.2 Page 2: Uncertain responses.....	42
A.3.3 Page 3: Altruism Scale	43
A.3.4 Page 4: Mathematical Question	44
A.3.5 Page 5: Demographic Variables.....	44
A.3.6 Page 6: Relationship with experimenter	44
A.4 Short Survey	45
A.5 Reminder Email.....	45
Appendix B: Robustness Checks	46
B.1 Test for Match Familiarity Measures.....	46
B.2 Tests for Demographic Differences between Versions and Conditions.....	46

1. Introduction

Throughout academic education, there is much focus on academic research. It is stressed that samples should be representative to be able to get results that are both statistically and economically relevant. This means that a careful selection of respondents and a high response rate is essential (Cook, Heath, & Thompson, 2000). In addition, motivation is essential to obtain good response rates and response quality (Herzog & Bachman, 1981). Earlier research on web-based surveys has shown that important factors for response rate and response quality of academic experiments are the length and the design of the survey, the type of monetary incentives, the saliency of the topic and the use of reminders (Deutskens et al., 2004; Sheehan & McMillan, 1999; Singer & Ye, 2013; Vicente & Reis, 2010). However, for academic research conducted by students the budget is generally very limited; students cannot provide monetary incentives and the experimental design cannot not be too costly. This means that student research often relies on voluntary participation in web-based surveys. For many individuals participation that is voluntary is associated with high burden because it costs time and effort (DeMaio, 1980). Therefore, participation comes at a cost for the respondent while it comes as a benefit for the experimenter. Since altruism is defined as “disinterested and selfless concern for the well-being of others” (Oxford Dictionaries, 2017), voluntary participation in web-based surveys is considered to be an act of altruism. This research aims to find which variables are of influence to this type of altruism.

The first factor that is analysed in this research is the length of a survey. Research on experienced burden has shown that longer surveys are perceived as more costly to respondents (Bradburn, 1978). If respondents become bored or fatigued, they are less engaged, make more mistakes and the error variance increases (Hess, Hensher, & Daly, 2012). Therefore, the time required to participate in a survey might influence response rate and response quality. Secondly, familiarity to the experimenter is analysed as well. Since participation in academic experiments by students is voluntary, for the experimenter it is easiest to ask familiar individuals to participate. However, research on altruism has shown that individuals behave differently when they personally know, like or can identify with the other person than when they do not (Cialdini et al., 1997). In general, individuals are more empathetic and altruistic when they are familiar with the other individual (Krebs, 1975). Therefore, it is of interest to also analyse how familiarity to the experimenter affects individuals in academic experiments. Survey length as well as familiarity seems to affect an individual’s choice of whether to participate in an experiment. Consequently, survey length and familiarity might be very relevant in respondents’ effort and performance. Therefore the following question is researched:

How are the response rate and response quality in voluntary web-based surveys influenced by the length of the survey and by personally knowing the experimenter?

To answer the research question, an experiment is designed. A random sample of members of Dutch student associations is used in which individuals are invited to participate in a voluntary web-based survey. All respondents are randomly allocated to either a short or a long version of a similar survey. In addition, all respondents are categorised as personally knowing the experimenter or not.

As the research question states, the purpose of this research is to establish how survey length and familiarity affect the number and quality of responses in academic experiments. The relevance of this research is thereby threefold. First of all, any response rate lower than 100% can give problems with statistical inference because of response bias (Groves, Cialdini, & Couper, 1992). Understanding how survey length and familiarity affect survey responses can be beneficial in increasing the response rate and quality of responses in future voluntary web-based surveys and thus in decreasing response bias. Secondly, while the effect of length of a survey has already been extensively analysed in combination with many effects, the scientific relevance of this research lies in analysing the effects of both survey length and familiarity on survey response. Thirdly, much research has been conducted on variables that influence incentivised participation. However, less is known about the variables that influence voluntary participation, an act of altruism. The research hereby aims to establish how survey length and familiarity affect different degrees of altruism, from initial helping behaviour to exerting much effort in helping.

The remainder of this thesis is structured as follows. First of all, relevant literature about web-based surveys and about altruism is reviewed for the purpose of hypotheses development. Secondly, the section Data and Methodology describes the experimental design and which statistical methods are used to test the hypotheses. Next, the results are presented. In the last section, the research is concluded and critically discussed.

2. Theoretical Framework

In this section, an overview and the definitions of the relevant concepts of this research are provided. Furthermore, by means of a literature review these concepts are connected to each other in order to provide a well-established theoretical framework.

2.1 Survey Length

There are many ways in which time can influence individuals in academic experiments. For example, by the expected time that the experiment takes, how much time the respondent has at that moment or the timing of the survey request. However for matters of measurability, in this research the expected length of a survey is analysed. Several variables that might be influenced by the length of a survey are discussed now.

2.1.1 Survey Length and Response Rate

One variable that has been stated to be largely affected by the length of a survey is the response rate, how many of the invited individuals start the survey. According to Bradburn (1978) participating in a longer survey can be perceived as more costly for the respondents. Therefore, participation in a longer experiment is associated with a higher degree of altruism than participation in a shorter experiment.

Crawford, Couper and Lamias (2001) researched how the length of a survey, reminders, the type of access to the survey and a progress indicator affect the response rate and the response quality of web-based surveys. They sent invitations to complete a web-based survey via email. Subjects were incentivised with a lottery incentive. The authors misrepresented the survey length to analyse its effects on the response rate. In one version of the survey, respondents were told that the survey would take 8 to 10 minutes to complete. In the other version, respondents were told that the survey would take about 20 minutes to complete. The actual length of the survey was somewhere in between. The results show that the response rate for the 'short' survey was 36.6%, and the response rate for the 'long' survey was 32.5% indicating that there is a negative relationship between the length of the survey and the response rate.

Similar research was conducted by Marcus et al. (2007). The authors analysed how the saliency of the topic and the length of the survey interact with different types of incentives. In this survey, a lottery incentive was used as a material incentive, no incentives were provided in the control group. In this research respondents also received an invitation by email. Before the actual invitation was sent, respondents received an email in which they were informed with the research purpose, the length of the survey and (if applicable) the incentives. The short survey consisted of 91 items and was expected to take 10 to 20 minutes to complete. The long survey consisted of 359

items and was expected to take 30 to 60 minutes to complete. Ten days later, the actual invitation to participate was sent to respondents. The authors found that the response rate was 30.8% for the short survey and 18.6% for the long survey.

Galesic and Bosnjak (2009) also researched the effect of survey length on response rate and response quality in a web-based survey. The researchers invited subjects by banner advertisements that shortly described the survey topic. There was one control group without incentives. Respondents were randomly divided into three different conditions: 10, 20 or 30 minutes. The length of the survey was mentioned on the introductory page of the survey. Results show that the shorter the stated length, the more respondents start the survey.

Another research by Deutskens et al. (2004) analysed the effect of the timing of the request and number of follow-up requests, the type of incentives, the length and the presentation of the survey on response rate and response quality. In an online survey that was conducted yearly since 1991, respondents were tested for their attitude of brands in different product categories. They were invited to participate via an email containing a brief introduction of the research, a request to participate and a hyperlink to the online survey. The short version of the survey contained 9 product categories while the long survey contained 19 product categories. The authors do not mention whether and how respondents were informed about the length of the survey. The results show that the short survey had a significantly higher response rate (24.5%) than the long survey (17.1%).

While these four researches show that the length of a survey negatively influences the response rate, there are also researches that find no significant relationship between survey length and response rate (Cook, Heath, & Thompson, 2000; Heerwegh & Loosveldt, 2006; Sheehan, 2001). However, in line with a large body of evidence, this research hypothesises the following:

H1: A longer survey leads to a different response rate than a shorter survey.

2.1.2 Survey Length and Response Quality

Besides the response rate, the length of a survey might also influence the respondents' effort during the experiment (Bradburn, 1978). As a consequence, it might also influence the quality of responses. Intuitively there are two ways in which the length of a survey can influence effort and therefore response quality. On one hand, it can be argued that effort is costly. Participating in a short survey only requires a little amount of effort (Bradburn, 1978). Participating in a longer survey requires a lot of effort so effort depletion is likely to take place (Krosnick, 1991; Galesic, 2006). This means that the longer a survey, the less willing individuals are to put a lot of effort in the survey. On the other hand it can be argued that the individuals who are willing to participate

in a long survey are more altruistic and more committed. Therefore, they might also exert more effort during the survey (Yan et al, 2010). Following this line of reasoning, in longer surveys effort depletion is little and the quality of responses does not change significantly compared to shorter surveys. To establish in which way the length of a survey influences response quality, earlier research on different measures of response quality is reviewed.

2.1.2.1 Survey Length and Dropout Rate

One measure of quality in a web-based survey is the amount of people that drop out of the survey after they have started it. In the earlier mentioned research Crawford, Couper and Lamias (2001) analysed how the length of a survey influenced dropout rate. The authors hypothesised that if the survey would be significantly longer than was mentioned, the dropout rate would increase. A Chi-squared test confirmed their expectations. The group of respondents that was told they would take a long survey, had a dropout rate of 9% while the group of respondents that was told they would take a short survey, had a dropout rate of 11.3%. The authors attribute the difference in dropout rate to the expected length of the survey. However, because the true length of the survey was misrepresented, the dropout rate could also be due to the fact that respondents felt they were deceived. Therefore, the result might not be a true estimate of the effect of the length of the survey on the dropout rate.

Research by Yan et al. (2010) analysed how progress indicators, stated length and actual length influence the completion rate of web surveys. The authors designed four different treatments concerning the length of the survey. Two treatment groups had to complete a long survey of 155 questions lasting approximately 25 minutes. One of these groups was told that the survey would take 40 minutes, the other group was told that the survey would take 10 minutes. The other two treatment groups had to complete a short survey of 101 questions lasting approximately 16 minutes. One of these groups was told that the survey would take 25 minutes while the other group was told that it would take 5 minutes. For the effect of actual length the authors hypothesised a lower dropout rate in the short version of the survey than in the long version of the survey. The results confirm this; the dropout rates were significantly different (i.e. 10.6% and 16.8% respectively). In addition, in contrast to what was hypothesised, the authors also found that respondents that expected a short survey (i.e. 5 minutes) had a lower dropout rate than respondents that expected a long survey (i.e. 25 minutes). However, this effect of stated length was only found for respondents in the short survey condition.

Similar research was done by Marcus et al, (2007) and Deutskens et al. (2004). However, both researches used only two conditions of survey length and they only presented the actual expected time to complete the survey. By using a Chi-squared test, they analysed whether there was a significant difference between the number of dropouts in the long and short version. The

authors found similar results as Crawford, Couper and Lamias (2001) and Yan et al. (2010). The shorter the length of the survey, the lower the dropout rate.

Hoerger (2010) used a slightly different approach to test whether the length of a survey affects the dropout rate. The author examined six different web-based surveys about personality and mood. The surveys varied in length from 243 to 543 survey items, participation in all six surveys was voluntary. A Kaplan-Meier estimator showed that 10% of the respondents drops out within the first 12 items. The dropout rate increased to a cumulative 13.2% and 20.7% after 100 and 500 items respectively. The results therefore show that shorter a survey causes the dropout rate to be lower. Since in all these researches dropout is expected to vary with the length of a survey, the following is hypothesised for this research:

H2: A longer survey leads to a different dropout rate than a shorter survey.

2.1.2.2 Survey Length and Uncertain Responses

A second measure of effort and therefore of response quality is the degree to which uncertain responses are expressed, that is reporting “don’t know” or “no opinion”. Krosnick (1991) reviewed the psychological mechanisms associated with response behaviour. It is argued that respondents sometimes give a satisfactory answer to a survey question if answering it otherwise requires too much cognitive effort. This behaviour is called ‘satisficing’ and can occur in many forms. One form Krosnick identified as a strong form of satisficing is reporting “don’t know” rather than reporting an opinion. Reporting an uncertain response does not require retrieval or judgement and therefore it hardly requires cognitive effort. This makes it a good response strategy if cognitive effort would otherwise be high.

Deutskens et al. (2004) experimentally examined this relationship between the length of a survey and the number of uncertain responses. The authors found that there are relatively more uncertain responses in the long version of their survey than in the short version of their survey. In accordance with these sources, the following is hypothesised:

H3: A longer survey leads to a different number of uncertain responses than a shorter survey.

2.1.2.3 Survey Length and Variability of Responses

A third measure of quality of responses in a web-based survey is the variability of responses to questions in grids. Krosnick (1991) argued that a limited variability of responses is a strong form of the beforementioned ‘satisficing’. It occurs when respondents have to rate different items on a common scale, i.e. multiple questions with the same response alternatives. Similar to giving

uncertain responses, this type of response behaviour costs less cognitive effort than optimal response behaviour. Therefore, it is likely to increase with the length of a survey.

Herzog and Bachman (1981) empirically analysed the response quality of a survey among high school seniors. One way in which the authors measure response quality is in terms of straight-line responding, a type of position bias in which a respondent uses an identical response for all questions in a grid. There were five versions of a survey all lasting 45 minutes but covering a different topic. In addition, there was one longer version of more than 120 minutes in which all topics were covered. Respondents in the long survey were motivated with a \$5 payment and with released time from class to participate in the survey. Respondents in the short surveys did not receive these extrinsic motivations. The authors identified that the amount of straight-line responses increased in later parts of the survey. Similar to Krosnick (1991) it is argued that this type of response behaviour is associated with decreased motivation and is therefore likely to occur near the end of a survey and when a survey increases in length.

Galesic and Bosnjak (2009) also analysed the effect of survey length on the variability of responses to questions in grids. Similar to the other researches, the authors found that the variability of responses decreases near the end of a survey. This behaviour occurred in both the version where the length of the survey was overstated and in the version where it was understated. This suggests that the variability of responses is affected by the actual length of the survey rather than the stated length. Based on these three researches, the length of a survey is expected to increase burden and therefore to change the variability of responses to questions in grids:

H4: A longer survey leads to a different variability of responses to questions in grids than a shorter survey.

2.2 Familiarity

Besides survey length, familiarity can also be an important factor of influence in academic research by students. As was mentioned before, due to the lack of money available to support students' academic research, voluntary participation in web-based surveys is common. Therefore, the experimenter often asks familiar individuals to participate. However, intuitively it can be questioned whether familiar individuals put the same amount of effort into the survey and answer as truthfully as unfamiliar individuals. If there are discrepancies between the responses of familiar and unfamiliar individuals, the results of the research may be biased. Therefore, this research analyses whether familiarity affects response rate and response quality. For matters of comparability, similar hypotheses as in Section 2.1 are tested. However, the literature about the effect of familiarity on response rate and response quality is not as extensive as the literature

about the effect of survey length. Therefore, this section reviews more general literature about initial helping behaviour and different levels of effort in terms of altruism. Please note that even though they are distinct, helping behaviour and altruism are used interchangeably here, because for the purpose of this research it is not relevant whether the motives for helping behaviour are purely altruistic.

2.2.1 Familiarity and Response Rate

The first relationship that is discussed is the relationship between familiarity and the response rate. That is, whether an individual's decision to participate in a survey is influenced by personally knowing the experimenter.

Extensive research by Batson and colleagues argues that purely altruistic actions can only occur if caused by empathetic concern for another. More specifically, it is argued that altruism originates from the feeling of attachment to another (Batson et al., 1981; Batson & Shaw, 1991). More recent studies support their findings that helping behaviour is indeed dependent on relationship context (Schenkler & Britt, 2001; Stürmer & Omoto, 2005; Maner & Gailliot, 2007). While these studies all considered general helping behaviour, typically of a person in need, Cialdini et al. (1997) researched different levels of severity of helping behaviour. The authors found that regardless of the severity of helping behaviour analysed, the results all show that increased perception of oneness is associated with increased levels of altruism. This provides evidence that for different types of helping behaviour, familiar individuals are considered more willing to help.

In contrast to the other mentioned researches, Groves, Cialdini and Couper (1992) researched the different factors involved specifically in the decision to participate in a survey. First of all, according to the authors there are several reasons why individuals agree to participate in a survey. One of those reasons is liking, i.e. individuals are more willing to comply with a survey request if they like individual who asks them to participate. Other reasons to comply with a survey request are reciprocity, consistency with earlier commitments, following the lead of other similar individuals, authority suggestions and scarcity of opportunity. Based on a combination of these results, in this research it is expected that personally knowing the experimenter has an effect on response rate. Therefore, the following is hypothesised:

H5: A survey in which the respondent personally knows the experimenter leads to a different response rate than a survey in which the respondent does not personally know the experimenter.

2.2.2 Familiarity and Response Quality

While response rate only allows to make inferences about individuals' initial helping behaviour, response quality also allows for analysis of the degree of willingness to help in survey responses. That is, to which degree respondents are willing to provide a good response rather than just any response. Therefore this section reviews literature in which varying degrees of willingness to help are researched. Research by Maner and Gailliot (2007) analysed how the degree of willingness to help differed for closely related individuals compared to strangers. The authors used the following design: respondents were randomly assigned to one of two scenarios of individuals needing help and within these two scenarios the individual needing help was either a close relative or a stranger. In each scenario and condition, respondents rated their willingness to help on a 7-point scale. Expected mediators of the relationship between familiarity and helping behaviour were empathetic concern, general negative affect and the degree of oneness. Results show that in both scenarios respondents were more willing to help when the individual was closely related than when it was a stranger. In addition, for closely related individuals empathetic concern was important for willingness to help while no such relationship was found for strangers. For strangers, the perception of being at one with another was a more important predictor of willingness to help. These findings therefore suggest that individuals have a higher degree of willingness to help familiar individuals than unfamiliar individuals. Therefore, it is expected that the response quality of familiar individuals differs from the response quality of unfamiliar individuals.

2.2.2.1 Familiarity and Dropout Rate

Similar to in Section 2.1.2, there are three measures of response quality. The first measure of response quality is dropout rate. As the earlier mentioned researches show, familiarity is associated with increased altruism. Therefore, it is expected that the dropout rate is different for familiar individuals than for non-familiar individuals. This leads to the following hypothesis:

H6: A survey in which the respondent personally knows the experimenter leads to a different dropout rate than a survey in which the respondent does not personally know the experimenter.

2.2.2.2 Familiarity and Uncertain Responses

The second measure of response quality is the number of uncertain responses, i.e. "don't know" or "no opinion" responses. As mentioned, individual willingness to help varies for different levels of familiarity. Therefore, the following effect is hypothesised:

H7: A survey in which the respondent personally knows the experimenter leads to a different number of uncertain responses than a shorter survey.

2.2.2.3 Familiarity and Variability of Responses

The last measure of response quality that is analysed is the variability of responses to questions in grids. Similar to the previous hypotheses, familiarity is associated with increased altruism and therefore with a different response quality. Thus, the following effect is expected:

H8: A survey in which the respondent personally knows the experimenter leads to a different variability of responses to questions in grids than a survey in which the respondent does not personally know the experimenter.

3. Data and Methodology

In this section, the data and methodology of this research are explained. Firstly, the experimental design is explained. Secondly, a description of the variables is given. Thirdly, the descriptive statistics are presented. Lastly, the methodology that is used to analyse differences in response rate and response quality is described.

3.1 Experimental Design

To answer the research question of how survey length and familiarity affect the response rate and response quality in voluntary web-based surveys, an experiment is designed. The experiment uses a Qualtrics survey design. There are two versions of the survey: a short survey lasting approximately 3 minutes and a long survey lasting approximately 10 minutes. Since research about the optimal number of items and survey length is not at all conclusive, the length of the survey is determined based on intuition and best practices in student surveys. The expected time to participate in the two versions is based on a pre-test of the survey. In addition to two versions of the survey, there are also two conditions: individuals either personally know the experimenter or not. Furthermore, all individuals (i.e. both the familiar and unfamiliar individuals) are randomly assigned to either of the two versions of the survey (i.e. long or short version). Respondents are invited to participate by email, email addresses are obtained from yearbooks of Dutch student associations. In total, 3007 email addresses were used. The survey period lasted 6 weeks in 2017: potential respondents received the email invitation on June 2 and the survey was closed on July 17. A reminder was sent on June 5. The email invitation and the reminder can be found in Appendix A.1 and Appendix A.5 respectively.

Since this research only considers objective measures of response rate and quality, the topic of the survey is irrelevant. However, to be able to check the robustness of the results, there are twenty questions in the survey about altruism based on the Altruism Scale of Rushton, Chrisjohn and Fekken (1981). Compared to the original scale, two questions are changed into a negative framing and one question is replaced by a question to check the respondents' attention. To analyse if response quality indeed differs with the length of a survey, it is important to have a set of the same questions in the beginning of the survey in one version and at the end of the survey in the other version. The Altruism Scale is the set of questions that is presented in both versions of the survey. In addition, to increase the length of the survey there are twenty questions about egoism following the Egoism Scale of Weigel, Hessing and Elffers (1999). Because the questions about egoism are not analysed and are only added to increase the length of the survey in the long version, they are added in the beginning of the long survey. Thereafter the Altruism Scale is presented. After the altruism scale is presented, respondents are requested to answer a

mathematical question to observe differences in effort. Lastly, information about the demographic variables age, gender, education and the relationship with the experimenter is requested at the end of the survey because these variables are expected to affect responses and altruism (Kaldenberg, Koenig, & Becker, 1994; Andreoni & Vesterlund, 2001; Malhotra, 2008; Cialdini et al., 1997). The full survey can be found in Appendix A.3.

To have a high response rate and to obtain reliable results, there are several design issues that are taken into account. First of all, as the research question already suggests, there are no monetary incentives offered. This is done to ensure that respondents are intrinsically motivated and to have a comparable setting as to what is often the case in academic research by students. Secondly, to promote a high response rate, it is of importance to invite individuals to a survey in an attractive way (Groves, Cialdini, & Couper, 1992). To increase the response rate the topic altruism is mentioned in both the email invitation and on the introductory page (Marcus et al, 2007). Furthermore, the Erasmus University and its logo are explicitly mentioned in the email invitation and on the introductory page of the survey because university sponsorship of the survey is expected to increase response rates (Fox, Crask, & Kim, 1988; Heerwegh & Loosveldt, 2006). In addition, to increase the response rate without affecting the response quality, a reminder is sent to all invited individuals who did not complete the survey yet (Crawford, Couper, & Lamias, 2001; Deutskens et al., 2004). The reminder is sent 3 days after the initial invitation.

3.2 Variable Description

This section describes the variables that are included in this research. Firstly, it is discussed how the response rate is measured and how differences in response rates are analysed. Thereafter, the measurement and analysis of the three measures of response quality are discussed.

3.2.1 Version

The variable Version is a binary variable that indicates whether the respondents participated in the short version or not. This variable takes value “1” if the respondent participated in the short version of the survey and it takes value “0” if the respondent did not participate in the short version, i.e. participated in the long version.

3.2.1 Familiarity

Similar to Version, the variable Familiarity is also a binary variable. This variable indicates whether respondents personally know the experimenter or not. Familiarity takes the value “1” if the respondent is considered to be familiar and it takes the value “0” if the respondent is considered to be non-familiar. Please note that this is judged highly subjectively by the researcher. Respondents also reported their degree of familiarity to the researcher, however, they were only

asked to do so at the end of the survey. This means that a lot of information about familiarity is lost if this so called Self-reported Familiarity is used. As can be seen in Appendix B.1, the two measures of familiarity are relatively similar. Therefore, to capture most information about whether the respondents personally know the researcher or not, the analyses that include Familiarity are performed by using familiarity reported by the researcher rather than familiarity reported by respondents themselves.

3.2.3 Response Rate

To measure the response rate, individuals received their invitation to participate by email. To make sure that respondents only participated once and to identify the familiarity condition of the respondents, all respondents received a personalised hyperlink. Since respondents only got to see the length of the survey on the introductory page, for the survey length analysis the response rate is calculated in a different way than for the familiarity analysis. For the survey length, all individuals who have opened the link and closed the survey immediately after seeing the length of the survey are considered non-respondents. The response rate is therefore calculated by dividing the number of respondents that continued after seeing the length of the survey by the total number of respondents that opened the link. For familiarity, the response rate is calculated by dividing the number of (partial) responses by the total number of invitations sent. Bounced and duplicate emails are not included in the calculation of the response rate. Since individuals are either responders or not, the variable response is a binary variable.

3.2.4 Response Quality

The measurement and analysis of the three measures of response quality - dropout rate, uncertain responses and variability of responses – are now discussed per measure.

3.2.4.1 Dropout Rate

The first measure of response quality, dropout rate, is the number of partial responses in relation to the total number of surveys started. Whenever an individual started the survey but did not finish it, this is considered to be a partial response. To be able to measure actual dropout rate, the partial completion time frame is set to Qualtrics' maximum of four hours, Qualtrics' minimum partial completion time frame. This means that if respondents close their web browser without completing the survey, they have four hours to continue their participation. If it takes more than 4 hours to continue participation, responses are stored as partially completed and these individuals are considered dropouts. Since respondents either finish the survey or drop out, dropout is a binary variable.

3.2.4.2 Uncertain Responses

The second measure of response quality, the number of uncertain responses, is measured by observing responses to the question in which respondents have the possibility to answer “Don’t know”. The total amount of “Don’t know” answers is counted and compared across versions and conditions. Respondents either give an uncertain response or they do not give an uncertain response, so this is a binary variable as well.

3.2.4.3 Variability of Responses

The variability of responses to questions in grids is measured by observing how much the responses to the twenty altruism questions vary. All questions about altruism are answered on a five-point scale which allows respondents to choose a response pattern that is identical for all questions, a straight-line response strategy. Respondents either choose a straight-line response strategy or not, this is a binary variable. In addition, variability of responses is also measured by comparing the responses to each question individually. This means that the response to each question is measured as a categorical variable with values that match the five response categories.

3.2.5 Control Variables

The first control variable in this research is Altruism. In the survey all respondents are presented with an adjusted version of the Altruism Scale of Rushton, Chrisjohn and Fekken (1981). Respondents have to report the frequency with which they have carried out 20 acts. Compared to the original scale, two questions are changed into a negative framing and one question is replaced by a question to check the respondents’ attention. As proposed by Rushton, Chrisjohn and Fekken (1981), the total altruism score of respondents is calculated by summing the frequency with which they have carried out the acts. A response of “Never” corresponds to a 1, “Once” corresponds to a 2, “More than once” corresponds to a 3, “Often” corresponds to a 4 and “Very Often” corresponds to a 5. The two questions that are negatively framed are reverse-scored. For the control question, respondents have to report the frequency with which they have had a fatal heart attack. Evidently, respondents have to answer “Never” to this question. Since the response to this question does not give information about altruism, it is excluded from the calculation of the respondents’ altruism score. Having 20 statements with the control question being excluded from the calculation of the altruism score, respondents could score a minimum of 19 and a maximum of 95. Please find the complete version of the Altruism Scale in Appendix A.3.3

The second control variable is Age. Respondents had to report their age by typing a number. Therefore, Age is a continuous variable. For the variable Gender, respondents reported one out of three options “Male”, “Female” or “Other”. Since no respondents reported “Other”, Gender is treated as a binary variable. Thirdly, respondents also reported their highest level of

completed education. There were five given options and one option where respondents could report their education if their highest level of completed education was not one of the given options. The variable Education is therefore a categorical variable with six categories. The variable Student Association was reported upfront for all individuals. There are five different student associations from which respondents were recruited. Therefore, Student Association is a categorical variable. Another control variable is Self-reported Relationship. In the survey respondents had to report being “Strangers”, “Acquaintances”, “Good friends” or “Family members”. The variable Self-reported Relationship is therefore also a categorical variable.

Furthermore, there are also two control variables that concern the duration and speed of response. Duration measures how long a respondent took to finish the survey or how long a respondent took before dropping out of the survey. Duration is measured in seconds and is a continuous variable. Response Speed in turn measures how long it took respondents to start the survey. That is, how many days after the initial invitation did individuals start the survey. Response Speed is a categorical variable with the values “Day 0”, “Day 1” “Day 2”, “Day 3” or “Day 4 or Later”. The descriptive statistics of all variables are presented in the next section.

3.3 Descriptive Statistics

This section presents the descriptive statistics of the response rate as well as of the variables of interest and the control variables. As can be seen in Table 1, the database contained 3007 email addresses, 5 of them were duplicates and 135 of the sent emails bounced. This means that 2867 individuals received an invitation to participate in the web-based survey. In total, 2487 individuals received the invitation without giving a response, 169 respondents completed the survey, 120 individuals chose to unsubscribe from the mailing list, and 91 respondents only completed the survey partially. Seven of the respondents that have the status ‘Opted Out’ partially completed the survey. These seven are not shown in the status ‘Partially Completed Survey’ but they are included in the calculation of response rate. Combing the finished surveys and the partially completed surveys this study contains 267 respondents. Overall, this is a response rate of 9.31%.

Table 1: Response Status

	Frequency	Percent
Email Bounced	135	4.49
No Response	2487	82.71
Finished Survey	169	5.62
Opted Out	120	3.99
Partially Completed Survey	91	3.03
Skipped As Duplicate	5	0.17
Total	3007	100

Table 2 contains descriptive statistics of the control variables. As can be seen, 39.70% of the respondents were categorised into personally knowing the experimenter. Almost half of the respondents participated in the short version and the other half participated in the long version of the survey. On average, respondents scored 43.37 on the Altruism Scale. Slightly more than half of the respondents is female and the mean age of the respondents is 23.55 years old. The largest amount of respondents is academically educated, 75.29% stated that their highest level of completed education is University. In addition, 20.59% reported high school as their highest level of education¹, for 2.94% of the respondents higher vocational education is the highest completed level of education and 1.18% indicated to have completed another level of education than stated. As mentioned before, the respondents are all members of a student association. Due to privacy concerns, the names of these student associations are not reported and the letters A to E are used to indicate the different associations. Exactly 50% of the respondents reported to be a stranger to the experimenter, 31.18% are acquaintances, 17.64% are good friends and 1.18% is family. Furthermore, the average duration of the survey was 421.36 seconds. Almost half of the respondents responded on the first day after receiving the invitation, 9.36% responded after one day, 5.62% responded after 2 day, 29.21% responded after three days and 9.36% responded after four days or more. This means that after the invitation, most individuals respond within the first days and the response rate gradually decreased per day. On the third day, the day of the reminder, response increased again. Only a small amount of individuals decided to participate later than four days after the initial invitation. All this information can be found in Table 2 on the next page. Please note that due to early dropout not all variables have 267 observations.

¹ In general, members of student associations are relatively highly educated. Presumably, the respondents indicating that high school is their highest level of completed education are currently studying to obtain a degree at university or at higher vocational education.

Table 2: Descriptive Statistics

		Number/Percent
Familiar	Non-Familiar	60.30%
	Familiar	39.70%
Version	Number of Observations	267
	Short	48.79%
	Long	53.53%
Altruism	Number of Observations	248
	Mean	43.37
	Standard Deviation	8.22
Gender	Number of Observations	171
	Male	46.47%
	Female	53.53%
Age	Number of Observations	170
	Mean	23.55
	Standard Deviation	2.09
Educational Level	Number of Observations	170
	Less than High School	0.00%
	High School	20.59%
	Intermediate vocational education	0.00%
	Higher vocational education	2.94%
	University	75.29%
Student Association	Other	1.18%
	Number of Observations	170
	A	8.24%
	B	54.68%
	C	10.49%
	D	8.99%
Self-reported Relationship	E	17.60%
	Number of Observations	267
	Stranger	50.00%
	Acquaintances	31.18%
	Good Friends	17.65%
	Family	1.18%
Duration	Number of Observations	170
	Mean	421.36
	Standard Deviation	1105.55
	Observations	267
Response Speed	Day 0	46.44%
	Day 1	9.36%
	Day 2	5.62%
	Day 3	29.21%
	Day 4 or Later	9.36%
	Number of Observations	267

3.4 Methodology

In this section, the methodology that is used to analyse the effect of time and familiarity on response rate and response quality is explained. The methodology is discussed in a separate section per variable.

3.4.1 Response Rate

The first variable of which the effect of survey and familiarity is analysed is the response rate. To analyse the differences across versions and conditions, a 2x2 Fisher's exact and a Pearson's chi-squared test are performed.

3.4.2 Response Quality

Response quality is measured in three ways. Most variables are tested by means of a 2x2 Fisher's exact and a Pearson's chi-squared test. For matters of completeness, the methodology for all three measures of response quality are now discussed separately.

3.4.2.1 Dropout Rate

The first measure of response quality, dropout rate, is compared across versions and conditions. Since dropout is a binary variable, and there are two versions of the survey a 2x2 Fisher's exact and a Pearson's chi-squared test are performed.

3.4.2.2 Uncertain Responses

Since uncertain responses are also measured in terms of a binary variable, this measure of response quality is also analysed by means of a 2x2 Fisher's exact and a Pearson's chi-squared test.

3.4.2.2 Variability of Responses

The effect of time and familiarity on the variability of responses to questions in grids is measured in two different ways. First of all, variability of responses is measured by analysing how many respondents choose a straight-line response strategy. The number of straight-line responses is compared across versions and conditions by means of a 2x2 Fisher's exact and a Pearson's chi-squared test. Secondly, the variability of responses is analysed by comparing the responses of each separate question. The differences in responses per question across versions and conditions are compared by using a Mann-Whitney U test.

The measures of response quality for which the abovementioned non-parametric tests show significant differences across versions or familiarity conditions, are analysed more in depth. A logistic regression that also includes the control variables is performed. For the response rate, this extra regression including control variables is not possible since there is very limited information available about the non-respondents.

4. Results

In this section, the results of the analyses are presented. The section starts with the effect of survey length on response rate and response quality. Thereafter, the effect of familiarity on response rate and response quality is presented. Analyses of a number of different control variables revealed no a priori differences between the versions and only some differences between the familiarity conditions. The tests for differences in control variables can be found in Appendix B.2.

4.1 Survey Length

This section contains the results of the analyses performed to understand the effect of time, on the response rate and response quality in web-based surveys. First, the effect of the survey length on the response rate is discussed. Thereafter, the results of the effect of the survey length on the three different measures of response quality are discussed.

4.1.1 Survey Length and Response Rate

This section contains the results of the analysis about the effect of survey length on response rate. The Fisher's exact and Pearson's chi-squared tests show that the null hypothesis of no differences is rejected. The observed differences are significantly different at 1% significance level. These results therefore provide evidence that supports H1 that a longer survey leads to a different response rate than a shorter survey. As can be seen in Table 3, there was relatively less response in the long version of the survey than in the short version. This means that relatively more respondents closed the survey after seeing the length of the long version than after seeing the length of the long version.

Table 3: Time and Response Rate

	Short Version	Long Version
Response	102	70
Non-Response	25	48
Number of Observations	127	118

Fisher's exact = 0.000

Pearson's chi2 = 12.8868***

4.1.2 Survey Length and Response Quality

As mentioned before, the three measures of response quality used in this research are dropout rate, uncertain responses and variability of responses. The results of the analyses are now discussed in this order.

4.1.2.1 Survey Length and Dropout Rate

To analyse the effect of the length of a survey on the first measure of response quality, dropout rate, a Fisher's exact is performed. In addition, a Pearson's chi-squared test is performed. The statistical tests show that the observed differences in dropout for the short and long version of the survey are significantly different at 1% significance level. These results therefore provide evidence that supports H2 that a longer survey leads to a different dropout rate than a shorter survey. In addition, as can be seen in Table 4, relatively more respondents dropped out in the long version of the survey than in the short version of the survey.

Table 4: Time and Dropout Rate

	Short Version	Long Version
Dropout	25	53
Finished	102	68
Number of Observations	127	121

Fisher's exact = 0.000

Pearson's chi2 = 16.7159 ***

4.1.2.2 Survey Length and Uncertain Responses

In comparing the number of uncertain responses across versions, one cannot use a Pearson's chi-squared test because the amount of uncertain responses is too low, there are only six uncertain responses in the short version and only three in the long version. Therefore, only a Fisher's exact test is performed to analyse uncertain responses. As can be seen in Table 5, it cannot be stated that the observed differences in uncertain responses are significantly different across versions at any conventional level of significance. This means that the amount of uncertain responses in the short version do not differ significantly from amount of uncertain responses in the long version. Not being able to find significant differences here is most likely to be explained by a small amount of uncertain responses reported in either version. These results therefore cannot confirm H3 that a longer survey leads to a different number of uncertain responses than a shorter survey.

Table 5: Time and Uncertain Responses

	Short Version	Long Version
Uncertain Responses	6	3
No Uncertain Responses	96	65
Number of Observations	102	68

Fisher's exact = 0.743

4.1.2.3 Survey Length and Variability of Responses

The third measure of response quality that is analysed here, is variability of responses to questions in grids. The effect of time on variability of responses is analysed in two different ways. Firstly, variability of responses to questions in grids is measured by observing the amount of respondents that choose a straight-line response strategy. However, there are no respondents who chose this response strategy. This could be explained by the fact that none of the respondents experienced enough burden to choose this response strategy. The second analysis of the effect of time on variability of responses is measured by comparing the responses of each question separately. As the first column of Table 6 shows, only the responses to the first and twelfth question differ significantly between the short and the long version. For the eighteen other questions, the null hypothesis of an equal distribution cannot be rejected at any conventional level of significance. Given the small amount of differences, the expectation that a longer survey leads to a different variability of responses to questions in grids than a shorter survey as described in H4 cannot be confirmed with these results.

Overall, this section has shown the effect of the length of a survey on response rate and response quality. The analysis of response rate has shown that there are significant differences in response rate across the different versions. The data shows a relatively higher response rate for the short version than for the long version. In addition, the analysis of response quality has also provided some insights in the effect of length of a survey. Dropout rate, the first measure of response quality, has been shown to be significantly different across the different versions. In this sample, more respondents dropped out of the long survey than out of the short survey. Analysis of the second measure of response quality, uncertain responses, revealed that the null hypothesis of no differences cannot be rejected at any conventional significance level. The third measure of response quality, variability of responses to questions in grids, only showed significantly different responses to two out of twenty questions. In conclusion, in this research dropout rate is the only measure of response quality that shows that there are significant differences between respondents in a long version compared to respondents in a short version. Therefore, a logistic regression of dropout rate will be performed after the results of the effect of personally knowing the experimenter are presented.

Table 6: Z-statistics of Mann-Whitney U Tests per Question

	Version	Familiarity Condition
Question 1	2.693***	-0.195
Question 2	-0.235	-1.253
Question 3	0.421	-2.527**
Question 4	0.675	-1.661*
Question 5	1.075	0.817
Question 6	-0.816	-0.451
Question 7	0.802	-0.931
Question 8	1.091	-1.101
Question 9	1.529	-0.300
Question 10	-0.969	-0.173
Question 11	0.980	0.378
Question 12	1.684*	-0.131
Question 13	-0.587	-1.524
Question 14	0.995	-0.133
Question 15	-0.149	-1.996
Question 16	-0.617	0.260
Question 17	-0.263	1.577
Question 18	0.739	-1.374
Question 19	-1.242	-0.159
Question 20	1.019	-0.786
Number of Observations	171	171

*Significant at 10%.

**Significant at 5%.

***Significant at 1%.

4.2 Familiarity

While the previous section discussed the effect of length of a survey on response rate and response quality, this section discusses the effect of personally knowing the experimenter on response. Similar to section 4.1, in this section the effect on response rate is discussed first. Thereafter, the effect of familiarity on three different measures of response quality is discussed. Again, an overview of the results can be found in Table 10.

4.2.1 Familiarity and Response Rate

The effect of personally knowing the experimenter on the response rate of the survey is analysed by means of a Fisher's exact and a Pearson's chi-squared test. As can be seen in Table 7, the observed differences in response rates across familiarity conditions are significantly different at

1% significance level. The results therefore provide evidence that confirms H5 that a survey in which the respondent personally knows the experimenter leads to a different response rate than a survey in which the respondent does not personally know the experimenter. Table 7 also shows that the response rate of familiar individuals is significantly higher than the response rate of non-familiar individuals.

Table 7: Familiarity and Response Rate

	Familiar	Non-Familiar
Response	105	155
Non-Response	392	2210
Number of Observations	2365	2365

Fisher's exact = 0.000

Pearson's chi2 = 105.6004***

4.2.2 Familiarity and Response Quality

The results of familiarity and response quality are presented in a similar manner as the results of survey length and familiarity. The results of the effect of familiarity on dropout rate, uncertain responses and on variability of responses respectively are discussed first. Thereafter, an overview of the results is given in Table 10.

4.2.2.1 Familiarity and Dropout Rate

The first analysed response quality effect is the effect of personally knowing the experimenter on the dropout rate. As shown in Table 8, the results of the Fisher's exact test and the Pearson's chi-squared test show that the observed differences in dropout for familiar and unfamiliar respondents are significantly different at 5% significance level. This confirms the expectation that a survey in which the respondent personally knows the experimenter leads to a different dropout rate than a survey in which the respondent does not personally know the experimenter as hypothesised in H6. In addition, Table 8 also shows that more non-familiar respondents than familiar respondents drop out of the survey before having finished it.

Table 8: Familiarity and Dropout Rate

	Familiar	Non-Familiar
Dropout	30	67
Finished	76	94
Number of Observations	106	161

Fisher's exact = 0.028

Pearson's chi2 = 4.8975**

4.2.2.2 Familiarity and Uncertain Responses

Similar to the analysis of the effect of time on uncertain responses, there are too little uncertain responses to perform a Pearson's chi-squared test. Therefore, only a Fisher's exact test is performed. As can be seen in Table 9, the observed differences in uncertain responses across familiarity conditions cannot be said to be significantly different at any conventional significance level. With these results, the expectation of H7 that a survey in which the respondent personally knows the experimenter leads to a different number of uncertain responses than a shorter survey cannot be confirmed.

Table 9: Familiarity and Uncertain Responses

	Familiar	Non-Familiar
Uncertain Responses	2	7
No Uncertain Responses	74	87
Number of Observations	76	94

Fisher's exact = 0.190

4.2.2.3 Familiarity and Variability of Responses

As mentioned in Section 4.1.2.3, there are no straight-line responders in this sample. So for familiarity it is also not possible to analyse the differences in variability of responses in this way. However, for the second measure of variability of responses, namely the difference in individual responses per question, an analysis of differences between familiar and non-familiar respondents can be performed. As can be seen in the second column of Table 6, the observed differences between responses in question 3 and question 4 are significantly different at 5% and 10% significance level, respectively. For the other eighteen questions, the null hypothesis of an equal distribution cannot be rejected at any conventional level of significance. Given the small amount of differences, based on these results a survey in which the respondent personally knows the experimenter does not lead to a different variability of responses to questions in grids than a survey in which the respondent does not personally know the experimenter. Therefore, H8 cannot be confirmed with these results.

This section has analysed the effect that personally knowing the experimenter has on response rate and response quality. The analysis of response rate has shown significant differences between individuals who personally know the experimenter and individuals who do not personally know the experimenter. The analysis of response quality has shown that there is only one measure of response quality for which significant differences between the familiarity conditions are observed. For the first measure of response quality, dropout rate, a significant difference is found between the different familiarity conditions. For the other two measures of

response quality, uncertain responses and variability of responses, no significant difference between the familiarity conditions are observed. An overview of these results including the effect of survey length, can be found in Table 10.

Table 10: Overview Results

		Version	Familiarity
Response Rate	Statistical differences	Yes***	Yes***
	Conform Hypothesis	Yes	Yes
Dropout Rate	Statistical differences	Yes***	Yes**
	Conform Hypothesis	Yes	Yes
Uncertain Responses	Statistical differences	No	No
	Conform Hypothesis	No	No
Variability of Responses	Statistical differences	No	No
	Conform Hypothesis	No	No

*Significant at 10%.

**Significant at 5%.

***Significant at 1%.

4.3 Extra Analysis Dropout Rate including Control Variables

As also shown in Table 10, Dropout Rate is significantly different across versions and familiarity conditions. Therefore, in this section a logistic regression of this measure of response quality is analysed. More specifically, a logistic regression of Dropout Rate on Version, Familiarity, Duration and Response Speed is performed. As can be seen in Appendix B.2, Duration is significantly different per version. Therefore, the interaction variable Duration_Version is also included in the regression. The results of the logistic regression are shown in Table 11. It can be seen that respondents of the short version have a lower predicted probability to drop out compared to respondents of the long version. This effect is significant at 1% significance level. In addition, it can be seen that respondents who personally know the experimenter have a lower predicted probability to drop out compared to respondents who do not personally know the experimenter. This effect is significant at 5% significance level. Since the coefficients of Duration and Duration_Version do not provide significant results at any conventional significance level, it cannot be said that the duration of the survey significantly alters the predicted probability of dropping out of the survey regardless of the length of the survey. Furthermore, in terms of response speed, respondents who started the survey on the first or second day after the initial invitation do not have a significantly different predicted probability of dropping out of the survey compared to respondents who started the survey on the day of the invitation at any conventional significance level. At 5% significance level, respondents who started the survey on the third day

after the initial invitation, the day of the reminder, have a significantly lower predicted probability to drop out of the survey compared to individuals who start the survey on the day of the invitation. Similar to respondents who started the survey on the second or third day after the invitation, respondents who started on the fourth day after the initial invitation or later do not have significantly different predicted probability of dropping out of the survey compared to respondents who started the survey on the day of the invitation at any convenient significance level.

In conclusion, it can be stated that in this sample, participating in a short survey and personally knowing the experimenter make it less likely to drop out of the survey. Furthermore, respondents who started the survey three days after the initial invitation are also less likely to drop out of the survey than respondents who started the survey on the day of the invitation.

Table 11: Logistic Regression of Dropout Rate on Version, Familiarity Condition, Duration and Response Speed

	Coefficient
Short Version	-1.365*** (0.325)
Familiar	-0.641** (0.311)
Duration	0.000 (0.000)
Duration_Version	-0.000 (0.000)
Response Speed	
Day 0	-
Day 1	0.811 (0.569)
Day 2	0.287 (0.660)
Day 3	0.813** (0.358)
Day 4 or Later	0.749 (0.524)
Constant	-0.465 (0.275)
Number of Observations	248
Pseudo R2	0.1002
Prob > chi2	0.0001

Note: Standard errors in parentheses.

*Significant at 10%.

**Significant at 5%.

***Significant at 1%.

5. Conclusion and Discussion

The aim of this research has been to analyse how the response rate and response quality in voluntary web-based surveys are influenced by the length of the survey and by personally knowing the experimenter. The research started by reviewing the literature on this matter. Based on this, eight hypotheses about the expected effects of survey length and personally knowing the experimenter were formed. Response rate is analysed by comparing the amount of respondents to the amount of non-respondents. Response quality is analysed by means of three measures: dropout rate, uncertain responses and variability of responses to questions in grids. To test the hypotheses, an experiment was designed in which members from different Dutch student associations were asked to participate in a survey. Upfront, all invited individuals were categorised as either familiar or non-familiar. When agreeing to participate in the survey, respondents were randomly allocated to either a short version with a stated expected duration of 3 minutes or to a long version with a stated expected duration of 10 minutes. This 2x2 design enabled multiple analyses of differences between the different versions and familiarity conditions. Methods included Fisher's exact test, Pearson's chi-squared test, Mann-Whitney U test and a logistic regression.

Firstly, as expected the length of the survey was found to be of significant influence to the response rate. In addition, there was only one measure of response quality for which the length of the survey was found to be of significant influence. The first measure of response quality, dropout rate, was found to be significantly affected by the survey length. For the second measure, uncertain responses, no significant differences were found. For the third measure of response quality, variability of responses to questions in grids, no respondent chose a straight-line response category suggesting that none of the respondents experienced enough burden. Furthermore, only two out of twenty questions showed significant differences in responses for the different versions.

Secondly, similar to the length of a survey, personally knowing the experimenter was also found to be of significant influence to the response rate. Furthermore, also for personally knowing the experimenter or not, only dropout rate showed significant differences between the two familiarity conditions. For the number of uncertain responses no significant differences were found between the different familiarity conditions. Since there were no straight-line responders, only the differences in responses to the twenty questions could be analysed for variability of responses to questions in grids. Again, only for two out of twenty questions significant differences were found between the different familiarity conditions. All in all, these results show that the length of a survey and personally knowing the experimenter significantly affected the response rate and dropout rate in voluntary web-based surveys. Further analysis of dropout rate showed that in this sample, participating in a short survey rather than in the long survey and personally

knowing the experimenter rather than not make it less likely to drop out of the survey. Furthermore, respondents who started this survey three days after the initial invitation, the day of the reminder, are also less likely to drop out of the survey than respondents who started the survey on the day of the invitation.

Taking everything into account, the findings of this research provide some new insights into the domain of survey design. While a lot of research has analysed the effect of survey length, to the best of my knowledge there has not yet been researched how personally knowing the experimenter influences response rate and response quality. Therefore, the finding that survey length and personally knowing the experimenter significantly affect response rate and the finding of how it affects dropout rate are interesting findings that add to current survey design research. In practice, these findings are also of relevance for students in academics that want to do research based on a voluntary web-based survey. While intuitively evident, this research has shown that for students it might be best to conduct a sufficiently short survey and to invite familiar individuals to participate to optimise response rate and to decrease dropout rate. In addition, in this sample, survey length and inviting familiar individuals did not affect the amount of uncertain responses and it only influenced the responses to two out of twenty questions. Therefore, it seems that inviting familiar individuals increases the amount of complete responses and does not affect further response quality.

However, there are several concerns that make that the findings of this research should be interpreted with caution. Therefore, the limitations of this research are also important to take into account. First of all, the measure of response rate in this sample is not completely accurate. One problem with the measure of response rate is that individuals were recruited from yearbooks of which it is highly probable that the email addresses were outdated. Therefore, it is not entirely clear how many individuals actually received the invitation and response rate is therefore not accurate. In addition, the effect of length of a survey on response rate is also not accurate. Respondents only got the information about the length of the survey after opening the survey link in the email invitation. Therefore, response rate in relation to survey length is measured in a different way than in relation to familiarity and it is biased.

Furthermore, the results of this research may not be generalisable to a different sample, a different experimenter or a different survey. First of all the sample is not representative of the total population as it only contains individuals that are members of a student association. This makes the sample similarly educated, of relatively the same age and maybe similar in their perception of altruism. In addition, the choice of this sample also makes the respondents more likely to be able to identify with the experimenter. Since being able to identify with someone makes individuals more willing to help, the results might be biased because of the sample (Cialdini et al., 1997). Furthermore, the way familiarity is measured here might really only reflect this

sample. For example, it is already known that responses are also affected by likability (Groves, Cialdini, & Couper, 1992). Therefore, if another experimenter would replicate the research even if it is with the same sample, results might be different from the results in this research. Thirdly, the results of the research might be different if there would be a different survey topic. In general, individuals' decision to participate in a survey is affected by survey topic (Marcus et al, 2007). In the email invitation, individuals were already informed about the survey topic. Another survey topic might therefore have affected the response rate, but might also have affected the response quality. In this survey respondents were asked to report their acts of altruism (and egoism), so they were directly confronted with altruistic behaviour. This, in turn, might increase their altruistic behaviour in terms of filling out the survey and it might therefore affect response quality.

Another limitation to this research is that, as mentioned in the previous paragraph, there are omitted variables that might have affected the response rate and the response quality of the sample. An example of this is the timing effects of the email invitation and the reminder (Crawford, Couper, & Lamias, 2001). Furthermore, saliency of the invitation might also affect the response rate (Sheehan & McMillan, 1999). In addition, because respondents were only asked to report their demographics in the end of the survey, analyses of control variables did not capture all respondents. If the demographics of the full sample would have been known, more conclusions could have been drawn about this. Moreover, privacy concerns have also been a problem in this research. Some individuals notified that they considered the invitation to participate in this research to be intrusive. This might have affected responses, especially of individuals who do not personally know the researcher.

In addition, only very little respondents responded with uncertain responses and no respondent chose a straight-line strategy. This caused that the hypotheses could not be tested as expected. It could be that the respondents did not experience enough burden to choose either of those response strategies. It could also be that there would have been better measures of response quality that are not taken into account here. Furthermore, analyses of the control variables revealed some differences in the distribution of gender and student associations across familiarity conditions. Therefore, the results of the effect of personally knowing the experimenter or not should be interpreted with some caution. Lastly it is important to take into account that the hypotheses of this research are two-sided. This means that no conclusions can be drawn about the size and the direction of the effects. Thus, the results only provide some preliminary evidence about the differences in response rate and response quality due to survey length and personally knowing the experimenter.

Given the reported limitations, the effect of survey length and the effect of personally knowing the experimenter on response rate and response quality have to be interpreted with caution and further research is required to provide a definite understanding of the relationship

between these variables. Most of the reported limitations can be overcome if an existing database in which individuals consented with participating in surveys would be used. Conditional on that respondents are presented the survey length upon invitation and on that email addresses are up to date, response rate calculations would be more accurate. Furthermore, a large database allows to do multiple analyses to test whether the findings of this research also hold in a different context i.e. with different samples, a different experimenter and with a different topic of the survey. In addition, more extensive analyses could be performed with a larger dataset. If the assumptions of performing parametric tests hold, parametric analyses would allow to identify the size and direction of the effect of survey length and personally knowing the experimenter or not. Furthermore, if the demographic information is known for the whole sample, including the non-respondents, more extensive analyses could be performed and results could be corrected for demographic differences. Also, privacy issues would be overcome if individuals in the database indeed consented to participate in surveys. In addition, the limitation that only a little amount of respondents chose to report uncertain responses or a straight-line response strategy could be tackled by exploring whether there are other measures of response quality that are more suitable. It could also be tackled by making longer surveys so that burden is increased. However, this might in turn negatively affect response rate. This is a trade-off that will have to be considered in later research.

All in all, because of the reported limitations the results of this research have to be interpreted with some caution. However, using an extensive database was not possible within the scope of this research. This research therefore only provides some preliminary evidence and can only act as a springboard to future research about the relationship between the length of a survey, personally knowing the experimenter and response rate and response quality.

Bibliography

- Andreoni, J., & Vesterlund, L. (2001). Which is the Fair Sex? Gender Differences in Altruism. *The Quarterly Journal of Economics*, 116(1), 293-312.
- Batson, D. C., & Shaw, L. L. (1991). Evidence for Altruism: Toward a Pluralism of Prosocial Motives. *Psychological Inquiry*, 2(2), 107-122.
- Batson, D. C., Duncan, B. D., Ackermann, P., Buckley, T., & Birch, K. (1981). Is Empathic Emotion a Source of Altruistic Motivation? *Journal of Personality and Social Psychology*, 40, 290-302.
- Batson, D. C., O'Quin, K., Fultz, J., Vanderplas, M., & Isen, A. M. (1983). Influence of Self-reported Distress and Empathy on Egoistic versus Altruistic Motivation to Help. *Journal of Personality and Social Psychology*, 45(3), 706-718.
- Bradburn, N. (1978). Respondent Burden. *American Statistical Association*, 35-40.
- Cialdini, R. B., Brown, S. L., Lewis, B. P., Luce, C., & Neuberg, S. L. (1997). Reinterpreting the Empathy- Altruism Relationship: When One Into One Equals Oneness. *Journal of Personality and Social Psychology*, 73(3), 481-494.
- Cook, C., Heath, F., & Thompson, R. L. (2000). A Meta-Analysis of Response Rates in Web-or Internet-Based Surveys. *Educational and Psychological Measurement*, 60(6), 821-836.
- Crawford, S. D., Couper, M. P., & Lamias, M. J. (2001). Web Surveys - Perceptions of Burden. *Social Science Computer Review*, 19(2), 146-162.
- Darley, J. M., & Batson, D. C. (1973). "From Jerusalem to Jericho": A Study of Situational and Dispositional Variables in Helping Behaviour. *Journal of Personality and Social Psychology*, 27(1), 100-108.
- DeMaio, T. J. (1980). Refusals: Who, Where and Why. *Public Opinion Quarterly*, 44(2), 223-233.
- DeRouvray, C., & Couper, M. P. (2002). Designing a Strategy for Reducing "No Opinion" Responses in Web-Based Surveys. *Social Science Computer Review*, 20(1), 3-9.
- Deutskens, E., De Ruyter, K., Wetzels, M., & Oosterveld, P. (2004). Response Rate and Response Quality of Internet-Based Surveys: An Experimental Study. *Marketing Letters*, 15(1), 21-36.
- Fox, R. J., Crask, M. R., & Kim, J. (1988). Mail Survey Response Rate. *Public Opinion Quarterly*, 52(4), 467-491.
- Galesic, M. (2006). Dropouts on the Web: Effects of Interest and Burden Experienced During an Online Survey. *Journal of Official Statistics*, 22(2), 313-328.
- Galesic, M., & Bosnjak, M. (2009). Effect of Questionnaire Length on Participation and Indicators of Response Quality in a Web Survey. *Public Opinion Quarterly*, 73(2), 349-360.
- Groves, R. M., Cialdini, R. B., & Couper, M. P. (1992). Understanding the Decision to Participate in a Survey. *Public Opinion Quarterly*, 56(4), 475-495.
- Heerwegh, D., & Loosveldt, G. (2006). An Experimental Study on the Effects of Personalization, Survey Length Statements, Progress Indicators, and Survey Sponsor Logos in Web Surveys. *Journal of Official Statistics*, 22(2), 191-210.

- Herzog, A. R., & Bachman, J. G. (1981). Effects of Questionnaire Length on Response Quality. *Public Opinion Quarterly*, 45(4), 549-559.
- Hess, S., Hensher, D. A., & Daly, A. (2012). Not Bored Yet - Revisiting Respondent Fatigue in Stated Choice Experiments. *Transportation Research Part A*, 46(3), 626-644.
- Hoerger, M. (2010). Participant Dropout as a Function of Survey Length in Internet-Mediated University Studies: Implications for Study Design and Voluntary Participation in Psychological Research. *Cyberpsychology, Behavior, and Social Networking*, 13(6), 697-700.
- Kaldenberg, D. O., Koenig, H. F., & Becker, B. W. (1994). Mail Survey Response Rate Patterns in a Population of the Elderly: Does Response Deteriorate With Age? *The Public Opinion Quarterly*, 58(1), 68-76.
- Krebs, D. (1975). Empathy and Altruism. *Journal of Personality and Social Psychology*, 32(6), 1134-1146.
- Krosnick, J. A. (1991). Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys. *Applied Cognitive Psychology*, 5(3), 213-236.
- Malhotra, N. (2008). Completion Time and Response Effects in Web Surveys. *Public Opinion Quarterly*, 72(5), 914-934.
- Maner, J. K., & Gailliot, M. T. (2007). Altruism and Egoism: Prosocial Motivations for Helping Depend on Relationship Context. *European Journal of Social Psychology*, 37(2), 347-358.
- Marcus, B., Bosnjak, M., Lindner, S., Pilischenko, S., & Schütz, A. (2007). Compensating for Low Topic Interest and Long Surveys: a Field Experiment on Nonresponse in Web Surveys. *Social Science Computer Review*, 25(3), 372-383.
- Oxford Dictionaries. (2017). *Altruism*. Retrieved April 21, 2017, from <https://en.oxforddictionaries.com/definition/altruism>
- Rosenbaum, J. R., Wells, C. K., Viscoli, C. M., Brass, L. M., Kernan, W. N., & Horwitz, R. I. (2005). Altruism as a Reason for Participation in Clinical Trials was Independently Associated with Adherence. *Journal of Clinical Epidemiology*, 58(11), 1109-1114.
- Rushton, J. P., Chrisjohn, R. D., & Fekken, G. C. (1981). The Altruistic Personality and the Self-Report Altruism Scale. *Personality and Individual Differences*, 2(4), 292-302.
- Schenkler, B. R., & Britt, T. W. (2001). Strategically Controlling Information to Help Friends: Effects of Empathy and Friendship Strength on Beneficial Impression Management. *Journal of Experimental Social Psychology*, 37(5), 357-372.
- Sheehan, K. B. (2001). E-mail Survey Response Rates: A Review. *Journal of Computer-Mediated Communication*, 6(2).
- Sheehan, K. B., & McMillan, S. J. (1999). Response Variation in E-mail Surveys: An Exploration. *Journal of Advertising Research*, 39(4), 45-54.
- Singer, E., & Ye, C. (2013). The Use and Effects of Incentives in Surveys. *The Annals of the American Academy of Political and Social Science*, 645(1), 112-141.
- Stürmer, S., & Omoto, A. M. (2005). Prosocial Emotions and Helping: The Moderating Role of Group Membership. *Journal of Personality and Social Psychology*, 88(3), 532-546.

- Vicente, P., & Reis, E. (2010). Using Questionnaire Design to Fight Nonresponse Bias in Web Surveys. *Social Science Computer Review*, 28(2), 251-267.
- Weigel, R. H., Hessing, D. J., & Elffers, H. (1999). Egoism: Concept, Measurement and Implications for Deviance. *Psychology, Crime & Law*, 5(4), 349-378.
- Yan, T., Conrad, F. G., Tourangeau, R., & Couper, M. P. (2010). Should I Stay or Should I go: The Effects of Progress Feedback, Promised Task Duration, and Length of Questionnaire on Completing Web Surveys. *International Journal of Public Opinion Research*, 23(2), 131-147.

Appendix A: Survey Design

A.1 Email invitation

Dear Sir, Madam,

As part of my Master thesis, you are kindly invited to participate in a research about altruism among members of student associations. I have tried to keep the survey as short as possible. Please either follow the link to the Qualtrics survey or copy and paste the URL in your Internet browser to participate.

Follow this link to the Survey:

[Take the Survey](#)

Or copy and paste the URL below into your internet browser:

https://erasmusuniversity.eu.qualtrics.com/jfe/preview/SV_6kWHEPI3owwMd5H?Q_CHL=preview

Thank you in advance for helping out.

On behalf of Erasmus University,
Samare Huls

A.2 Introductory Page

Dear participant,

Thank you for the taking the time to participate in this survey. The survey will last approximately 3² minutes. Good luck!

Kind regards,
Samare Huls
Erasmus University

² Based on a pretest, the estimated duration of the survey is 3 minutes for the short survey and 10 minutes for the long survey.

A.3 Long Survey

A.3.1 Page 1: Egoism Scale

Please rate the extent to which you agree with the following statements. You can rank items on a five-point scale ranging from “Strongly Disagree” to “Strongly Agree”.

1. The best way to handle people is to tell them what they want to hear.
2. A person should obey only those laws that seem reasonable.
3. Nowadays a person has to live pretty much for today and let tomorrow take care of itself.
4. It is hard to get ahead without cutting corners here and there.
5. It is alright to “bend” the law if you do not actually violate it.
6. It’s hardly fair to bring children into the world with the way things look for the future.
7. Never tell anyone the real reason you did something unless it is useful to do so.
8. A person should obey the law no matter how much it interferes with their personal ambition.*
9. These days a person doesn’t really know whom he can count on.
10. All in all, it is better to be humble and honest than important and dishonest.*
11. A person is justified in giving false testimony to protect a friend on trial.
12. Most people don’t care what happens to the next fellow.
13. Generally speaking people won’t work hard unless they’re forced to do so.
14. Laws are so often made for the benefit of small, selfish groups that a person cannot respect the law.
15. Next to health, money is the most important thing in life.
16. The biggest difference between most criminals and other people is that criminals are stupid enough to get caught.
17. It is alright for a person to break the law if he doesn’t get caught.
18. You sometimes can’t help wondering whether anything is worthwhile.
19. General rules about lying are useless; whether a lie is moral or immoral depends on the situation.
20. To make money there are no right and wrong ways anymore, only easy and hard ways.

* Items marked with an asterisk are reverse scored.

A.3.2 Page 2: Uncertain responses

1. When is the last time you helped someone to make yourself feel better?
 - a. This week
 - b. This month
 - c. More than a month ago
 - d. Don't know

2. When is the last time you helped someone to make another person feel better?
 - a. This week
 - b. This month
 - c. More than a month ago
 - d. Don't know

A.3.3 Page 3: Altruism Scale

Please select the category that conforms to the frequency with which you have carried out the following acts. You can rank items on a five-point scale ranging from “Never” to “Very Often”

1. I have helped push a stranger's car that was broken down or out of gas.
2. I have given directions to a stranger.
3. I have donated goods or clothes to a charity.
4. I have given money to a charity.
5. I have given money to a stranger who needed it (or asked me for it).
6. I have deliberately put waste in the wrong recycling bin.*
7. I have done volunteer work for a charity.
8. I have donated blood.
9. I have helped carry a stranger's belongings (books, parcels, etc.).
10. I have delayed an elevator and held the door open for a stranger.
11. I have allowed someone to go ahead of me in a lineup (in the supermarket, at a copy machine, at a fast-food restaurant).
12. I have given a stranger a lift in my car.
13. I have noted a clerk's error (in a bank, at the supermarket) in undercharging me for an item and kept it to myself.*
14. I have let a neighbor whom I didn't know too well borrow an item of some value to me (e.g., a dish, tools, etc.).
15. I have bought 'charity' holiday cards deliberately because I knew it was a good cause.
16. I have helped a classmate who I did not know that well with an assignment when my knowledge was greater than his or hers.
17. I have had a fatal heart attack.³
18. I have offered to help a handicapped or elderly stranger across a street.
19. I have offered my seat on a bus or train to a stranger who was standing.
20. I have helped an acquaintance to move households.

* Items marked with an asterisk are reverse scored.

³ This question is a control question for respondents' effort. Evidently, everyone should answer this question with 'Never'.

A.3.4 Page 4: Mathematical Question

1. What is $12 \times 7 \times 2 - 5$?
 - a. 185
 - b. 168
 - c. 163
 - d. Don't know

A.3.5 Page 5: Demographic Variables

1. What is your age?
 - a. [open]
2. What is your gender
 - a. Male
 - b. Female
 - c. Other
3. What is the highest level of education you completed?
 - a. Less than high school
 - b. High school
 - c. Intermediate vocational education (in Dutch: MBO)
 - d. Higher vocational education (in Dutch: HBO)
 - e. University
 - f. Other, namely [open]

A.3.6 Page 6: Relationship with experimenter

1. What is the relationship between you and the experimenter?
 - a. Strangers
 - b. Acquaintances
 - c. Good friends
 - d. Family members

A.4 Short Survey

The short survey is similar to the long survey. However, in the short survey there will be no questions about egoism and the two uncertain questions. This means that directly after the introductory page the respondent will be presented with page 4 of the long survey; page 2 and 3 will be skipped in the short version.

A.5 Reminder Email

Dear Sir, Madam,

This is a friendly reminder to participate in my research about altruism as part of my Master Thesis. It would be very helpful if you can fill out the survey. Please either follow the link to the Qualtrics survey or copy and paste the URL in your Internet browser to participate.

Follow this link to the Survey:

[Take the Survey](#)

Or copy and paste the URL below into your internet browser:

https://erasmusuniversity.eu.qualtrics.com/jfe/preview/SV_6kWHEPI3owwMd5H?Q_CHL=preview

Once again, thank you in advance for helping out.

On behalf of Erasmus School of Economics,
Samare Huls

Appendix B: Robustness Checks

B.1 Test for Match Familiarity Measures

To check whether familiarity reported by me is similar to familiarity reported by respondents (Self-reported Familiarity), these two measures of familiarity are compared. All respondents that stated that I am not a stranger to them, are considered to be familiar. Results show that the two measures of familiarity match in 155 of 170 cases. That is, 91.18% of the individuals I reported to personally know also indicated that they personally know me. Regardless of which measure of familiarity is used to analyse response rate and response quality, the statistical tests give similar results. Therefore, considering the relatively large amount of respondents that did not complete the survey up to the point where they had to report familiarity, familiarity reported by me is used in all analyses.

B.2 Tests for Demographic Differences between Versions and Conditions

To analyse if there are systematic differences between the two different versions and the two different conditions, several Mann-Whitney U test are performed. Differences in altruism levels, Age, Gender, Education, Student Association, Duration and Response Speed are compared across versions and conditions. At a 10% significance level, the Mann-Whitney U test fails to reject systematic differences in Altruism, Gender, Education, Student Association and Response Speed between the versions. For Age however, at 5% significance level there is some evidence that suggests differences between the long and short version of the survey. However, since the versions are assigned randomly by Qualtrics this result is not considered to be a problem. Duration also shows significant differences between the versions at a 5% significance level, but since one version is longer than the other this is evident.

In addition, the Mann-Whitney U test fails to reject systematic differences in Altruism, Age, Education, Duration and Response Speed between the familiarity conditions at any conventional significance level. Therefore, for the purpose of this research it is assumed that there are no differences in these variables. However, based on the Mann-Whitney U test, the null hypothesis of no differences in Gender and in Student Association between the different familiarity conditions is rejected at 5% and 10% significance level, respectively. This could be explained by the fact that I know more female individuals and more individuals from the student association that I am a member of myself. All Mann-Whitney U test statistics can be found in Table 12.

Table 12: Mann-Whitney U Test Statistics

	Version	Familiarity Condition
Altruism	1.530	-1.339
Age	2.002**	0.288
Gender	-0.751	-2.565**
Education	1.147	-1.500
Student Association	1.034	5.941***
Duration	2.577**	-1.359
Response Speed	-0.368	-1.530

*Significant at 10%.

**Significant at 5%.

***Significant at 1%.