ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Master Thesis in Behavioral Economics

Incentives for truth telling while testing the assumptions of choice-matching in student evaluation forms during the COVID-19 pandemic

Name student: Patrick Rego Titze

Student ID number: 447566pr

Supervisor: Dr. Tong Wang

Second assessor: Pr. Dr. Aurelien Baillon

Date finale version: 23.01.2021

Table of Contents

1. Introduction

To the question "why do you lie?", individuals often ashamedly and defenselessly reply "lying is a part of being human". This statement in fact was challenged by a BBC team back in 2018 in which three individuals tried to live lie-free for a week (Ramaswamy, 2018). The experiment which was, of course, doomed to fail, however, showed the seriousness of the issue at hand. Lying regardless of the context is part of being human, but in many cases, it only creates discomfort or becomes truly aware to us once it affects us in a monetary, ethical, or moral way. This thesis, therefore, explores the true well-being of students during the pandemic lockdown, while inducing honesty in survey questions regarding their academic process. It furthermore explores the possibility to strengthen the survey or questionnaire methodology used in many institutions which simply rely on their participant to be as honest as possible. Honesty in students' study process or beliefs has been difficult to account for in the past (Matteucci, 2014). One of the ways to counter for example the social desirability bias experienced in evaluation surveys is a truth mechanism. The bias often observed in these circumstances can be considered a form of lying.

Lying has a large economic impact in cases of asymmetric information, in which one agent decides to deliberately hide information and as a result, lead to a market failure. This market failure results in an overall lower social welfare, in which the transaction simply did not occur, or one party feels betrayed. This market failure can be often observed in the famous market for lemons, for example in cases of car dealers (Kessler, 2001). This classic and often used example demonstrated the power of honesty, in which a more honest seller of a used car could overall benefit from overall higher welfare simply being truthful. However, the same case can also be asked from buyers. This becomes clear in cases of insurances. If insurance companies would have full transparency of their customers, then most consumers would not have to pay a heavy fee (Wilson, 1977). It is simply the uncertainty in many cases that forces lower welfare and higher cost for many consumers.

In a very new but dangerous time such as the COVID-19 pandemic which affects everybody regardless of age, gender, or race honesty and consideration for others are of the utmost importance. This has not only been repeatedly praised by politicians but also by our own social norms. It is, therefore, important from the government level but also institution level to receive valid and honest information about the well-being of their people and members. When face-to-face interactions are not possible or constrained, it is important for researchers to find insights into the behavior of people. This all has to be explored through digital channels

and can therefore not be 100 percent controlled for. In a globalized and fast-moving world that we are living in, it is difficult to keep track of individuals, and therefore more often than not we have to take a word of someone as the truth. The pandemic we are currently facing came as a shock for all of us, and its effects can be observed and lived up till now. Since the outbreak of the pandemic which lead to a quarantine in many countries around the world, governments have actively worked in order to reduce the spread and consequently the capacity of our health system. One order which most governments in Europe such as Germany, The Netherlands, or Italy praised was the "stay at home campaign" ((RIVM, 2020), (bmg, 2020)). This was achieved by closing all non-essential buildings and institutions and sending their citizen back to their home. As a result, families and students have been restrained to their homes for weeks or months. Even though, many countries were successful in achieving a stop of high new infections, many scientists in May 2020 still warned of a second wave later in 2020 unless a vaccine would be provided (WHO, 2020). This change in our environment is something new which hasn't been explored yet by our generation, which makes new discoveries even more important for future generations. It therefore essential to gather enough information in order to be prepared for the future as COVID-19 might not be the last time.

The problem institutions or governments face is the level of honesty they receive when sensible questions are asked. Especially in a time such as the one we are facing negative behavior is often ignored or deliberately lied about. This can mainly be explained by social desirability bias in which people try to represent them more socially desirable in surveys. This behavior can hurt however struggling students who might then not receive the attention they require once the pandemic has slowed down. Even though Erasmus university has started surveys in order to get an overview of not only students but also staff members during this time ((Univeristy, 2020), (ErasmusMagazine , 2020)). These surveys did not use any incentivizing mechanism to identify if this was the true belief of students. They did, nevertheless, show the enormous change students experienced due to the isolations and digital working environment. Therefore, this is the reason a choice matching mechanism is used combined with more personal questions. As a result, this paper is the first to use the true incentivizing mechanism exploring not only the validity of evaluation procedures but also to give some insights on true learning behavior experienced by students during their quarantine.

Therefore, the main research question is "How is the COVID-19 pandemic affecting students' true learning behavior". This paper furthermore explores the use of "choice-matching" proposed by Cvitanić, Prelec , Riley , & Tereick (2019) an incentivizing mechanism

for evaluation forms or questionnaires. It also further investigates the assumptions set Cvitanić, Prelec , Riley , & Tereick (2019) and if they truly hold in the real world.

The contribution of this thesis is two-fold. First, it will shed more light into the well-being of students and their study behavior during the lockdown. It will also show how the lockdown affected students due to the constraints set by governments. Second, it will test the assumptions of the truth mechanism. Something most researchers take for granted when using truth-telling mechanisms is the assumptions set by them. Therefore, this thesis tries to investigate if the theoretical assumptions hold when used in a real-life application. Beyond showing the gap between experimental setup and real-life setting, this thesis will also inspire future mechanism design in order to increase the use for it in a commercial field. The rest of the thesis is organized as follows. Section 2 reviews the literature on the fundamentals of lying, elaborating the psychological aspects involved in lying. Section 3 explores choice matching and other truth inducing mechanism. Section 4 covers the methodology used in the experiment conducted in order to answer the research question. Section 5 presents the results and tests the assumptions of the choice-matching mechanism. Lastly, we will conclude with advice and guidelines for future institutions in order to be better prepared and to improve the studying environment for their students during this new age of digital learning and after.

2. Literature

2.1. The fundamentals of lying

Lying is an aspect of being human. Dictionary (2020) in fact defines a lie as “a false statement made with the intention of deceiving”. This act of deceiving can be often found in personal surveys but also in other situations. One of many examples can be found in a study conducted back in 2003, in which researchers estimated the cost of stealing office supplies to be 52 billion dollars annually (Pascual-Ezama, Jacobsen , & Fosgaard , 2018). This example shows the enormous impact of small dishonest behaviors. Therefore, behavioral economists and social psychologists have been working for decades on measures to reduce and explain this economic loss created on an individual and common level (Pascual-Ezama, Jacobsen, & Fosgaard, 2018). The explanation of lying itself has not been something new to the research world, but not much has been done on the causes and prevention of it. In fact, one definition of lying can be seen as a social interaction between a deceiver and a who has been deceived (Lindskold & S. Walters , 1983). Even though an accidentally untruth told statement may not be seen technically as a

lie, an internationally deceptive statement in which one party benefits can be. Which this paper built upon. The complexity in which dishonest behavior can be observed and influenced by not only external but also internal aspects makes this topic so broad and interesting. Especially companies and government institutions hope to find fast and cost-effective ways to receive more honest data. Therefore, this section will be elaborating on some of the literature and results collected in the last years on the act of "lying".

2.2. Why do we lie ?

This question has been asked by many researchers and philosophers for many years. As mention already would most people respond say that everybody lies and it's part of being human, however some studies such as from Halevy, Shalvi, & Verschuere , (2014) found a correlation between untruthful behavior and psychological tendencies. In fact, they claimed that most people actually tell the truth and only a minority lies every day. Nevertheless, in the context of this paper, we are not interested in the act of lying to your parents, partner, or friends, but in the act of deliberately falsifying information to be more socially desirable. This is referred to in behavioral economics as "social desirability bias". This bias is the main cause of distorting survey data (Krumpal, 2011). This as a result can influence future policy changes, consumer needs, or even a student's life as we know it. This is mainly due to the fact that the data provided could not be 100 percent reflecting the actual situation but rather a situation that is more socially preferred. However, before we disentangle the issue of what social desirability bias is and what factors play a role, we have to look at the aspect behind the act of lying itself. The social desirability bias is just one phenomenon that reflects dishonestly behavior.

While many theories have been developed to explain the dishonest behaviors of people, three models especially capture the potential process used by an individual. These three models can also be combined with falsifying deliberately surveys or questionnaires

The first and most basic one builds upon "expected utility theory", in which the subjects weigh the expected cost and benefits from lying or committing a crime (Pascual-Ezama, Jacobsen, & Fosgaard, 2018). In the case of questionnaires or surveys, we can expect due to anonymity that expected loss equals zero, whereas the benefits are collectively positive due to the personal image displayed and the task completed.

The second model goes one step further than the previous model and is referred to in the literature as the "self-concept maintenance theory ". It states that people will act dishonestly if the opportunity sets itself, but it will only be in a certain framework in which they maintain

their self-concept (Pascual-Ezama, Jacobsen, & Fosgaard, 2018). This framework works as a certain boundary which individual morals don't pass. Hence, in the case of filing in an outside survey, dishonest behaviors will not affect their moral balance too much.

The last theory builds referred to as "Bounded Ethicality and Ethical Blindness" uses ethical blindness in order to accept unethical behavior. Assuming that students are aware of the use of the survey for future purposes, many students might not see their behavior as unethical and therefore unconsciously act on it. This action is again strengthened by the anonymity set in many surveys and questionnaires, by ignoring the action it coexists with their self-interest (Pascual-Ezama, Jacobsen, & Fosgaard, 2018).

2.3. Social Desirability bias

Social desirability bias is referred to as one of the most cited reason for distortion of survey data in social science (Börger, 2012). The reason for the occurrence, however, is not due to the sensitivity of the question but the sensitivity of the answer (Krumpal, 2011). It is the social norms set in different societies or environments which set the social preferred view. In the case of university students, this is reflected in the name of the school and the expectation set by the institutions and peers. We therefore can expect students to change their self-presentation in the context of personal development, growth, and image by using social desirability which then affects the data collected in evaluations forms distributed in institutions (Matteucci, 2014).

In the paper by Matteucci (2014) previous studies on the changing behavior of people in front of an audience are touched upon. This behavior is observed in a person changing his roles, hence using a self-presentation strategy. This strategy consists of a person's effort to transfer a certain image or information about him or herself. This behavior of course is being manipulated in order to leave a positive impression or minimizing the social disapproval depending on the situation. However, in order for the strategy to be successful, individuals have to know what the audience is looking for. In the case of social desirability, the research by Matteucci (2014) claims that it is an affective value that makes people pleasant, socially attractive, or vice versa. The author also refers to social utility as the second approach in the dimension of self-presentation in which however the individual tries to meet the requirements of a particular social context in order to have the chance to be successful within this group or organization. However, in the context of the study process or development, the requirements are not set and the group, in general, does not find out. From another perspective on this theory Chung & Monroe (2003) claims that in social desirability bias, individuals hold two beliefs

that contradicts each others. These beliefs are built upon ethical evaluation in which they have beliefs about themselves and the other about their peers. These beliefs are indirectly used in this paper by letting people set probability the same way. In their paper, they claim that the belief about themself is more socially acceptable, hence participants tend to underestimate their use of drugs, alcohol and overestimate their use of given to charity or helping the poor. In the context of this paper, we would expect the same result without a truth mechanism.

While the act of lying itself is considered cognitively low as we expect to lie every day, students or participants of surveys still select specific answers which maximize their own personal value (Krumpal, 2011). This is mainly due to the shame and embarrassment observed by participants or potential fevering negative consequences from their answers (Krumpal, 2011). This is the main part that the choice matching methodology used in this paper tries to eliminate or overcome. The overall social norms set by institutions or classmates which reflect a productive lifestyle even during a pandemic might create discomfort for some people who have simply not been following this. These norms are not based on individuals but a social system that runs on social approval. Especially struggling students, which might have taken a less focused and cared view towards university duties might have given more freedom due to the lockdown without strict guidelines or preset protocols. This, therefore, leads to too more promising results in forms of productivity and well-being than actually observed or experienced. This false viewed given to institutions will influences the actions taken in the followed semester and not give students the help they truly need.

When revisiting the social desirability bias, it is important to keep in mind that it shouldn't be confused with other often experienced bias such as "acquiescence or yea-saying" (Börger, 2012). These biases are especially popular in the academic field in which an individual rather agrees than disagrees with a statement. Similar to the social desirability bias this bias is impulse and triggered by the personality and situation factors (Börger, 2012). It is the competitive and globalized world we live in in which "good is not good enough" attitude impacts students' and individuals' peoples' task every day. In the case of the COVID-pandemic institutions are as well interested in the personal development and growth of students. This is especially important as it cannot be personally seen or controlled for except in exams. In the paper of Bowman & Hill (2011) especially the social desirability bias within self-reported gains is highlighted. In which students overreport or estimated desirability behavior and vice versa desirability behavior such as cheating. As a result, students' evaluation forms show a misleading conclusion which influences the future implementation of intuitions.

3. Incentivizing truth-telling

Most of what we have learned from human behavior such as taste, preferences, or beliefs comes from self-reporting measures. Unfortunately, the basic human tendency is to present itself in its most desirable way (Fisher, 1993). Methodologies used, nevertheless, to take the truth out of people can already be followed back to the 15th and 16th centuries, excluding torture (Krumpal, 2011). Priests were already trained to follow protocol in order to find out the truth. Therefore, it is no surprise when studies in the past have been conducted on honesty, religious groups have always shown a higher level of honesty than non-religious groups (Arbel , Bar-El, Siniver, & Tobol , 2014). This claim is mainly built upon that religious groups hold higher morals and hence, therefore, a more reluctant to dishonesty. Therefore, methods such as the Bayesian truth serum (BTS) has been widely implemented in various forms in the field of inducing honesty. The methodology has since been expanded and modified in many papers, some of which will be elaborated on in this section. However, in the case of the survey conducted in this paper, a new approach has been used, referred to as the choice-matching mechanism. This new method, which has similarities to BTS is more appropriate in this context and has better criteria's in order for institutions to implement more.

3.1. Choice-Matching Mechanism

The question which still stands at this point is, what makes the usage of choice-matching so desirable for institutions compared to previous methods such as the Bayesian truth serum. Since this method is only truth-inducing, it is not able to identify false responses individually. The behavior method used in this process is by making participants play in a game in which a high "score" display potentially high honesty. This choice matching score is nevertheless only used as a reference and not to identify the actual level of honesty. Given the non-existing truth incentives in the current survey procedures, choice matching still in theory introduces a truth-telling incentive. Since it is published only recently, there have not been any empirical studies using choice matching, therefore its efficiency and variability compared to other methods still has to be explored. The paper, therefore, is the first in its kind and starting point for it.

The choice-matching mechanism has been originally introduced by (Cvitanić, Prelec , Riley , & Tereick , 2019). The method, similar to previous mechanisms can be used to elicit the true believe or honest behavior of certain actions. Each question is built upon two sub-questions. The first question asked the participant to choose an option that fits them the best. The second question asks them to set a prediction, in which the individual gives his best guess

on the distribution of each option given the sample of the survey. What is meant with the sample is, that the person needs to be aware of who else has received the survey, only then his guess can be reliable. In most cases, the sample will be constructed based on a classroom of students, a working station, or simply a target group that know each other. The individual of the survey then has the chance of winning a monetary compensation if his score is among the highest and is drawn from a lottery. This gives the individual a chance to be compensated for his effort. The score is computed using the actual percentages of each individual option and the prediction set by each participant. For the researcher or institution which has distributed the survey, the mathematical process is easy to compute and has the benefit that they don't require prior knowledge of the actual distribution. Furthermore, the sample does not have to be considerably high as the following section will show is an important aspect especially in research projects.

3.2. Bayesian Truth Serum

Past research as from John, Loewenstein, & Prelec (2011) or Weaver & Prelec (2013) have however used Bayesian truth serum in order to test the effectiveness in knowledge questionnaires and scientific misconduct in academic papers. The main idea behind the Bayesian truth serum is that its rewards individuals with a higher score if the answers is more common than the group collectively predicts and wise versa if the answer are less common (Weaver & Prelec , 2013). Similar to the choice matching mechanism are individuals asked to answer the options which fit them best or they believe to be true and set a prediction on the distribution in the sample. This final score referred as BTS score is calculated using an information and prediction score. Further information on the mathematics can be seen in Appendix 4.2. Whereas the information score is different in the choice-matching method, the prediction score is similar to computing parts of the choice-matching score. It is, however, important to keep in mind that a low score does not necessarily mean that the individual is dishonest, it simply reflects his or her beliefs, which are based upon expectation.

In the paper by Weaver & Prelec (2013) five separate small experiments were conducted in order to test the usage of BTS on knowledge questions such as historical name, authors, electronic or life sciences. All of the experiment showed promising result in which participants were exposed to the BTS recognized fewer nonexistent foils than participants who were simply asked to be honest. On top of being more reluctant of admitting to knowing a foil, students also spent more time on the questions. All experiments were conducted on a PC by

university students. The BTS in fact eliminated some untruthfulness, nevertheless, it also introduced concerns on scoring surveys. The authors claimed that changing the nature of the questionnaire or survey to a scoring system in fact gives participants the chance to learn in advance and hence in theory score better if they are prepared. The creation of a score therefore introduced competitiveness and reputational stakes but also filter the thoughtful responses. The paper by Howie, Wang , & Tsai (2011) in fact used the mechanism on a company level in which they successfully extracted the true belief of adopting a new product line. They found that a modified usage of the BTS improved the prediction performance by 36%. The paper by John, Loewenstein, & Prelec (2011) showed that using truth mechanism actually increased the number of scientific misconducts done by researcher from a list of ten items. The BTS group compared to the control group showed a higher admission rate for "more questionable" behavior, however showed similar rates to less questionable behavior. All control groups in the previously mentioned experiments simply were asked to be truthful. In the paper by John, Loewenstein, & Prelec (2011) participants on top of the mentioned prediction also had to give one extra information about the "admission rate", which was used to predict the participants who "would actually admit" to have done scientific misconduct. This extra information was used to further understand the prevalence distribution of the different options. However, none of these past papers have used the mechanism in order to test the validity of evaluation forms or questionnaires. Therefore, many surveys send by universities or companies simply take the word of their participants as a given. However, the truth mechanism not only improves the lying aspect of surveys but also has shown that participants have taken the act of participating in a survey more seriously.

As mentioned in the previous section is the mathematical approach to identifying more honest responses very similar between the choice-matching and the often-used BTS. However, the choice matching approach compared to the BTS score is easier in the understanding and execution. Participants in both methods decide if they agree with the statement and what they expected their fellow participant's overall distribution to be across the options. The difference to the BTS method lies in the matching of the score to fellow participants who are within the group rather than the overall group prediction. More detail on the difference can be found in appendix 4.2 which shows that the prediction score part of BTS is similar to the error term in choice matching and hence a perfect prediction which in the BTS is referred to as $\bar{y}_k^r = \bar{x}_k$ gives a zero for both the error term in Choice-matching and prediction score in BTS.

The choice matching method in itself compared to other methodologies gives the experimenter an overall easier way to elicit honest responses. The previous study from Weaver

& Prelec (2013) has mentioned the difficulties for participants to understand the BTS methodology. Therefore, choice-matching gives participants more transparency within the survey or questionnaire they are participating in, and hence they have a stronger relationship to the validity of responses they give. Furthermore, the mechanism as mentioned already can induce honesty with a very low response rate compared to other Bayesian truth serums which require a big data set in order to become more valid. In fact, the mechanism only requires a minimum number of participants of N> M+2 where M reflects the number of possible responses (Morgan R., Cebrian , Pickard , & Rahwan , 2017). Hence, the method also doesn't require all participants to answer prediction questions. This as a result reduces the effort for participants. This becomes especially important, as people often lose concentration and effort once the tasks sum up. The Bayesian Truth Serum on the other hand only becomes effective once a large sample has been constructed. Overall is the choice-matching a more transparent and easier approach to implement for both the researcher and participant.

3.3. Assumptions in the choice matching mechanism

The choice-matching mechanism is based upon four assumptions. These assumptions are sufficient but not essential in order for choice matching to be having an honest inducing effect (Cvitanić, Prelec , Riley , & Tereick , 2019). Each assumption will be tested if it holds in the experiment conducted in this paper. The importance of this testing lies in the often-easy approach taken by researchers towards assumptions set by them. Mechanism or formulas are often tested or controlled for in safe and assumption-oriented fields, however, not much has been done to prove if these assumptions truly hold when used in a real-life situation.

The first assumption is the "common prior" assumption, which assumes that individuals participating in the survey or questionary are aware of the people for which the prediction are based upon. It does not assume that the person conducting, in the case me knows the exact distribution but that students need to have some prior knowledge. This assumption in the case of this survey can be assumed to hold due to the fact that all participants were taken from a friend bases, therefore a common prior knowledge did exist. Nevertheless, as will be pointed out in more detail in the limitation section, this assumption would have been stronger if the pool of subject would have been taken from the same academic year and same course.

The second assumption establishes the need that each respondent is fully aware that each option within a question is represented in the survey and has a chance of receiving an honest endorsement. Consequently, this puts a constraint in the survey which says that the

answer options cannot be lower than the number of responses. This is not the case the in the survey constructed. Each answer options have been careful reviewed and adapted based on similar survey set by institutions, this further strengthen the possibility that each option can be represented in the survey.

The third and most common assumption in this field is the “stochastic relevance”. The assumption in respect too choice-matching requires that the prediction of different types is not equal to each other. However, the assumption will be violated if the distribution of the variable or characteristics is common knowledge. As it requires that individual must have different prediction for an answer option if they response differently. Mathematical seen it can be written as follow: $p^{r,k} \neq p^{s,i}$ where p reflects a prediction, k and i reflect an answer options s and r reflect individuals. An example for this in relation to this paper would be, knowing the exact distribution of people who have failed an exam. Given the nature of the question asked which a very general, you can be certain that the distribution is not common knowledge.

The fourth and last assumption, which is referred as “Impersonal Updating” is the most demanding assumption. ". It states that the prediction of individual is strong correlated to their personal characteristics, hence individual which answer the same type (answer options) will have the same prediction. Mathematical seen it can be shown as follow: $p^{r,k} = p^{s,k}$. Similar to the previous assumption, you can see that now the prediction need to be similar for individuals (r and s) answering the same option (k).The last assumption of course brings difficulties with it, as its difficult to imagine that individuals truly have the same prediction about an answer options. The paper by Cvitanić, Prelec , Riley , & Tereick (2019) is fully aware of this and proofs that if the assumption is approximately true it is sufficient. Both assumptions, nevertheless, will be tested in the next section.

4. Methodology

The following section will be elaborating on the experiment conduct in which incentives truth mechanisms were used to receive more truthful responses. Therefore, this section will be covering the data collection aspect, the experiment distribution, and experiment construction. It will also be outlining methods to facilitate future research, who plan on using the same methodology in evaluation systems. The first study examining the learning behavior of students during the pandemic has already shown a 50% reduction in study time dedicated to school duties during the lockdown, mainly due to the lack of interactions and control (Institut, 2020). Further research conducted by Univeristy (2020) early in the pandemic has shown similar changes especially in the stress and pressure level toward students and members.

When it comes to the environment in which students had to study and learn during this pandemic, the most important change was observed in the intensity of digitalization. Institutions all around the world were challenged to adapt to a new way of mediating the necessary information to continue the study. Many factors that impact the learning behavior of students have been identified in past research. The list of conditions, environment, social heritance, and personality which influence the learning behavior of students have been extensively covered in the past in paper such as from (Entwistle & R. Peterson , 2004). In which especially the development of students learning behavior is emphasized. The paper by Entwistle & R. Peterson (2004) states that students in later years in their study become more resistant to outside factors compared to study beginners. Therefore, especially students which have only recently entered the university phase or are in the early years have the biggest struggle to face. Furthermore, does the usage of digitalization create a much more distracting field, which someone would not experience during a standard lecture.

The survey also which will be presented also tackles the psychological aspect of the lockdown, faced by the monotonicity. A student-life is often also built upon a strong work-life balance which not only companies often built upon but also institutions. Unfortunately, the usage of digital interactions plus social distance at a global level has never been experienced or researched before, therefore the uncertainty of the true effect on students is yet to come.

4.1. Survey Design

The survey consisted of three parts. Part one consisted of small explanations of the survey setup and the reward calculation, and it also included the motivation of answering the survey truthfully. The second part identified the subject's demographic characteristics such as age, nationality, gender, and level of study. The third part which was the main part of the survey asked eight different questions with eight follow-up prediction questions. Each question was used to identify actions, experiences and beliefs of the pandemic on the study behavior of the subjects. In total eight questions which are displayed bellowed have been asked, more specific information can be found in Appendix 2.

Q1. Please indicate the number of courses you have failed in this academic year 2019/20
Q2. Do you believe, that due to the Coronavirus, you will be at a disadvantage compared to students of previous years?
Q3. Have you been consuming more intoxicating substances like alcohol or marijuana, compared to the situation prior to the Coronovirus outbreak?
Q4. What seemed to be the biggest struggle when faced with the home-office lifestyle? Please indicate which option has affected you the most.
Q5. Which statement best describes your work time dedicated to your weekly school duties compared to a school week prior to the Corona outbreak?
Q6. During the pandemic, specifically the lockdown, did you have serious confrontations with roommates or family members.
Q7. Have you taken advantage of online examination in a way which can be considered cheating due to the lack of control normally experienced in an exam at the University campus.
Q8. Response rates for course evaluations or school surveys are typically low. Why do you think that is the case ?

Table 1: Survey Questions identifying students learning behavior and factors

Question one is used to detect “struggling" students from the sample. These struggling students are classified by answering that they have failed one or more exams this year. Question three and five are used to identify if students have been actively consuming more intoxicating substances and hence have maybe reduced their attention to the work duties. Question four and six are used to identify the work environment confrontations students had to experience during the lockdown. Question two is used to identify if students felt an overall difference due to the pandemic. Question seven is specifically constructed for the Dutch system in which any cheating behavior has been observed. Question eight rounds up the reason for the ongoing struggle of researchers such as me facing low response rates.

4.2. Incentives methodology

The incentive created in the experiment were communicated throughout a detailed explanation of the choice-matching mechanism and the potential payout for the participants. Students were asked to give their best prediction of the distribution of different questions. These predictions should be as honest as possible. A more honest response and hence a better prediction score would give them the possibility to enter a lottery which would reward them with monetary compensation. Compared to the previous approach of intermediation or the more complex BTS scoring, a better understanding of the process and methodology would give students a better feeling of actually winning the reward. Students had the chance prior to starting the survey to read a detailed explanation of the mechanism and task requested. A detailed explanation can be found in Appendix 2.

The honesty score of each individual can be estimated using four steps. Step one is identifying the "true" distribution of types within the sample. Step two is calculating the prediction score which is mainly computed using the difference between the prediction and the true distribution. This error is subtracted from any given number which in this case is 10000. This number however can be higher or lower. Hence the closer you are to 10000 the better your prediction is to the true percentage. Step three is used in order to compute the matching score. This approach averages the prediction score of different types. A more detailed explanation of the mathematic can be found in Appendix 4.2.

5. Data analysis and Results

This section reflects on the responses given in the survey. It will shed more light into the behavior of students during the lockdown and potential confrontation they have experienced. It will further explore limitations and potential improvements for future research projects in this field. All responses are displayed graphically in Appendix 1. The main data collection was done on students currently attending a university in the academic year 2019/20. The survey was distributed from Bachelor 1 to Master students. Participants were asked throughout social media platforms such as Facebook and LinkedIn. In total, 36 participants responded. The sample consists of 39% female and 61% male. The age of participants ranges from 18-26. The most representative population is German (39%) and Master students (50%). The survey itself was distributed between 3 to 4 weeks after the housing lockdown requirements where lifted in the Netherlands, Germany and Belgium.

5.1. Collective result

Figure 1 below demonstrates that most students who have taken the survey in fact have not failed any exams. As a result, all followed up questions are built upon students who were academically stable in the sense of not experiencing any re-sits. The second highest responses, however, are students who have failed two exams. This very high responses of students which have failed no exams can be to some extent be explained that they are master students, which only have studied for one year. Therefore, these students have better adapted to their learning behavior compared to first year or second year students as the study by Entwistle & R. Peterson (2004) showed.

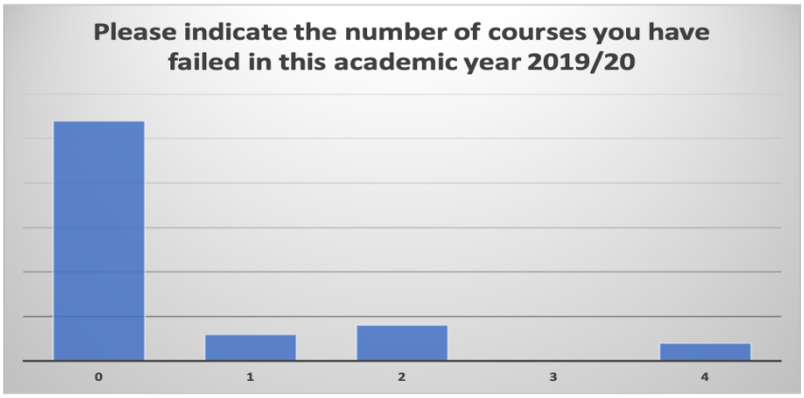

Figure 1: Question 1 of Survey

Even though, most students believe the coronavirus has given them a disadvantage compared to previous-year students as can be seen in Figure 2, the number of students who don't believe this is also almost equal. This shows a surprising result. Some would expect a high majority of people to answer yes due to this very strange and new experience.

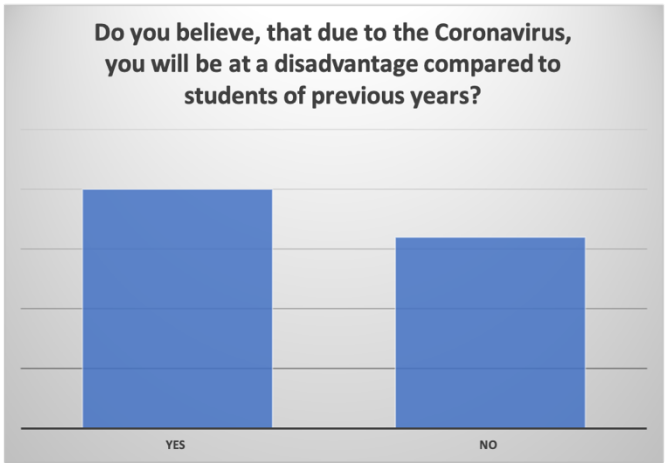

Figure 2: Question 2 of Survey

Figure 3, which identifies if student have taken more intoxicating substances, shows that the majority of students have answered no. One must not forget though that the average score in this question compared to other questions in the survey is the lowest, which show some sort of disagreement when prediction where set.

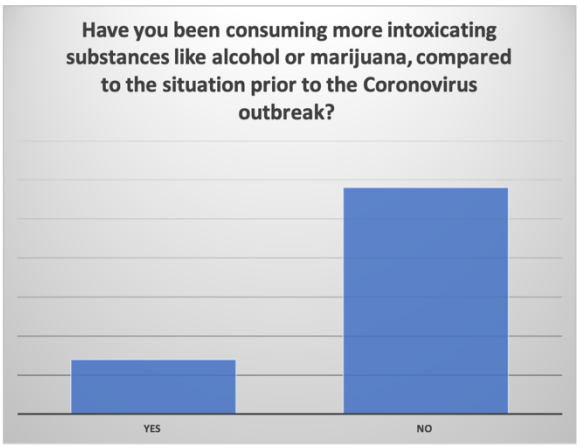

Figure 3:Question 3 of Survey

When it comes to the work experiences, as can be found in Figure 4, the lockdown has created an no time constrained which most students found struggling. This result was followed by an equal split of privacy and adapting to online classes. This result shows that students faced more psychological issues rather than technical issues. For the future, therefore, university have to take this into account when creating a teaching schedule. Privacy caused by roommate is difficult to change and also requires adaptation and communication.

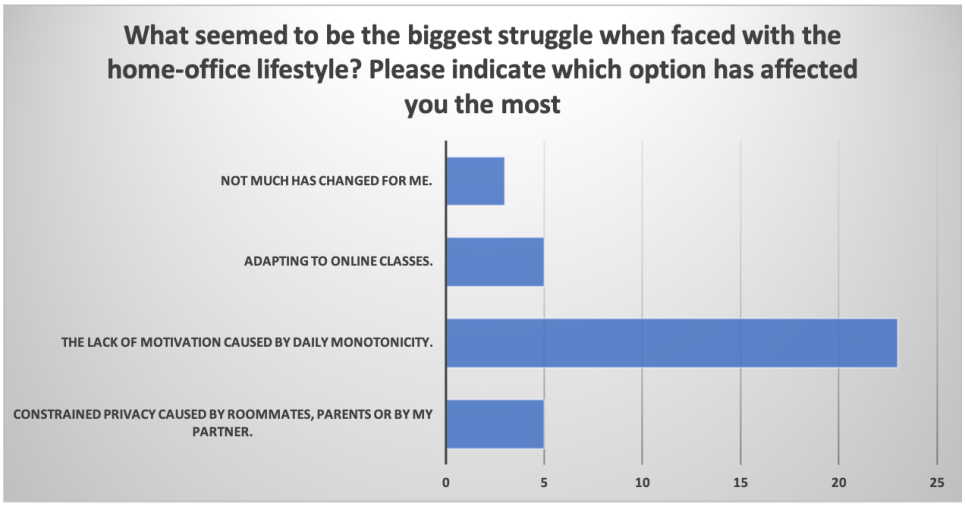

Figure 4: Question 4 of Survey

When asked about the work spent on duties, the biggest group answered that they have in fact worked more as can be seen in Figure 5, however just with one-point different students also have worked less. On third place students answered they have not changed their routine. This result again shows an interesting behavior when it comes to students' work ethics. While

most online sources demonstrated an increase in the work pressure such as from Univeristy (2020) this small survey showed inconclusive results.

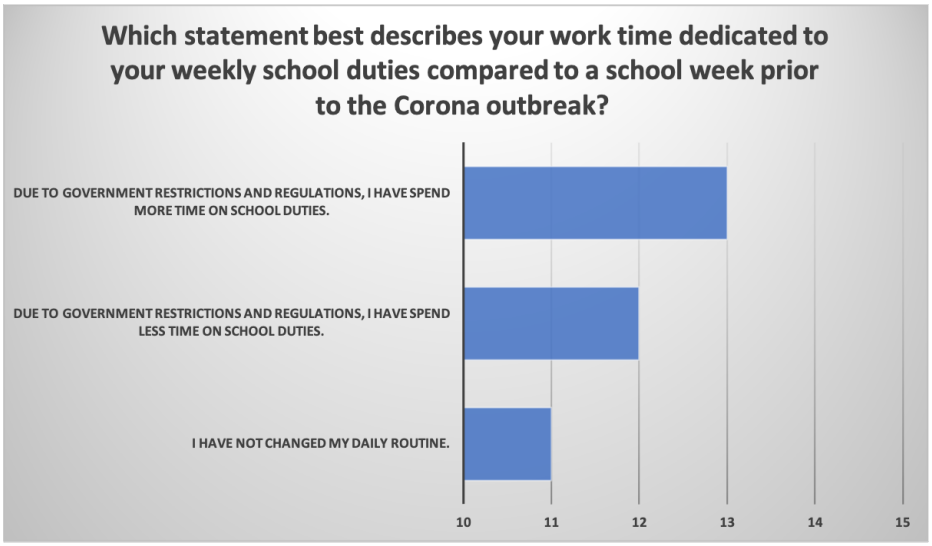

Figure 5:Question 5 of Survey

When asked about confrontation within the friends or family circle, most students were lucky enough to experience none. Some might think, however that some sort of confrontation is possible as a study from Germany in fact showed some small increase in violation during the lockdown (Steinert, 2020). Being constrained by four walls for a longer period of time can create after some time often some disagreement.

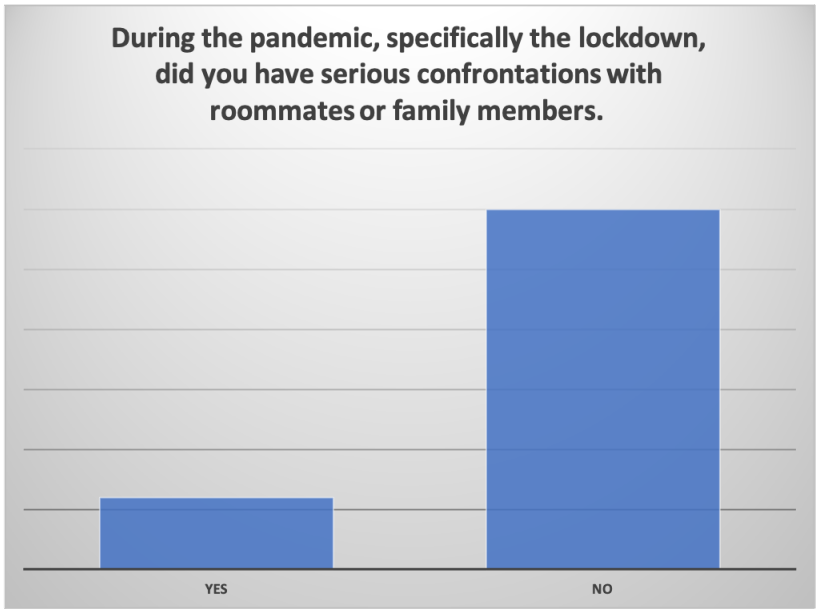

Figure 6:Question 6 of Survey

A more surprising result was concluded for students in the Netherlands, which accounted for 27 of the total responses collected. As can be seen in Figure 7, a surprising high number of students admitted having "cheated" in some way in the online procedure. This result is surprisingly high given this ethically unaccepted behavior. This cheating mostly arises from the change in examination procedure. Institutions such as Erasmus University and other universities should work closely with course coordinator and private companies in order to not only create a more representative but also fair exam for everybody.

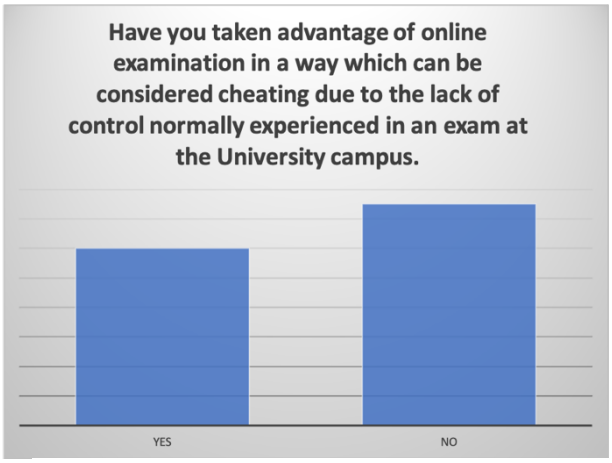

Figure 7:Question 7 of Survey

The final question which was used to deepen the insights into the general low response rates in surveys or questionnaires, whose result is shown in Figure 8, showed that time was not the problem. But rather students felt no incentives was given and that they felt mostly distanced to the survey itself. For future surveys, this means that researchers should find ways to improve the bridge between participants and the projects conducted. Students or survey-takers need to get the feeling that they are not just another number but that their input is very much valued.

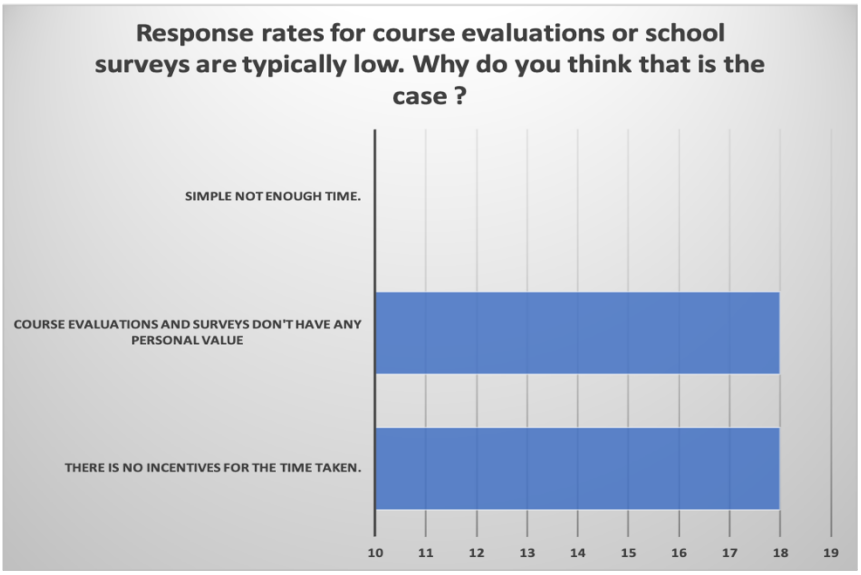

Figure 8:Question 8 of Survey

Finally, in total a score of 80000 could have been achieved in the survey of this paper, for which the average 63354 was conclude. Taken away question seven which was only asked to students actively in the Dutch system an average score of 54827 out of 70000 was concluded. It is furthermore to keep in mind that a higher score does not necessarily mean a higher honesty. It shows rather a more understanding of the true distribution, whereas a low score can signal less effort or understanding. Further detail on the score distribution between questions and participants can be found in Appendix 4.1.

5.2. Testing Assumptions

The third assumption tested in this paper is displayed in the table below. The referred p-values for each question have been constructed by the usages of a Mann-Whitney U test and a Kruskal-Wallis test. Both methods were used to identify similarities between the samples of prediction given a certain answer option (Type).

Survey Questions	**P-Value**
Question 1	0.0158
Question 2	0.0020
Question 3	0.1608
Question 4	0.2227
Question 5	0.0096
Question 6	0.0374
Question 7	0.0919
Question 8	0.0000

Table 2: Testing if prediction set given different types are similar

For questions 2,3,6, 7, and 8 a Mann-Whiney U test was used due to the binary nature of the answer options, which consisted of Yes or No. For the question 1, 4 and 5 a Kruskal-Wallis test, this is due to the fact that these questions have more than two types of responses. However, it is important to keep in mind that when testing for assumption three the following $H_0: \theta_1 = \theta_2 = \theta_3 = \cdots = \theta_k$ was set. Hence in order for the assumption to be true, we would accept a low p-value and reject any p-value higher than 5%. This 5% level can vary however the smaller the value the stricter the assumption. Furthermore, even though Question 8 consist of three answer options, the split of responses was equally distributed only for two answer options. As a result, for question 8 a Mann-Whiney U test was also used.

Table two above demonstrates that the majority of questions aligned with assumption three. All questions except 3 ,4, and 7 have a very low p-value or close to zero, hence the probability that the predictions mean between samples is the same is can be rejected on a 5% level. The reason questions 3,4 and 7 are experiencing a high p-value can mainly be explained by the very uneven distribution of responses. More detailed results can be found in appendix 5.1.

The fourth assumption explores if there are any similarities between prediction given the same type. In order to test the validity of assumption four, histograms were constructed to display the prediction distribution within each type. These histograms are displayed in appendix 5.2. Even though the response rate is fairly low which for some questions made the conclusion difficult to construct, some questions such as 8 show a clear pattern in the prediction set by each type. It can be observed the most responses in question 8 range within a 20% difference between each other. For the other question, the same pattern can be seen, however often only in one type such as in questions 2,3, 4 and 6. In these questions on type often the one with a majority of responses showed similar predictions. Hence, some can expect once the response rate becomes higher and equally distributed between answer options would lead to assumption four to hold stronger. This conclusion is similar to the paper by Cvitanić, Prelec , Riley , & Tereick (2019) who stated that the prediction doesn't have to be exactly the same for the method to work.

5.3. Limitations/Suggestion

One of the most struggling aspects of this thesis project was creating a prediction population. In other words, creating a group that was familiar with which the prediction question could be based on. Universities or future institutions should, however, not struggle too much in this aspect as they have the resources and classes for which each survey can be sent to. Furthermore, like in many other surveys conducted by students is the response rates were very low, this could be reduced in future experiments with a higher incentive given, this would give students a high push to not only take the prediction question serious but also respond more. Lastly, the question asked on the questionnaire were to some extent very general. Future research studies would benefit from a more personalized survey which is either based on the specific course, product, or duties participants have experienced. In the sample used in this survey, this would mean beliefs and well-being created by Erasmus University and German universities. Erasmus currently conducts surveys with the chance of winning often 50 euros by chance if the survey in hand is completed. Therefore, a simple change of the survey distributed to students would improve honesty received by students without creating more costs. The original experiment planned can be found in Appendix 3, which elaborates on a more controlled environment. Unfortunately, due to the stress experienced by many professors, it was not able to be done.

Lastly, future researchers could create several groups in order to test the efficiency between the BTS, choice-matching, or simple traditional approach. In that way, a more transparent comparison could be created between the different approaches.

Conclusion

This thesis has provided several deeper insights into the effect of the COVID-19 pandemic on student's study behavior. Furthermore, it has provided a setting in which the choice-matching method could be tested and validated. The most obvious finding to emerge from this study was that students did experience a significant change in their study behavior, especially on the psychological level. Even though the confrontation and struggle with roommates or family were fairly low, the monotonicity created by the restrictions was the biggest obstacle. Many students are aware that this pandemic will have consequences for their future carrier, however, the situation has given them new skills to cope with not only the online classes but also the constrain of staying at home. To a surprise for institutions such as Erasmus was the cheating behavior fairly high which of course shows that most students felt insecure about the study progress they have done. The thesis as well shaded light into the well-known phenomena of "social desirability bias "and its frequent occurrence in especially development and process surveys, something truth mechanism can minimize.

The study as well provided a perfect set up to test the assumptions set by Cvitanić, Prelec , Riley , & Tereick (2019) in a real-life survey. Especially assumption three seemed to hold even for a very low response rate. Assumption four, however, showed some irregularity. This was mainly due to the majority of responses focused on one answer option. Nevertheless, questions 2, 3, 4, 6, and 8 showed some pattern. This aligns with the statement set in the original paper by (Cvitanić, Prelec , Riley , & Tereick , 2019). As a result, for future research and surveys, it's important not only to receive more responses but also to slowly begin implementing methods such as the Choice-matching in order to cheap and efficiently receive more honest responses. This is especially the case with respect to sensitive questions such as mental health or drug abuse. The well-being of students is especially important to be followed during and after this pandemic and in order to implement target group-oriented solutions such as for academically weak but also sensible students. The future is still uncertain and the time we will return to normality still seems unknown, nevertheless, universities have tried the most to not only act fast but efficient to this erupt change. It is now in the hand of government institutions to provide the necessary tools to combine it with the knowledge of schools to provide a safe and promising environment for current and future students.

References

Arbel , Y., Bar-El, R., Siniver, E., & Tobol , Y. (2014). Roll a die and tell a lie – What affects honesty? *Journal of Economic Behavior & Organization* , 153-172.

Börger, T. (2012). Social Desirability and Environmental Valuation. Frankfurt am Main : Peter Lang .

bmg. (2020, July 23). *Informationen zum Coronavirus*. Retrieved from bundesgesundheitsministerium: https://www.bundesgesundheitsministerium.de/coronavirus.html

Boring , A., Ottoboni , K., & B.Stark , P. (2016). Student evaluations of teaching (mostly) do not measure teaching effectiveness. *ScienceOpen Research* .

Bowman, N. A., & Hill, P. (2011). Measuring How College Affects Students: Social Desirability and Other Potential Biases in College Student Self-Reported Gains. 73-85.

Chambel, M. J. (2005). Stress in Academic Life: Work Characteristics as Predictors of Student Well-being and Performance. *Applied Psychology* , 135-147.

Chung, J., & Monroe , G. (2003). Exploring Social Desirability Bias . *Journal of Business Ethics*, 291-302.

Cvitanić, J., Prelec , D., Riley , B., & Tereick , B. (2019). Honesty via Choice-Matching. *AER: Insights*, 179-192.

Dictionary, C. (2020). *Lie*. Retrieved from Cambridge Dictionary : https://dictionary.cambridge.org/dictionary/english-german/lie

Drouin, M., Miller , D., Wehle , S. M., & Hernandez , E. (2016). Why do people lie online ? " Because everyone lies on the internet". *Computers in Human Behavior* , 134-142.

Entwistle, N. J., & R. Peterson , E. (2004). Conceptions of learning and knowledge in higher education: Relationships with study behaviour and influences of learning enviroments . *International Journal of Educational Research* , 407-428.

ErasmusMagazine . (2020, May 18). Retrieved from High levels of study-related stress and work pressure throughout the corona crisis: Working and studying from home creates pressure and stress among staff and students of Erasmus University

Fisher, R. J. (1993). Social Desirability Bias and the Validity of Indirect Questioning. *Journal of Consumer Research* , 303-315.

Halevy, R., Shalvi, S., & Verschuere , B. (2014). Being Honest about Dishonesty: Correlating Self-Reports and Actual Lying . *Human Communication Research* , 54-72.

Howie, P. J., Wang , Y., & Tsai , J. (2011). Predicting new product adoption using Bayesian truth serum. *Sage* , 6-16.

Institut, I. (2020, August 3). *ifo Institut: Coronakrise halbierte bei Kindern die Zeit für die Schule*. Retrieved from Ifo Institut: https://www.ifo.de/node/57298

John, L. K., Loewenstein, G., & Prelec, D. (2011). Measuring the Prevalence of Questionable Research Practices With Incentives for Truth Telling . *Pychological Science* , 524-532.

Kessler, A. S. (2001). Revisiting the Lemons Market. *International Economic Review* , 25-41.

Krumpal, I. (2011). Determinants of social desirability bias in sensitive surveys: a literature review . *Springer Science+Business Media* , 2025-2047.

Lindskold , S., & S. Walters , P. (1983). Categoeries of Acceptability of lies . *The Journal of Social Psychology* , 129-136.

Matteucci, M. C. (2014). Social utility versus social desirability of students' attributional self-presentation strategies . *Springer* , 541-563.

Meyer, J. (1995). Gender-group differences in the learning behaviour of entering first-year university students . *Higher Education* , 201-215.

Morgan R., F., Cebrian , M., Pickard , G., & Rahwan , I. (2017). Validating Bayesian truth serum in large-scale online human experiments. *PLoS ONE*.

Pascual-Ezama, D., Jacobsen , C., & Fosgaard , T. R. (2018). Why do we lie ? A practical guide to the dishonest literature. *Journal of EconomicSurveys*, 357-387.

Ramaswamy, C. (2018, August 29). *Horizon: A Week Without Lying – The Honesty Experiment review – a social experiment that was bound to fail*. Retrieved from The Guardian : https://www.theguardian.com/tv-and-radio/2018/aug/29/horizon-week-without-lying-review-social-experiment-honesty

RIVM. (2020, July 23). *National Institute for Public Health and the Environment Ministry of Health, Welfare and Sport*. Retrieved from Current information about the new coronavirus (COVID-19): https://www.rivm.nl/coronavirus-covid-19/actueel

Sheard, J., Ceddia , J., Hurst , J., & Tuovinen , J. (2003). Inferring Student Learning Behaviour from Website Interactions: A Usage Analysis . *Education and Information Technologies* , 245-266.

Steinert, P. D. (2020, June 2). *Domestic violence during the coronavirus pandemic* . Retrieved from TUM : https://www.tum.de/nc/en/about-tum/news/press-releases/details/36053/

Univeristy, E. (2020, May 8). *Working and studying from home creates pressure and stress among staff and students of Erasmus University*. Retrieved from Erasmus Univerisity Rotterdam : https://my.eur.nl/en/eur/news/working-and-studying-home-creates-pressure-and-stress-among-staff-and-students-erasmus-university

Weaver, R., & Prelec , D. (2013). Creating Truth-Telling Incentives with the Bayesian Truth Serum . *Journal of Marketing Research* , 289-302.

WHO. (2020). *Preparing for the second wave: lessons from current outbreaks*. Retrieved from World Health Organization: https://www.who.int/csr/disease/swineflu/notes/h1n1_second_wave_20090828/en/

Wilson, C. (1977). A model of insurance markets with incomplete information. *Journal of Economic Theory* , 167-207.

Appendix

1. Survey sample descriptive summary

1. Gender	Freq.	Percent	Cum.
Female	14	38.89	38.89
Male	22	61.11	100
Total	36	100.00	

Table 3. Gender distribution

Nationality	Freq.	Percent	Cum
Belgium	3	8.33	8.33
Dutch	3	8.33	16.67
French	2	5.56	22.22
German	14	38.89	61.11
Others	14	38.89	100.00
Total	36	100.00	

Table 4. Nationality distribution

Studylevel	Freq.	Percent	Cum.
1 Year Bachelor Students	4	11.11	11.11
2 Year Bachelor Students	3	8.33	19.44
3 Year bachelor Students	11	30.56	50.00
Master Students	18	50.00	100.00
Total	36	100.00	

Table 5. Study level distribution

Mean estimation			**Number of obs** = **36**	
	Mean	Std. Err.	(95% Conf. Interval)	
Age	22.63889	.3987138	21.82946	23.44832

Table 6. Age distribution

2. Survey

Start of Block: Description

Q1 Dear Participant, Thank you for taking the time to answering this survey. This survey is intended to shed more light into the learning behavior of students during this new pandemic.

The survey will take approximately 3 minutes and we would appreciate if you could answer the survey as honestly as possible. We will be using a so called choice matching mechanism to reward honest responses. This mechanism has shown promising results in past research. This method has been developed by researchers at Caltech, MIT and Erasmus University Rotterdam.

The purpose of this method is to encourage participants to pay close attention to the questionnaire. Using this method, we will calculate your payoff based on your responses to questions 1 to 8 in this questionnaire. In each response, you select an answer option and make a prediction. Please keep in mind when answering the prediction question that the options have to add to 100% together. Also your predication should be based on other students who study in the same year as you.

Q2 Your score is determined via two factors: - Firstly, you receive a score for how good your prediction was. - Secondly, you receive the average prediction score of other respondents that selected the same answer option. At the end of the survey, a unique code will be published which you have to keep for yourself. Once all responses have been evaluated, my colleagues and I will be publishing the table with the score rank. You can compare your score to others. You will also be receiving a second code below the first one. This code is a private in order to receive the reward if you win the lottery. The survey is 100% anonymous: the random code can not be linked to you.

The top 25% of respondes with the highest score will be entering the lottery in order to have the chance of winning a price. This price is monetary. This reward can also be donated to a charity of your choice. The reward will be 15 euros. Therefore, next to the scoreboard my colleagues and I will be publishing the winning code. Thank you again for your help and good luck!

Start of Block: Demographic characteristics

Q3 Please indicate your gender

- Male (1)
- Female (2)

Q4 Please indicate your age

_____

Q5 Please indicate your nationality

- ○ Dutch (1)
- ○ German (2)
- ○ Belgium (3)
- ○ French (4)
- ○ Others (5) _____

Q6 Please indicate your level of study

- ○ 1 Year Bachelor Students (1)
- ○ 2 Year Bachelor Students (2)
- ○ 3 Year Bachelor Students (3)
- ○ Master Student (4)
- ○ Phd (5)

Q8 **Q1**. Please indicate the number of courses you have failed in this academic year 2019/20

- ○ None (7)
- ○ 1 (8)
- ○ 2 (9)
- ○ 3 (10)
- ○ 4 or more (11)

Q9 In the previous question, which percentage of respondents chose each answer option (not including your own answer)?
_____ None (1)
_____ 1 (6)
_____ 2 (2)
_____ 3 (3)
_____ 4 or more (4)

Q10 **Q2**. Do you believe, that due to the Coronavirus, you will be at a disadvantage compared to students of previous years?

- ○ Yes (1)
- ○ No (2)

Q11
In the previous question, which percentage of respondents chose the answer option "Yes" (not including your own answer)?

	0 10 20 30 40 50 60 70 80 90 100
Yes ()	

Q12 **Q3**. Have you been consuming more intoxicating substances like alcohol or marijuana, compared to the situation prior to the Coronovirus outbreak?

- ○ Yes (1)
- ○ No (2)

Q13 In the previous question, which percentage of respondents chose the answer option "Yes" (not including your own answer)?

	0 10 20 30 40 50 60 70 80 90 100
Yes ()	

Q14 **Q4**. What seemed to be the biggest struggle when faced with the home-office lifestyle? Please indicate which option has affected you the most.

- ○ Constrained privacy caused by roommates, parents or by my partner. (1)
- ○ The lack of motivation caused by daily monotonicity. (5)
- ○ Not much has changed for me. (2)
- ○ Adapting to online classes. (3)

Q15 In the previous question, which percentage of respondents chose each answer option (not including your own answer)?
_____ The constrained for privacy caused by roommates, parents or partner (1)
_____ The lack of motivation caused by daily monotonicity. (2)
_____ Not much has changed for me (3)
_____ Adapting to online classes. (5)

Q16 **Q5**. Which statement best describes your work time dedicated to your weekly school duties compared to a school week prior to the Corona outbreak?

- I have not changed my daily routine. (1)
- Due to government restrictions and regulations, I have spend more time on school duties. (2)
- Due to government restrictions and regulations, I have spend less time on school duties. (3)

Q17 In the previous question, which percentage of respondents chose each answer option (not including your own answer)?
_____ I have not changed my daily routine. (1)
_____ Due to government restrictions and regulations, I have spend more time on school duties. (2)
_____ Due to government restrictions and regulations, I have spend less time on school duties. (3)

Q18 **Q6**. During the pandemic, specifically the lockdown, did you have serious confrontations with roommates or family members.

- Yes (1)
- No (2)

Q19 In the previous question, which percentage of respondents chose the answer option "Yes" (not including your own answer)?

	0 10 20 30 40 50 60 70 80 90 100
Yes ()	

Q20 **Q7**. Have you taken advantage of online examination in a way which can be considered cheating due to the lack of control normally experienced in an exam at the University campus.

- Yes (1)
- No (4)

Page Break

Q21 In the previous question, which percentage of respondents chose the answer option "Yes" (not including your own answer)?

	0 10 20 30 40 50 60 70 80 90 100
Yes ()	

End of Block: Study progress

Start of Block: Survey question

Q22 **Q8**.Response rates for course evaluations or school surveys are typically low. Why do you think that is the case ?

- Simple not enough time. (1)
- There is no incentives for the time taken. (3)
- Course evaluations and surveys don't have any personal value (4)

Q23 In the previous question, which percentage of respondents chose each answer option (not including your own answer)?

_____ Simple not enough time. (1)
_____ There is no incentives for the time taken. (2)
_____ Course evaluations and surveys don't have any personal value (3)

Start of Block: Final closure

Q24 Thank you again for filling out this survey. Your response will be carefully evaluated and will have great impact into the insights of the pandemic aftershock.

Q25 **Please save this number** in order to be eligible for receiving the reward.

${e://Field/RandNumber}

Q26 **Please save this number** in case you win the lottery. This number is needed in order to receive the reward.

${e://Field/RandNumber2}

3. Original Experimental Set up

The original setup controlled for a better environment, however, due to the regulations and changes in teaching methods, it became considerably difficult to involve students from the same university and course. The original experiment would have asked students from bachelor 1 till Master. The survey would have been distributed between different study programs. For each study program and year one class would have been chosen. One half of the class would have received the treatment survey, which included prediction question and the other half would have received the same survey however without prediction questions. Both surveys would have given students the possibility to receive a monetary amount as an incentive.
The survey preferrable would be distributed throughout emails by professors and teachers.

4. Honesty score distribution

4.1 Total final score for each question

Student ID	Question 1	Question 2	Question 3	Question 4	Question 5	Question 6	Question 7	Question 8	Total Score out of 70 k	Total Score out of 80 k	ID	ID 2
1	7868.29356	8410.64876	4508.17857	8147.68367	9764.12857	7052.84096	8660.05541	8793.20408	54544.97818	63205.03359	64356	93844
2	6564.17268	9374.58776	5041.79252	8980.66512	8852.79406	8066.3582	9353.36475	8739.96038	55620.33072	64973.69547	47580	96189
3	7994.39246	8511.88185	5685.57143	8102.36224	8844.87925	3283.55102	9340.86475	8824.26291	51246.90116	60587.76591	53073	98421
4	7733.26609	8244.58861	5685.57143	9048.11224	8350.53401	7476.70303	9340.86475	8428.18727	54966.96269	64307.82744	2951	65400
5	6929.83203	9521.25442	5685.57143	8960.04082	9080.59354	4411.18578	9319.43618	9083.64946	53672.12747	62991.56365	61329	89599
6	7180.79356	7409.8217	6981.57143	8832.30148	9550.18571	7882.63406	9292.86475	9418.0108	57255.31875	66548.1835	99969	28105
7	3410.87598	9254.58776	5179.32143	8744.57653	9222.41497	8192.84096	5740.82464	8008.42257	52013.04019	57753.86482	94026	77703
8	7553.05442	8559.65628	3637.42857	8849.4833	8681.10575	6452.5651	6417.81764	9404.69988	53137.99331	59555.81095	19577	34599
9	8101.81005	8904.36305	7493.60714	9202.93785	9193.60544	7918.08234	8826.57904	9461.97719	60276.38306	69102.96209	57857	51213
10	4845.21664	8873.64125	5041.79252	8654.89239	8754.61224	3910.22027	9024.79332	9392.59904	49472.97435	58497.76767	34654	11734
11	7835.60126	9471.25442	6637.03061	9315.84694	9234.98639	5320.42716	8637.04841	9201.83433	57016.98112	65654.02953	32369	90378
12	8065.9584	8511.88185	5685.57143	9458.21058	8609.67718	6014.63406	9112.29332	9056.62425	55402.55774	64514.85106	63713	92372
13	8057.53061	8511.88185	6553.42857	8917.18367	9484.31973	5528.42716	8719.43618	9351.994	56404.76559	65124.20177	36862	22311
14	8316.91994	8244.58861	4624.85714	9453.71058	8432.14471	4411.18578	8140.86475	8914.64106	52398.04782	60538.91257	81230	11364
15	4593.06122	9186.85442	7535.0068	7902.11224	9399.20068	4343.32245	8645.78967	9163.91837	52123.47619	60769.26586	89298	82156
16	8182.71664	8984.85442	7710.57143	9150.43785	9256.46259	5213.53061	8207.07904	9265.96879	57764.54232	65971.62136	90170	20217
17	4707.854	9095.59613	6915.03571	9059.89239	9653.17143	7421.94441	8378.09736	8845.26291	55698.75698	64076.85435	45080	56441
18	4576.02041	9125.23523	5588.17857	8947.52876	8866.04082	4650.08234	7371.10436	9160.42257	50913.50869	58284.61305	11167	83077
19	5157.20408	8170.98776	8034.57143	8427.57421	9035.12857	7845.25475	7568.51694	8561.43097	55232.15177	62800.66872	78960	36784
20	7970.21664	9391.25442	7710.57143	9286.66512	9618.32857	5528.42716	8378.09736	9005.53181	58510.99516	66889.09252	64734	43844
21	4264.06122	9528.18776	5179.32143	9330.84694	8107.46939	6195.59958	9176.57904	9552.06122	52157.54754	61334.12657	34631	32855
22	7770.57823	7929.92696	8072.17857	8560.52876	9582.58571	7721.53061	7139.07904	8504.93517	58142.26402	65281.34306	90255	32665
23	7212.79906	9374.58776	7999.85714	9453.71058	9454.01429	6842.22027	5636.83862	9207.85954	59545.04863	65181.88725	62229	95463
24	6922.96389	9455.65442	8018.17857	9748.53061	9081.49536	5138.75102	8637.04841	9442.7503	57808.32418	66445.37259	79143	25461
25	7522.41444	9418.85442	7824.53061	8417.89239	8924.09276	3780.1513	8581.93618	8509.77551	54397.71145	62979.64763	87625	61597
26	6260.67347	8709.06982	7357.03061	9639.88776	9476.46259	6288.35102	8352.79332	9178.14526	56909.62052	65262.41384	2676	35681
27	6043.84301	8511.88185	6553.42857	8340.57421	8763.03401	8218.08234	6507.2582	9002.14526	55432.98925	61940.24746	42555	76695
28	7299.47488	7158.49839	4538.81633	9209.07421	9063.96259	4254.97959		9470.58223	50995.38822		85035	13394
29	7322.13972	8736.18776	4624.85714	8782.25603	9742.27143	5629.53061		9253.03601	54090.2787		16256	54378
30	6554.41994	9524.58776	7957.42857	7949.14796	9645.32857	7721.53061		8853.33013	58205.77354		33705	57048
31	2911.97488	7409.8217	5874.57143	9132.86224	9716.04286	4073.49388		9265.56543	48384.33241		77993	82633
32	4576.02041	8985.16004	6055.85714	9140.07421	8613.96289	7918.08234		9267.96038	54557.11742		206	60240
33	7247.27708	9481.25442	8092.42857	9103.80148	9554.81429	7845.25475		8390.80912	59715.63972		96800	38260
34	7584.22763	7567.89689	8034.57143	7897.18367	9204.60544	8200.42716		8666.70828	57155.62051		21374	81423
35	6394.66719	6991.25442	4624.85714	9772.03061	8209.1577	7721.53061		7943.34694	51656.84462		28171	72218
36	7826.12323	8579.55102	4624.85714	9290.07421	5925.65121	6014.63406		8693.73349	50954.62437		83233	51575
Average Score	6648.8458	8697.82766	6315.77778	8922.51927	9026.36848	6235.78798	8241.01205	8980.14853	54827.27551	63354.57868		

Table 3: Choice-Matching score distribution

4.2. Mathematical formula of Incentivizing Truth Mechanism

Bayesian Truth Serum: Basic intuition is that someone receives a high score for answers for which the actual frequency is higher than the predicted frequency.

$$BTS\ score = information\ score + prediction\ score$$

$$r = u(x^r = j, y^r) = \log\frac{\bar{x}_j}{\bar{y}_j} + \sum_{k=1}^{m} \bar{x}_k \log\frac{y_k}{\bar{x}_k}$$

$$information\ score\ = log\frac{answer's\ actual\ relative\ frequency}{(geometric)\ mean\ of\ answer's\ predicted\ frequency}$$

$$r = respondent\ , j = answer\ option\ , k = answer$$

Choice-Matching: Steps of computing the final score for one individual

$\boldsymbol{Step\ 1}$: $The\ True\ Distribution$

COUNTIFS = counts cells that match certain criteria's

$= \left(\frac{Number\ of\ similar\ type}{N-1}\right) \times 100$ N = Number of survey Participants

$\boldsymbol{Step\ 2}$: $Calculating\ the\ Prediction\ Score$

$= (10000) - \sum_{i=1}^{X}(Prediction_X - True\ Distribution_X)^2$

X = Types (Multiple choice question)

The second part of the equation is also referred as "error "term.

$\boldsymbol{Step\ 3}$: $Calculating\ the\ Matching\ Score$

AVERAGIFS = Returns the average (arithmetic mean) of all cells that meet multiply criteria

$= (Average\ of\ all\ Prediction\ scores\ with\ similar\ types)$

The intuition behind this step builds upon "matching" the prediction scores of people which have a similar type.

$\boldsymbol{Step\ 4}$: $Final\ Score$

$$= \frac{Prediction\ Score + Matching\ Score}{2}$$

The closer the final score is to 10000 the closer the prediction of the individuals is to the true distribution of the participant was.

5. Testing Assumptions

5.1. Assumption three

Kruskal-Wallis equality-of-populations rank test

Response	Obs	Rank Sum
0	27	582.50
1	3	13.50
2	4	52.50
4	2	17.50

chi-squared = **10.350** with **3** d.f.
probability = **0.0158**

chi-squared with ties = **10.503** with **3** d.f.
probability = **0.0147**

Figure 9: Question 1

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

Response2	obs	rank sum	expected
0	16	199	296
1	20	467	370
combined	36	666	666

unadjusted variance **986.67**
adjustment for ties **-2.29**
adjusted variance **984.38**

Ho: predic~n(Respon~2==0) = predic~n(Respon~2==1)
z = **-3.092**
Prob > |z| = **0.0020**

Figure 10: Question 2

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

Response2	obs	rank sum	expected
0	29	501.5	536.5
1	7	164.5	129.5
combined	36	666	666

unadjusted variance **625.92**
adjustment for ties **-3.06**
adjusted variance **622.86**

Ho: predic~n(Respon~2==0) = predic~n(Respon~2==1)
z = **-1.402**
Prob > |z| = **0.1608**

Figure 11: Question 3

Kruskal-Wallis equality-of-populations rank test

catrep	Obs	Rank Sum
Adapting to online classes.	5	75.50
Constrained privacy caused by roommates, parents or by my partner.	5	132.50
Not much has changed for me.	3	37.50
The lack of motivation caused by daily monotonicity.	23	420.50

chi-squared = **4.386** with **3** d.f.
probability = **0.2227**

chi-squared with ties = **4.398** with **3** d.f.
probability = **0.2215**

Figure 12: Question 4

Kruskal-Wallis equality-of-populations rank test

catrep	Obs	Rank Sum
Due to government restrictions and regulations, I have spend less time on school duties.	12	149.50
Due to government restrictions and regulations, I have spend more time on school duties.	13	327.00
I have not changed my daily routine.	11	189.50

chi-squared = **9.292** with **2** d.f.
probability = **0.0096**

chi-squared with ties = **9.322** with **2** d.f.
probability = **0.0095**

Figure 13:Question 5

```
Two-sample Wilcoxon rank-sum (Mann-Whitney) test

   Response2 |       obs    rank sum    expected
-------------+---------------------------------
           0 |        30         506         555
           1 |         6         160         111
-------------+---------------------------------
    combined |        36         666         666

unadjusted variance      555.00
adjustment for ties       -0.79
                         ------
adjusted variance        554.21

Ho: predic~n(Respon~2==0) = predic~n(Respon~2==1)
             z =  -2.081
    Prob > |z| =   0.0374
```

Figure 14: Question 6

```
Two-sample Wilcoxon rank-sum (Mann-Whitney) test

   Response2 |       obs    rank sum    expected
-------------+---------------------------------
           0 |        15       175.5         210
           1 |        12       202.5         168
-------------+---------------------------------
    combined |        27         378         378

unadjusted variance      420.00
adjustment for ties       -0.90
                         ------
adjusted variance        419.10

Ho: predic~n(Respon~2==0) = predic~n(Respon~2==1)
             z =  -1.685
    Prob > |z| =   0.0919
```

Figure 15: Question 7

```
Two-sample Wilcoxon rank-sum (Mann-Whitney) test

      catrep |       obs    rank sum    expected
-------------+---------------------------------
Course evalu |        18         171         333
There is no  |        18         495         333
-------------+---------------------------------
    combined |        36         666         666

unadjusted variance      999.00
adjustment for ties       -1.03
                         ------
adjusted variance        997.97

Ho: predic~n(catrep==Course evaluations and surveys don't have any personal value) = predic~n(catrep==There is no i
> ncentives for the time taken.)
             z =  -5.128
    Prob > |z| =   0.0000
```

Figure 16: Question 8

5.2. Assumption four

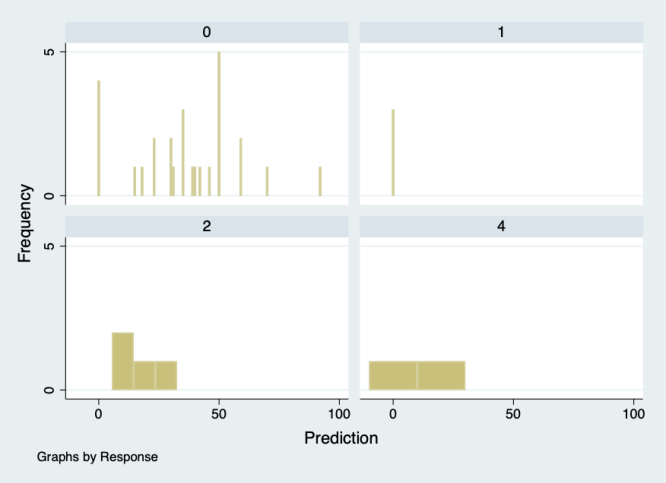

Figure 17: Question 1

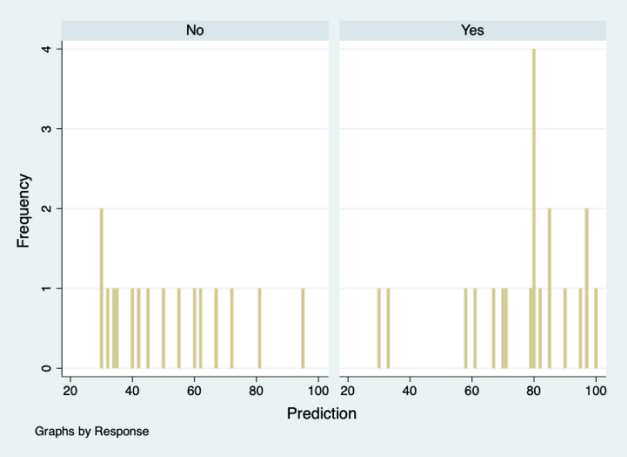

Figure 18: Question 2

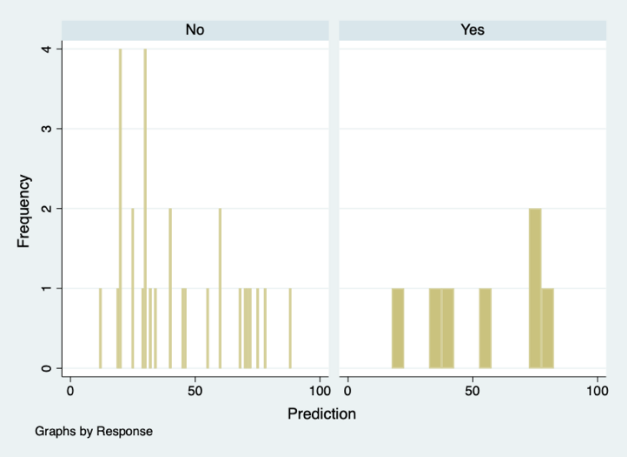

Figure 19: Question 3

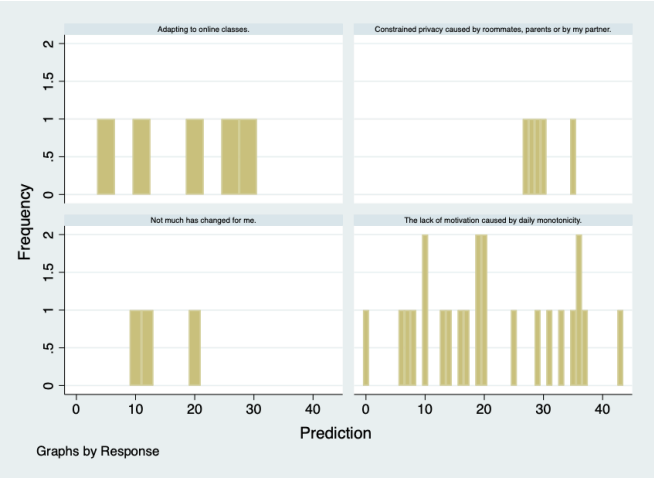

Figure 20: Question 4

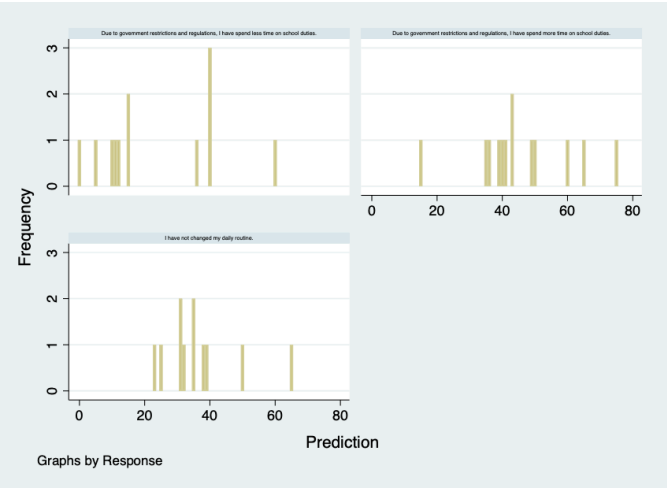

Figure 21: Question 5

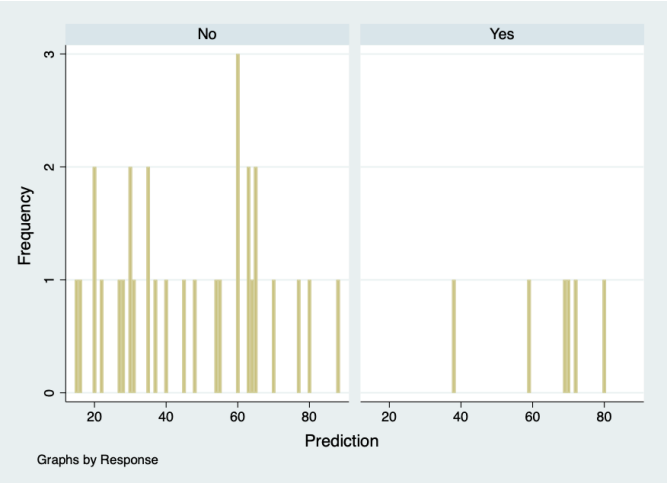

Figure 22: Question 6

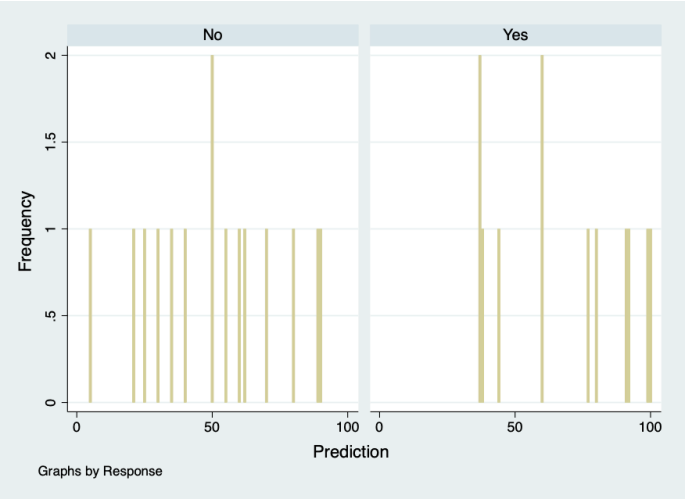

Figure 23: Question 7

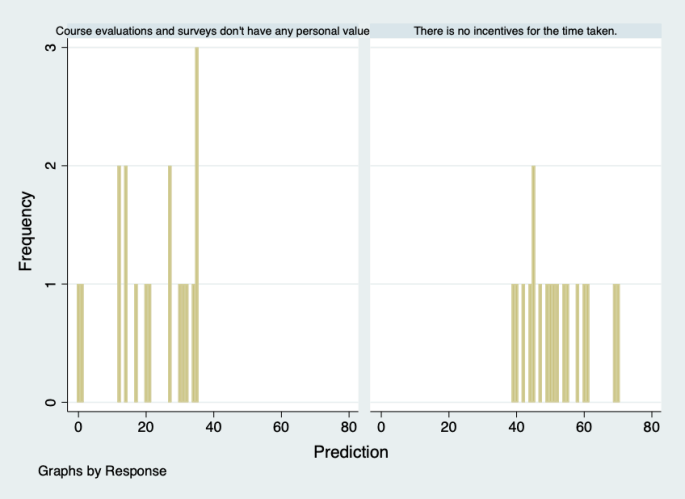

Figure 24: Question 8