“Come if You Are Good…”: An Empirical Evidence on Personalised Mails in the Hungarian Extracurricular Sector

by

Péter Lánczi

Submitted to

Central European University

Department of Economics and Business

In partial fulfilment of the requirements for degree of Master of Arts

Supervisor: Professor László Mátyás

Budapest, Hungary

2017

ABSTRACT

The aim of this thesis is to contribute with some statistical evidence to the growing literature showing that personalised e-mails are more effectively galvanizing actions than pure informative ones. My hypothesis has been that receiving an e-mail with the name of the recipient in the heading creates an “inappropriately” large reaction. With implementing a field experiment in a secondary school programme of a Hungarian extracurricular institution, I investigate whether students are more likely visiting events if personalised mails are received. I find that in the first round of the experiment, the participation rate is 9 percentage points higher among those who received an invitation letter with their name on the headings. I also examine an alternative mailing technique, in which the results are not different significantly from the standard personalised letters or the control group either. Due to the randomised controlled trial methodology, the internal validity of my research is quite high. The research work in my thesis could be extended by better measurement i.e., by assessing whether students actually read those mails, and not just receive it. Despite the unique population, I think this work may contribute to the broader policy implementation of personalised mails.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

In recent times, commonly referred to as the "information age", analysing what raises the attention of people has become more and more researched in social sciences. Convincing consumers to buy products,[1] voters to answer their duty as citizens[2] or pay their debt on time[3] might be more challenging nowadays due to the vast amount of available opportunities. Incentivising people to make better decisions is not about tricking them into situations where they never would like to be. Its intention is to help them avoiding the biases of their own behaviour.[4] Most of these biases are intertemporal ones: humans fail to follow the commitments they make to themselves, like saving more money today or writing a thesis systematically, day by day. Incentives, like reminders, or well-tailored text messages can have the same effect on helping people to keep their commitments as financial ones.[5]

Monitoring the attitude of people towards letters, messages and e-mails from an economical view is not unexplored in the literature (Duflo and Saez, 2003; Bertrand et al., 2005). The purpose of this study is to provide further empirical evidence on the growing literature of shaping behaviour with different mailing techniques. My aim is to measure the

[1] Simester, Hu, Brynjolfsson and Anderson (2009)
[2] Dale and Strauss (2009)
[3] Morten, Karlan and Zinman (2012)
[4] Karlan, McConnell, Mullainathan and Zinman (2010) agrgues that helping to make better intertemporal decisions increases the wellfair of society (p. 3).
[5] Cadena and Schoar (2011)

differences in the attendance rate of a secondary school program (a program of a Hungarian extracurricular education institution), where events are held monthly.

My research methodology is field experiment. At the beginning of the procedure, students were divided into groups according to the length of their e-mail addresses. One control and two treatment groups were created with this method. The first treatment group received a simple, “name added letter” (the heading contained the full name of the recipient) while the second received a kind, reminding letter about their obligation to participate in at least two events per semester. During the analysis I used probit models, and made predictions about the marginal effects of the different variables on the participation rate. The randomisation was successful, and the internal validity of the research was high.

I find that students receiving letters with their names in the heading participate more: they participate 9 percentage points higher than others, and the impact is statistically significant. Considering that the average participation rate is 9%, it is a 50% impact on probability of participation. This result is in line with the findings of the literature, where adding name to letters is a useful strategy to galvanise actions.

This thesis also measures whether reminding someone about their goal in a kind manner changes behaviour. Haynes et al. (2013) found that reminding a person about her obligation or goal, increases the effect of the letter significantly, but not as much, if the recipient was addressed by name. On the other hand Sanders and Kirkman (2014) found that letters with kind remainders – which may increases the level of reciprocity – help to increases participation more than letters mentioning their names. My estimates in this “reminder treatment group” are tend to be lower than in the case where recipients were addressed by name, and this group iw not statistically different from either the control or the other treatment group.

A grand question about the non-monetary, behavioural incentives (sometimes called nudges) is whether the effects are measurable on longer timespan. I find that effects of the personalised letters are uneven: it has a greater impact in the first round of the experiment and it declines rapidly over time. Although the significant effect of personalised letters wears off after the first event, the impact is still measureable on the total sum of participations: the average participation at the end of semester is 0.27, while students who received the name added letter participates 0.1 more, which is an approximately 40% increase.

The following chapter reviews the background literature of the thesis from the political mobilization literature to the collection of delinquent fines with the help of personalised letters. In Chapter 3, I briefly summarise the field of experiment and the experimental design, and I also introduces the methodological setup. Chapter 4 presents the results. First I describe the monthly results of a very elementary regressions, then I build a model to measure the short term effect of personalised mails. Finally, I test how long the power of personalisation can last. I end my thesis with a summary of results, and I also formulate some further possible opportunities to research, based on my dataset.

CHAPTER 2

LITERATURE REVIEW

How can personalised e-mails persuade people to act differently? The idea of my thesis is rooted in the political science literature. Political mobilization, where the target is to inform the voters and galvanise an action (like voting), provides vast amount of empirical results. In their paper Dale and Straus (2009) summarise this wealth of literature, and deduce two leading strategies of the political entities to rise the attention of the voters. The first one, labelled as *Noticeable Reminder strategy*, focuses on the amount of political messages. The more the voter receives the core message, the higher the rate of voting turnout. On the contrary, *Social Occasion strategy* claims that a well-tailored, highly personalised message is more useful for mobilisation. I find this dichotomist approach enlightening about the core problem of attracting attention via different ways of communication.

While the Noticeable Reminder strategy is more intuitive to understand (based on the ancient idea of "repetition is the mother of learning"), the effectiveness of personalisation may need some further investigation. The idea originated in the seminal study of Cherry (1953) about the later labelled "cocktail party effect". In his study, Cherry shows that it is easier to notice things, which are previously known, in a multi audio stimuli environment, than to notice completely unknown ones. Moray (1959) describes that this effect is even stronger when the stimuli is somehow meaningful for the person, like his own name. [6] According to Bargh (1982) it is less energy consuming to pay attention to self-related subject, and much harder to avoid them and focus the attention to other things. Further researches on the topic describe the

[6] The popular version of this theory is rooted from this observation. People can recognise their name spoken out loud even in a crowded, noisy room, where a cocktail party is happening.

existence of cocktail party phenomenon in written texts (Caldwell and Sorenson, 1997, Mack and Rock, 1998).

People do not just notice expressions, which are more related to them, but they are willing to attend any content with greater care and shape their actions to it. Kreuter et al. (1999) show that taking personalised, self-relevant information creates a connection between the text and the recipient, while Dijkstra (2005) shows that even a very elementary level of self-relevance, like using the name of the recipient in a text, can increase the level of understanding, and shape the behaviour later on.

Measuring the effects of personalisation have become common in economics during the recent years. The energy consumption schema introduced by Bargh (1982) coincides with the logic of rationality theory of economics, and empirical researches in this field are reasonably wide spread. Bertrand et al. (2005) show that mails, which are tailored (according to characteristic like gender) can significantly change the investment behaviour of recipients. Even further, personalised mails help to save energy (Gleerup et al., 2010), collecting delinquent fines (Haynes et al., 2013), save more money for their pensioner ages (Duflo and Saez, 2003), quit smoking (Free et al., 2011) or pursuit people to attend job fairs (Sanders and Kirkman, 2014). There is even evidence that pure reminding messages can be as efficient as monetary incentives (Cadena and Schoar, 2011). Of course there are some counter examples as well: there is no conclusive evidence found by Gerben and Green (2000) in political mobilization or Morten, Karlan and Zinman (2012) in collecting delinquent liabilities.

Simple personalisation and persuasion through tailored communication are the two sides of the same scale. The range varies from adding the name of the recipient to a general text (Dijkstra, 2005; Sanders and Kirkman, 2014), to the tailoring of the whole text to the characteristic of the recipient (Betrand et al., 2005). In this study I examine what is the effect of adding the person's full name in the heading of the invitation, so it is useful to introduce some further researches where the procedures is similar.

In his paper Disjkstra (2005) finds that personalised information texts increased the number of attempts to give up smoking. He measured a 20% plus increase in the quitting activities of the volunteers who received preventive information containing their own name, compared to a standard general preventive test. He used three different techniques: *personalisation*, when the name of the volunteer is added to the general text; *adaption*, when self-related text elements are emphasised (like female participants received information how harmful smoking is for babies etc.); and *feedback*, where results of the previously made tests are used (like you smoke 50% more than others). The most powerful strategies are personalization and feedback. The weakest point of his research is that this results are based on self-reporting, so it is not clear, either they actually tried to quit more than other groups or they were persuaded by the treatment to report more than others.

Heynes et al. (2013) discover that people receiving a personalised letter pay 41% more delinquent fines than those who relieves standard mails. In their paper, they use four different type of mails: (1) standard; (2) standard plus: containing the precise amount of the fine; (3) name added to the standard mail; and finally (4) both name and the amount of the fine. The most successful by far is the personalisation approach without the amount of the fine. They also ask the question whether they would have the same results if they resend the mails to those who don't pay on time. Although they suggest that the effect of recent text messages would be smaller in these cases, they have not implemented it in their paper.

Sanders and Kirkman (2014) use different levels of personalised mails to invite unemployed individuals to job fairs. They introduce the reciprocity exchange idea of Falk and Fischbacher (2006), and assume that writing a kinder and more personal letter helps to increase participation. The first treatment group contains just the name of participants, while the second contained the name of the manager, who sent the text message, and finally the third was finished with a kind and encouraging sentence. They find that attendance rate among the third group was the highest: 26.8%, which is approximately 16% more than the control group. On the other hand, the "just the name" strategy exceeded the control with only 4%, and it is not significant at all.

From the literature I deduce three relevant research questions. First of all, can personalisation with names galvanise the expected action? Disjkstra (2005) and Heynes et al. (2013) argue it can, Sanders and Kirkman (2014) measures its existence just in one setup, while Morten et al. (2012) find no evidence of it when they send reminders for loan repayment.

Secondly, all of the reviewed articles use numerous variations of messages which expand the possibilities of personalisation. I divide these strategies into the following groups: "social pressure" (Gerber and Green, 2010; or the feedback technique of Disjkstra, 2005), where the expected action is motivated by the performance of others; "goal setting" as in the case of Haynes et al. (2013), where they show the amount of delinquent fines, or Karlan et al. (2010) where they introduce the achievable goal to increase the willingness to save up, or "reciprocity exchange" like in Sanders and Kirkman (2014). In my thesis I use a combination of the latter two: the goal setting and kindness exchange approaches.

Finally, Cahill and Perera (2008) and Oliver (2012) remark that small, non-monetary attempts to influence behavioural incentives, the so called "nudging", usually achieves a "single shot" change, and after the withdrawal of the incentives the effect may erode, or even make things worth than before. Heynes et al. (2013) also point out that the effect of the personalised mails may wear off in the longer period. In the marketing literature, the phenomenon that short run success may ruin the long run revenue is an extremely critical issue. Simester, Hu, Brynjolfsson and Anderson (2009) examine the short and long run effects of a personalised advertisement campaign. They found that these campaigns – although they are successful in short run – haven't increased the revenue on the long run. I also intend to measure long-run effect of personalised mails in my research.

To sum up, I have found an avalanche of evidences and some counterexamples from various fields of life which make the personalisation phenomena convincing for research. People tend to listen to a known thing, especially if is related to them either it is audible or it is written down. Furthermore, there is a high probability that people act in accordance with the content of a personalised message.

CHAPTER 3

EXPERIMENTAL SETUP

3.1 FIELD OF EXPERIMENT

Mathias Corvinus Collegium (MCC) is an extracurricular education institution in Hungary which offers trainings to students from the elementary school until the end of their master thesis. MCC consists of approximately 2000 students from all over the Carpathian Basin, among which more than 1500 students are members of the Secondary School Programme (SSP). The education program for secondary school students consists of three elements: an e-learning platform, monthly education gatherings (called "SSP Saturdays") and talent camps. The e-learning platform includes video based, 8 week long courses – similar to Coursera – in different disciplines of social sciences with regular tests. SSP Saturdays are monthly organised events which consist of a morning lecture and seminars during the afternoons. Talent camps are organised every half year; they are four-day long training and meeting sessions.

The program is divided into semesters. A semester is completed when a student finishes one e-learning course (achieved 50% at the tests) and participates in at least two SSP Saturday events. Dates of the SSP Saturdays are scheduled at the beginning of the semester; the students are informed about it, but the detailed content is sent just two weeks before the occasion via e-mail. The usual number of participants on SSP Saturdays is around 120 students.[7]

While it is a non-degree programme, MCC SSP helps students to gain better achievements in school or at matriculation exams; however, it never promises any short term

[7] See in **Table 3**.

gain, like exact points in the university application. This means that the success of the programme is based on the level of the students' engagement.

3.2 EXPERIMENTAL DESIGN

My thesis examines the effects of personalised invitation on the participation level of students at MCC SSP Saturday events. According to the previously reviewed articles, students may increase their attendance rate on SSP Saturdays due to such personalisation. The research strategy was to send different mails to students, who are randomly assigned into the following different categories.

- The CONTROL group received an invitation letter, containing all the details of the event, but with the heading: "Dear Student". (Before the timespan of the experiment, all letters were addressed like that.)
- The TREATMENT 1 group received the very same letter like CONTROL except the heading with the full name of the student.
- The TREATMENT 2 group received the same letter like TREATMENT 1, but with additional reminder sentences, how important it is for the students to participate in at least two events per semester to face with the requirements, and some personal ending words.

Multiple treatment is commonly used in the literature: researchers can compare the results of different mailing strategies if the different treatments are under the same condition.[8] **Table 1** summarize about the received treatments and the subgroups.

[8] Gerber, Green, Kaplan and Kern (2010) p. 302 explains why one can actually do this.

Table 1: Summary of treatments for the different subgroups

	Control	Treatment 1	Treatment 2
Name of the student		X	X
Reminder & farewell			X

During the spring semester of 2017, I collaborated with the managers of the program to run this experiment for our mutual benefit. In January we agreed upon terms, the dataset was anonymised and the date of events was settled. **Table 2** shows the schedule of the experiment, which was separated into four rounds. Reminder letters were sent a fortnight before the events. In Round 0, everybody received the control letter. In Round 1 randomisation started, and students received letters according to their group of randomisation. In Rounds 2 and 3 the treatments of Round 1 were repeated. The translations of the generic messages are part of the Appendix[9].

Table 2: Schedule of experiment (2017 February – May)

Round 0: every participant in the control	
10 February	invitation to the 1st event of the semester
25 February	1st event of the semester
Round 1: randomisation	
10 March	invitation to the 2nd event of the semester
25 March	2nd event of the semester
Round 2: Round 1 repeated	
14 April	invitation to the 3rd event of the semester
22 April	3rd event of the semester
Round 3: Round 1 repeated	
8 May	invitation to the 4th event of the semester
20 May	4th event of the semester

[9] See Appendix Table 5.

3.3 DATA

The dataset consists of 1190 students who were assigned to Budapest program of SSP from February to May 2017. I received the anonymised data including the following variables: e-mail address, gender, ZIP code of accommodation, ZIP code of school, number of semesters spent in SSP, secondary school grade and coded name of the school. **Appendix A** summarize the most important variables.

Table 3 introduces the participation rate of the previous year events of the SSP programme. The average participation rate was 13% in 2016, and during the spring it was slightly higher, 15%. After interviewing the managers of the programme, it emerged that there is a higher participation rate during the spring semester. The peak of the participation rate is at the first event of the semester; later on it decreases and during the end of the semester it climbs back.

Table 3: Participation data from 2016

	Eligible	Participation	Participation rate
Spring Semester			
2016 February	996	242	24%
2016 March	996	173	17%
2016 April	996	97	10%
2016 May	996	137	14%
2016 June	996	113	11%
Average			*15%*
Autumn Semester			
2016 September	877	156	18%
2016 October	877	123	14%
2016 November	877	82	9%
2016 December	877	50	6%
2017 January	877	89	10%
Average			*11%*
Year average			*13%*

During the spring semester 2017 students could participate in four training events. As I described previously, students learned the date of trainings before the beginning of the semester, and they received an additional invitation letter two weeks before the actual events. **Table 4** shows the results of the participation. Comparing it with **Table 3**, the number and the percentage of participation is lower than one year before, but higher than in the previous semester. The level of participation follows the usual pattern – starting with a high participation rate which then declines but climbs back in the end. [10]

Although the number of participants at SSP Saturday events is limited, during the research period it did not matter. The capacity of the building is up to 250 people, and **Table 4** shows that the number of participation remained below this. This means that no participants dropped due to this technical reasons.

Table 4: Participation data from 2017 Spring Semester

	INVITED STUDENTS #	# OF PARTICIPANTS	PARTICIPATION RATE	EXPERIMENT
FEBRUARY	1190	173	14,5%	NO
MARCH	1190	134	11,3%	YES
APRIL	1190	91	7,6%	YES
MAY	1190	100	8,4%	YES

Most features which are required for a well-behaving experimental design are reasonably well satisfied in my field experiment. One of the frequent problem is the quality of data collection and treatment. I received the best, most up to date data from the managers, and they regularly refreshed the database at the beginning of the semester. The results are also complete: since all outcome data is available, there is no attrition.

[10] It is worth mentioning some other side effects of the experiment. While the exact number of participation was lower month by month, when letters were sent according randomisation, the operators of the program told me that they received an enormous amount of excuse responses compared to previous years.

3.4 RESEARCH METHODOLOGY

The intention of most economic analysis is to find causal effects instead of statistical associations among variables (Angrist, Imbens, Rubin, 1996). The common strategy to measure the impact of a treatment is to compare the group average of treated and untreated units, [11] which is called ATE, or average treatment effect (Duflo et al., 2007). ATE measures a causal effect, if treated and the untreated outcomes behaves like the potential outcomes (Rubin 1974) of each other. The idea is to compare the outcome related to an individual if he had been treated ($Y_i(1)$) and the very same individual if it hadn't been treated ($Y_i(0)$). Since it is unrealistic to measure both at the same time,[12] some assumption is needed to find a realistic, counterfactual counterpart. Angrist et al. (1996) show[13] that these assumptions are satisfied if the treatment status is decided upon a randomised experiment.

Gerber and Green (2012) point out two assumptions that are crucial to satisfy (after the randomization): excludability and non-interference. Excludability means that the difference between the treated and non-treated is the fact of the treatment, so the measured difference is due only and *excludible* to the treatment. Non-interference means that participants in the experiment may not adjust their behaviour due to the fact that they are in the treated or in the non-treated group. If one can assure these two conceptions, the randomised field experiment can be unbiased.

My randomisation design is on the individual participant level. I took all SSP students who live in Hungary, and created three groups of them. I measured the length of their e-mail

[11] This is also called as difference-in-means estimation strategy. (Gereben and Green, 2012)
[12] Holland (1986) calls this as the „fundamental problem of causal interference".
[13] Precisely in Section 3.1. The assumption set is called SUTVA (standard unit treatment value assumption) and it requires the ignitability of the treatment.

addresses and divided it by three. If the remainder was zero, the student went to the Control group. In line with that, remainder 1 created the Treatment 1 group and remainder 2 the Treatment 2. Participants were "blind", none of them knew that they were part of a research. As I have shown in the paragraph above, it is crucial to avoid the risk of interference among participants, also known as Hawthorne or John Henry effect,[14] because then the treatment or the control group adjusts their behaviour.

Table 5 shows the results of the randomisation. Approximately one third of the population took part in each and every group of the experiment. **Appendix Table 1** shows the distribution of different variables between the control and treatment groups compared to the total population. However, there are minor differences between the total population and the subgroups, they are not decisive. As **Table 2** shows, in February none of the eligible students received any treatment, while in March, April and May students received a group number from the above described randomisation procedure.

Table 5: Result of randomisation

	CONTROL	TREATMENT 1	TREATMENT 2	TOTAL
STUDENT #	386	392	412	1190
	32.4%	32.9%	34.6%	100.0%

Despite the effective randomisation procedure, I am not able to exclude the possibility of communication between the control and treatment groups, which may ruin the non-interference assumption. The treatment group may influence the behaviour of the control group (or vice versa) via personal relations (i.e. friends calls each other's attention to participating) or location can cause an undetected bias (i.e. participants attending the same school may organise their trips to the event together). The school spillover is easier to measure since school variable

[14] Duflo, et al. (2007), section 8.2.

is available in the dataset. On the other hand, there is no guarantee that people coming from the same school know each other at all. The case of social spillover is harder to detect, because I found no conclusive strategy to identify friend relations. If my hypothesis is right and a personalised letter causes a positive change in participation, then under- or unidentified social or geographical spillovers cause a bias. The sign of this bias is hard to determinate, since a group of students can convince each other to come and not to come with equal probability. This is the weak point of my research.

Table 6 shows that there were undelivered e-mails: 1-4% of all sent mails bounced back. Although there is an increase in the amount of bounced back mails, I found no student who did not receive the e-mail more than once, so the eligible number of students remained the same during the timespan of the experiment. Unreceived mails endanger the proper measurement of messaging (Haynes et al., 2013), and it can hurt the assumptions ensuring that randomisation brings casual results. Since there are some individuals who should have received the treatment, but they actually never received it then the result is the intent to treat effect (ITT) instead of ATE.

Table 6: Summary of undelivered e-mails

	FEBRUARY	MARCH	APRIL	MAY
# ELIGABLE STUDENST	1190	1190	1190	1190
# PARTICIPANTS	173	134	91	100
# BOUNCED BACK MAILS	15	16	43	42
% BOUNCED BACK MAILS	1.3%	1.3%	3.6%	3.5%
# OF PARTICIPANTS WITH BOUNCED BACK MAILS	0	0	0	0

To measure the short run effect of the treatment I compared the outcomes of the different treatment groups, but ensured that they were the best possible potential contrafactual outcomes of each other. This is not so easy as it sounds. Angrist et al. (1996) defines four latent groups which can exist among participants: always-takers, never-takers, defiers and compliers. I denoted with z=1 if a student was assigned to a treatment group, z=0 if he was not and define d=1 if the student received the letter, while d=0 if he did not. An always-taker is somebody, who always acts as he received a personalised letter, no matter whether he received it or not (d=1 either if z=1 or z=0). The opposite of always-takers are the never takers, because they never receive the treatment (d=0 either if z=0 or z=1). The most interesting group is the so called compliers, because they are the ones who act in line with the treatment: if z=1 then d=1 and if z=0 then d=0. The opposite of compliers are defiers: they act always the opposite of the treatment, so if z=1 than d=0 or z=0 than d=1.

In the case of my research never-takers are easy to define. They are the ones who did not received their e-mails due to technical reasons.[15] They are measured in **Table 6**: 1-4% of the population are never-takers. Since a letter was sent before the beginning of the semester containing the precise date of events in the examined period, there is theoretical chance that some students may have appeared despite no invitation, so I can not exclude the existence of always-takers a priori. However, according to the last row of **Table 6** none of the students participated in the events who did not receive the invitation letter before. Thus there is no evidence of the existence of always-takers. Due to the fact of the single blind research design setup, nobody knows actually about participating in an experiment, so it is reasonable to rule out defiers a priori.

Why it is so important to identify these latent groups? Gerber and Green (2012)[16] show that the possible diference in distribution of the latent groups among the treated and non-treated cause a barrier in identifing the causal effect which is more accurate than intend-to-treat effect. The final aim is to idenfiy complier average causal effect (CACE), because compliers always

[15] Haynes et al. (2013) and Sanders and Kirkman (2014) use the same strategy to identify never takers.
[16] Precisely in Chapter 3 and 4.

act in accordance with the treatment, so the unbiased, consistent effect of treatment could be measured the best among them. In my experiment, as I have shown, always-takers or defiers do not exist and never-takers are measurable, so I conclude that the remaining 96-99% of the students are the compliers.

Another approach to identify the four latent groups is also possible. Previously, *z* was defined as being assigned while *d* was about receiving the mail. Let us take the idea that *d* represents whether they actually opened these letters, not just received it. Of course, it would increase the accuracy of the measurement, shrink the number of compliers and increase the number of never-takers. However, tracking of mailing habits in the current e-mail system of the programme is unavailable; this is a reasonable field for further research. It is worth pointing that according to the managers of the program, "newcomers" (with zero or one finished semesters) are usually more active, so there is a possibility that taking the semester variable into the regressions may reduce this kind of variability.

Since my research is measuring the effect of personalised letters over time, it is an important question whether one can consider the different rounds of the experiment as separated periods of the treatment; in other words, is it right to consider that I measure the same kind of effect in March and May? Gereben and Green (2012) in section 8.4 specify two criteria to adjust the "classical" non-interference and exclusivity assumptions to interrupted time series. The first one is the *no-anticipation assumption*, which is the alternative of the non-interference assumption. In a nutshell, imagine any given *i* individual as two different individuals in *t* and in *t+1* period of the experiment, and apply the non-interruption assumption to them: it means the communication between *t* and *t+1* may ruin the measurement of the causal effect. This issue typically arises in medical experiments, where someone receives the treatment later than others – due to shortage of medicine – but he anticipates the fact of treatment which may change his behaviour compared to his behaviour if he received the treatment at the beginning of the experiment. In my research setup this assumption is satisfied, because there was no delay in treatment, treatment groups were steady during the experiment (i.e. none of the participant is moved from Control to any treatment ad vica versa) and the students never knew that they were part of an experiment.

The second assumption is the *no-persistence outcome*. It means that treatment in one period is unaffected by the treatment in another period. If it is true, then participants of the experiment in each period can be treated as different individuals, and the researcher can enjoy the advantages of a large, pooled dataset. In my research the satisfaction of this assumption is undecided a priori. As I show later on, I find conclusive evidence that the first round of treatment had a long lasting effect on later rounds of treatment. To ensure non-persistency, I treat the whole three rounds of treatment as one treatment with three doses. According to this, I define treatment as receiving letters in all three periods of the experiment.

Continuing the above mentioned line of thought, comparing individuals with different levels of treatment can ruin the accuracy of my estimations.[17] Partial treatment is a problem because treatment in Round 2 may be different for those who receive it as a first doze or a second one. Individuals with at least one bounced back e-mail can't receive the full scale of treatment, so they are not comparable with others.

3.5 HYPOTHESIS

The base hypothesis of my thesis is that when students receive an invitation letter with their name in the heading they are more likely to take part in the event in the short run. Using the nomination from the previous section, Hypothesis 1 is true if participation is significantly greater in TREATMENT 1 than in CONTROL in March.

[17] Gerber and Green (2012)

The empirical results of earlier researches in the field of personalised letters show that there is a difference between the outcomes of the message variants. I expect that students in Treatment 2, with the additional kind reminder will participate more than the students from Treatment 1. The idea is that these sentences are deepening the reason why it is good if they would come and enhance the expected action, namely the participation. I assume that Treatment 2 is significantly different from Treatment 1, and it is also larger than that.

My third research question is the duration of the above described effects. Haynes, et al. (2013) point out that receivers may get used to the treatment, so the effect just wear off after a while. One possible reason behind this phenomenon is that students optimise on the whole term. If students already decided their preferred level of participation before the semester started, then this experiment may affect only the dates, but not the total amount of participation during the semester. Hypothesis 3 is that the effect wears off on the long run, so the total amount of participation is not significantly different among the various groups.

CHAPTER 4

RESULTS

4.1 SHORT TERM RESULTS

4.1.1 MONTHLY RESULTS

Since the monthly results of my experiment are binary variables, I use a probit model estimator on the dataset. While linear probability models also can be used, they are not the best linear estimators. The basic problem of linear portability model is the heteroscedastic residual, which is present due to the limited values of the dependent variable. With robust residuals the heteroscedasticity can be avoided, but the range of estimation is still out of the zero to one interval, so latent variable estimation methods like logit or probit are much commonly used in economics (Wooldridge, 2008). The advantage of the probit model is the better estimated significance, while the drawback is that the coefficient of the point estimate is not the marginal result of treatment, but a probability number. Despite the fact that the magnitude in the result of a probit model is not comparable at once, the sign and the significance are conclusive.

The first estimated model is the following:

$$part_{mi} = P\left(c + \beta_j \; treatment_{ji}\right) + u_i \qquad (1)$$

where $part_{mi}$ is the dependent variable with binary outcome of the participation, and it takes 1 if student *i* participated at the event of *m* month. This variable depends on the treatment dummy variables, where *j* can take the value 1 or 2 in accordance with Treatment 1 and Treatment 2. If student *i* received any of the treatments, then the variable takes 1, otherwise 0.

Table 7 shows monthly results of the personalised letters. *Column 1* shows that in February the Treatment 1 and 2 group are not significantly different from the Control group, but both of them are positive and the point estimates are close to each other. This pre-experiment result is somewhat puzzling, but there is no evidence that these similarities are due to some omitted explanatory variable which may bias the success of randomisation. Because the coefficients of the treatments are not significant I conclude that this regression serves as the exogeneity check of my randomisation procedure. Since the treatment groups have no a priori explanatory power, the randomisation is successful.

Table 7: Estimation results of Model 1

VARIABLES	(1) February	(2) March	(3) April	(4) May
Treatment 1	0.141	0.272**	-0.0466	0.173
	(0.112)	(0.122)	(0.137)	(0.129)
Treatment 2	0.140	0.218*	0.115	0.0527
	(0.111)	(0.122)	(0.129)	(0.131)
Constant	-1.154***	-1.386***	-1.457***	-1.457***
	(0.0819)	(0.0919)	(0.0957)	(0.0957)
Observations	1,190	1,190	1,190	1,190
Pseudo R-squared	0.00211	0.00659	0.00260	0.00281
χ^2 Test	0.991	0.631	0.216	0.336

Standard errors in parentheses
*** p<0.01, ** p<0.05[18]

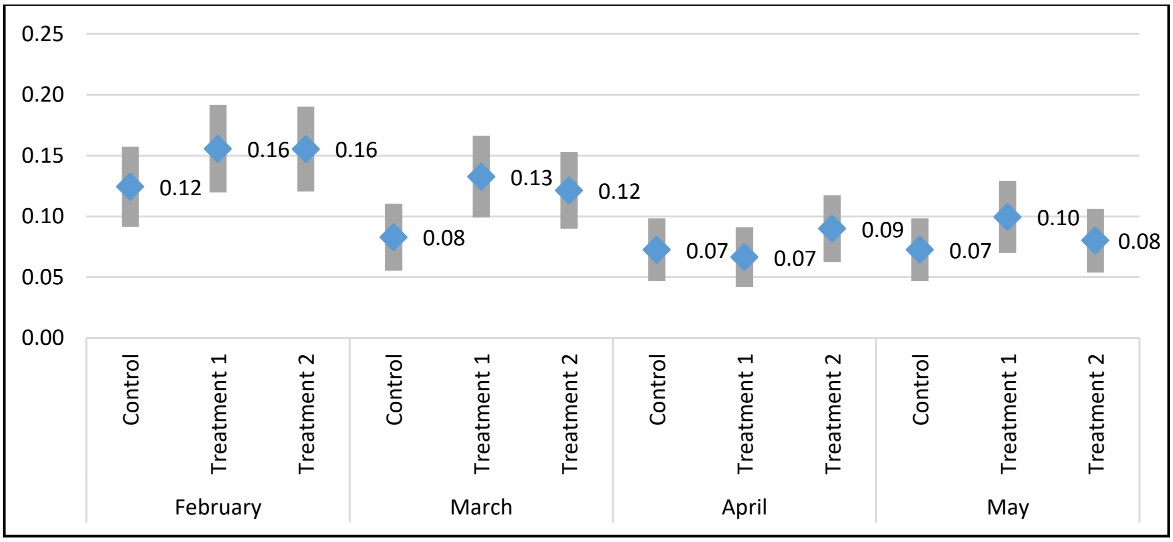
Column 2 shows that in March the participation in treatment groups are significantly larger than in the control. Treatment 1 seems to be more significant and larger in point estimate than Treatment 2. This result suggests that Hypothesis 1 is satisfied, and in March the treatment has a significantly positive effect on participation.

[18] One-sided probability of the treatments can be found in the Appendix Table 2. Since the one sided significance tests are in the same line of interpretation like the two-sided ones, I just introduce them in the Appendix.

Column 3 contains results from the April round. According to this, none of the treatments have any significant effect on participation, while the point estimate in the case of Treatment 1 turns negative. *Column 4* shows the participation in May. Although the sign of the coefficients are positive, they are not significant. These results strengthen the existence of the intertemporal substitution, so support Hypothesis 3.

The last row of **Table 7** contains χ^2-tests which measure if the point estimate of Treatment 1 and Treatment 2 are significantly different from each other. The estimates are the closest to significance in April, but it is still insignificant in any reasonable level. According to Model 1, there is no difference between the two treatment groups, which is against Hypothesis 2.

Figure 1: Marginal effect estimations in Model by months



Grey lines are representing the 95% confidence intervals around the blue squares, which are the point estimates

While the results of **Table 7** are giving a summary about the signs and significances of the treatment variables, **Figure 1** shows the marginal results at the means of Model 1.[19] In February the participation rate among students in the Control group is around 12%, while in

[19] See also Appendix Table 3.

both treatment groups it is around 16%. The grey line shows the 95% confidence interval around the point estimate. In February the point estimates of the treatment groups are in the grey area of Control, so it also shoes that the randomisation procedure is successful. Treatment groups are different from control groups only in March. Here the participation rate in the Control is around 8 percentage points, while it is 4-5 percentage points more in the treatment groups. In later months, the point estimate of Treatment 1 and Treatment 2 are always in the 95% confidence interval of Control. It is worth mentioning that in April the point estimate of Treatment 1 is actually lower than the control, which is in-line with Hypothesis 3. Apart from the fact, that Treatment 1 and 2 are never different from each other significantly, one can note that the point estimates are not necessarily moving together.

Examining the a priori hypothesis of my thesis through the lens of **Table 7** and **Figure 1**, it can be concluding that personalised letters have a significant effect on participation, at least in March, so Hypothesis 1 can not be rejected and personalised letters can galvanise actions. According to the Hypothesis 2 students would react differently to the shorter and longer version of the letters. From the results of Model 1 this hypothesis is falsified, since χ^2 tests show that point estimates are not significantly different from each other in any months. However, it is worth noting that results of the longer version of the letter are less significant than the shorter letters. Since Hypothesis 1 holds, a discussion whether long lasting effect exists is reasonable. In **Table 7** *Column 3* and *Column 4* all variation of letters lose their explanatory power, which supports Hypothesis 3.

4.1.2 OTHER EXPLANATORY VARIABLES

Table 7 illustrates that the level of pseudo-R^2 (or McFadden R^2) in the treatment dummy setup is very low, which suggests that other explanatory variables may exists. If randomisation is done properly, involvement of additional explanatory variables does not change the point estimate while it can reduce its variance.[20]

From the dataset the possible explanatory variables are the following: *grade,* which is the secondary school grade of a student; *bp,* which is a dummy and takes 1 if a student lives in Budapest, and 0 otherwise; *semester* which shows how long a student has been part of the programme; and *school* which is a categorical variable.

I estimate the following regression:

$$part_{mi} = P(c + \beta_1 BP_i + \beta_2\, girl_i + \beta_3\, semest + \beta_{3+j}\, treatment_{ji} + \beta_{5+k}\, school_{ki}) + u_i \qquad (2)$$

where

$part_{mi}$ is a binary variable representing the participation of student *i* in month *m*;

BP_i is 1 if student *i* lives in Budapest, 0 otherwise;

$girl_i$ is 1 if student *i* is a girl, 0 otherwise;

$treatment_{ji}$ is 1 if student *i* Treatment *j* if $j = \{1; 2\}$; and

$school_{ki}$ is 1 if student *i* attending school *k* if $k = \{1; ...; 254\}$.

Heuristically, there is a great chance that students who live in Budapest are more willing to participate, because they face lower travelling and time costs than others from the countryside, so β_1 is positive, but not necessarily large. The higher the number of *semester*s spent in the programme, the lower the probability to participate in the events. The reasoning goes like this: students are interested in a few e-learning topics and when they finish them in the first or the second semester later they just cherry-pick from the programs. Based on this assumption, β_3 should be negative and large by magnitude. From the experience of the

[20] Duflo, Geinnerster, and Kremer (2007)

managers, 10th and 11th grade students are the most enthusiastic participants, so regarding grades I aspect a reverse U turn. On the other hand, grades and semesters are highly correlated,[21] so one of them should not be included into the regression. I leave the number of semesters in the models because it is more related to the programme, so it is probably better measured. The array of school dummies is responsible to treat the possible geographical spillovers, discussed in Section 3.4.

Table 8: Estimation results of Model 2 in February

VARIABLES	(1) February	(2) February
Budapest	0.127	0.323
	(0.117)	(0.266)
Girl	0.0459	0.105
	(0.108)	(0.150)
Semester	-0.416***	-0.396***
	(0.0424)	(0.0639)
Constant	-0.592***	-1.056**
	(0.107)	(0.492)
Observations	1,190	635
School dummy	NO	YES
Pseudo-R^2	0.127	0.200

Standard errors in parentheses
*** p<0.01, ** p<0.05

Table 8 is a variation of Model (2). It represents the pre-experiment explanatory power of the above described variables. In *Column 1* the dependent variable is the participation rate in February, while the explanatory variables are gender, living in Budapest and number of semesters. In *Column 2* the array of school dummies are added to the regression. From **Table 8** it can be seen that neither Budapest, nor the gender dummy has any significant effect on participation, so I decided to leave them out from the further regressions. The *semester* variable

[21] The exact number is 0.51.

– on the contrary – is very significant in both setups, and has a negative sign. This finding is in line with expectations. Hence pseudo-R^2 is higher in *Column* 2 and school dummies increase the precision of the models. On the other hand, if the number of students from a given school is smaller than two, or there is not a single student from a given school, who participated in February, the probit procedure omits those schools. As a result of this, a substantial number of students are dropped out from the regression. [22]

Table 9 introduces estimations of Model 2 on March. *Column 1* shows the results, if *semester* variable is added to the regression. Compared to **Table 7** *Column 2*, the point estimate of Treatment 1 is almost the same, while Treatment 2 shrinks in both value and significance. The a priori assumption about semester is satisfied: the larger the number of finished semesters the lower the chance of participation. The explanatory power (pseudo-R^2) of this model is stronger than any previous one.

In *Column 2* school dummies are added to the estimation. The result – as previously described – is a drop in the number of observations. Despite the shrinkage in the dataset, introducing school dummies helps to deal with the previously mentioned social and geographical spillovers. Compered to *Column* 1, in *Column 2* point estimate of Treatment 1 increases, while the significance remains high. The absolute value of the *semester* variable also increases.

Complier average casual effect (CACE) in *Column 3* shows the results among those, who actually received the mails. The estimated model is the same as in *Column 2*. Due to the low number of bounced back mails (see **Table 6**), the result is close to the findings of *Column 2*. The marginal effect prediction shows that the participation rate is 9% in the Control group

[22] Precisely 198 schools and 550 students are dropped from the regression.

while it is 18% in Treatment 1 and 13% in Treatment 2. The outcome of Treatment 1 is significant at 5% level, while Treatment 2 is not significant.

Table 9: Estimation results of Model 2 in March

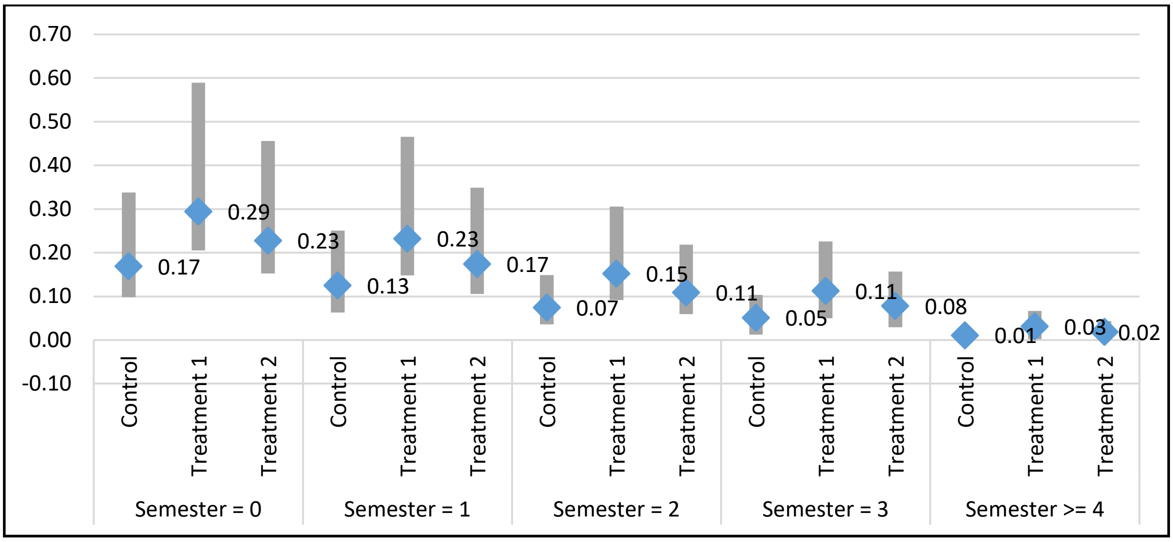
VARIABLES	(1) March	(2) March	(3) March	(4) Wrong
Treatment 1	0.277**	0.417**	0.420**	-0.110
	(0.126)	(0.175)	(0.176)	(0.177)
Treatment 2	0.196	0.215	0.218	-0.0197
	(0.126)	(0.171)	(0.171)	(0.170)
Semester	-0.272***	-0.329***	-0.322***	0.219***
	(0.0407)	(0.0646)	(0.0654)	(0.0387)
Constant	-1.011***	-1.535***	-1.541***	-2.223***
	(0.106)	(0.554)	(0.554)	(0.152)
Observations	1,190	662	640	1,190
Model	ITT	ITT	CACE	Exogeneity
School dummy	NO	YES	YES	NO
Pseudo-R^2	0.0709	0.162	0.158	0.0853
χ^2 test	0.495	0.201	0.203	0.607

Standard errors in parentheses
*** p<0.01, ** p<0.05[23]

As *Column 4* shows, being a non-receiver of a mail has no relation with the treatment. This is crucial to ensure that there is no systematic bias among non-receivers of the mails, because that may ruins the internal validity of my model. The thread is theoretical because of the low number of bounced-back mails, but one can conclude that it increases the precision of the point estimates. It is also important to point out that non-received mails are highly correlated with the number of finished semesters. One possible reason behind this phenomenon is that those who do not follow closely the on-goings of SSP have changed their e-mail addresses.

[23] One-sided probability of the treatments values can be found in the Appendix Table 2.

Figure 2 Marginal effect estimations in Model 2
at different levels of the *semester* variable



Grey lines are representing the 95% confidence intervals around the blue squares, which are the point estimates. All the other variables are at the mean.

Figure 2 introduces the marginal effect estimations of the treatment and control groups while keeping the *semester* variable constant at different levels.[24] One can conclude that the fewer the number of finished semester the higher the effect of the different treatments. It is also worth noting that in every different semester subgroup, the level of participation is higher in the treatment groups than in the control one. In the subgroups from semester=0 to semester=2, the point estimate of Control is out of the 95% confidence interval of Treatment 1. The same is not true about Treatment 2 and Control nor the Treatment 1 and Treatment 2 relations.

Neither of the measured effects of **Table 9** can be treated as average treatment effect (ATE). As shown in Chapter 3, there are students (with bounced back mails) who are actually never-takers of the experiment, so all the results including them can be treated as just intend to treat effect (ITT). All outcomes of the introduced models in **Table 7** and the first 2 columns of **Table 9** are intend to treat effects. Complier average causal effect (CACE) only measured in *Column 3* of **Table 9**, and I treat that as the short run final result of my thesis.

[24] See also Appendix Table 4.

As I wrote previously, there is one more major danger in calling the results of *Column* 3 of **Table 9** a complier causal effect, namely whether the students read the e-mails at all. There is no solid proof that any of the students actually opened and read the letters. Because this is the only source from they can learn what is going to be the agenda of weekend I assume that no other channels can exist. However, there are students who never read the mails (let us call them No-Readers), so they never received the treatment. To handle this issue, I had to make sure that No-Readers are uniformly distributed in the treatment and control groups. Impressions and information from the managers of the programme show that this statement is not necessarily true: No-Readers were not uniformly distributed among the subgroups because their number increases with the number of semesters spent in the programme. This explains the negative sign of semester variable, so statistically it answers the problem of non-uniformity of No-Readers. On the other hand, in section 3.3 I redefined the treatment in such a way that personalisation was responsible not just for galvanising the action, but also for gathering attention to read the letters. Until any of these two assumptions hold, the result of *Column 3* still remains the complier average causal effect.

To sum up, I claim that there is a significant effect of personalised mails on the participation of the students. **Table 9** Column 3 clearly shows, that the shorter letter version (Treatment 1) has a significant impact on participation immediately after implementation. Unlike the short letter, the longer version has very low significant effect on participation, and as χ^2-test results show, there are no significant differences between the two treatment groups. So, for this subsection Hypothesis 1 holds, while Hypothesis 2 is rejected. The next subjection provides a deeper analysis about Hypothesis 3.

4.2 SEMESTER RESULTS

Table 7 and **Figure 1** show a remarkable decline in significance and magnitude of treatment effects after the first month of the experiment. In March, both are significant, and they are larger with 5% and 4% more than the control, while in the following months, effects shrink rapidly to 1-2 percentage points and loses all explanatory power. This is in line with the results of previous studies, namely that nudges, like personalised mails, wear off very fast. On the other hand, some further viewpoints should be taken into consideration. To go deeper into this question, one has to keep in mind that I use within subject randomisation strategy over time, which means that all background information of the individuals are constant over time, except the variation in the treatment. [25] This approach limits the usage of typical panel estimation strategies.

Hypothesis 3 questions whether the effect of personalised letters change the number of participation, or students just reschedule their participation in time. This is a common question in the marketing literature, i.e. whether the effect of the campaign has a long lasting positive revenue effect at all.[26] In the case of the participation is SSP, the sum number of participation is the "revenue", so my question is whether the sum of participation is higher among treated than in the control.

Table 10 uses the approach of Model 2, but here the dependent variable is the sum of the participation of a given individual in the semester after the start of the experiment. *Column 1* shows that without further explanatory variables, the treatments themselves have no convulsive results. The average sum of participation is 0.27, and Treatment 1 increases it with 0.07 while Treatment 2 with 0.06, which is approximately a 25% change. Introduction of

[25] This approach is also called as interrupted time series in Gereben and Green (2012) p. 273
[26] Simester, Hu, Brynjolfsson and Anderson (2009)

semester variable does not change the magnitude of the variables of interest, but increases the level of significance in *Column 2*.

Table 10: Estimation results of Model 2 on the sum of participation

VARIABLES	(1) Sum	(2) Sum	(3) Sum	(4) Sum
Treatment 1	0.0705	0.0789*	0.102**	0.107**
	(0.0446)	(0.0440)	(0.0512)	(0.0517)
Treatment 2	0.0633	0.0613	0.0490	0.0538
	(0.0449)	(0.0443)	(0.0497)	(0.0504)
Semester		-0.0782***	-0.0696***	-0.0723***
		(0.0117)	(0.0174)	(0.0168)
Constant	0.228***	0.359***	-0.0320	-0.0345
	(0.0302)	(0.0363)	(0.0502)	(0.616)
Observations	1,190	1,190	1,190	1,145
R-squared	0.002	0.038	0.303	0.305
Model	ITT	ITT	ITT	CACE
School dummy	NO	NO	YES	YES
p(F-test)	0.877	0.700	0.313	0.287

Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$[27]

In *Column 3* and *Column 4* school dummies are also added to the regressions. This action significantly increases the level of R^2, the magnitude and the significance in the case of Treatment 1. The difference between *Column 3* and *Column 4* is the same as between *Column 2* and *Column 3* in **Table 9**: individuals with bouncing back mails in any round of the experiments are dropped out from the sample. This means that the complier average causal effect of Treatment 1 is 0.10; that is, being in the Treatment 1 group increases the participation by almost 40%, significantly. The magnitude of Treatment 2 is half of Treatment 1, and it is not significant at all, and based on the F-tests I reject that the two treatment groups are significantly

[27] One-sided probability of the treatments values can be found in the Appendix Table 2.

different from each other. According to these results, Hypothesis 3 is rejected: the personalisation letter have a long lasting effect.

CHAPTER 5

CONCLUSION: COMPARING THE RESULTS AND FURTHER RESEARCH POSSIBILITIES

Comparing the results of my thesis with the previously reviewed articles is not an easy task. Due to the randomisation, the internal validity of all the introduced studies are usually very high – this is true to mine as well. Nonetheless, comparability of the results with each other is much harder, since the introduced field experiments are too much imbedded into a unique cultural (i.e. developing world) and situational (i.e. invitation to a job fair) environment. This is definitely against the generalisation of my findings. Hence, comparing the outcomes of my thesis with other outcomes seems to be the most tenuous part of my study.

The short run outcome of my experiment is quite conclusive: adding the name of the recipient to a standard mail increases the participation rate by 9 percentage points, which is a 50% increase, compared to the control group. So the shorter version of the letters, which contained no reminder of goals or reciprocity, has higher significance and larger point estimates than Treatment 2. These all together are closer to the findings of Haynes et al. (2013), where they found 41% increase in willingness to pay delinquent fines due to the personalisation with name, but against the findings of Sanders and Kirkman (2014), who suggest that pure name added mails have much lower impacts.

Despite this result, one would be mistaken to interpret that reciprocity approach or the goal targeting theory does not work. There is a great chance that this very specific subgroup of people (talented students who would like to learn more) is just not receptive to either of them.

Another possibility is that the standard (control) letter is actually at the same level of reciprocity as the medium reciprocity variant of Sanders and Kirkman (2014), so adding a higher level of reciprocity does not change the results as previously accepted. It also possible that I chose the goal unsuccessfully: students were maybe not interested in completing the semester, but in achieving more knowledge or gaining more points in the matriculation exam.

The results of investigating the long-run effects were somewhat contradictory. Monthly outcomes showed that the effect of the treatment vanishes after the first month of implementation. On the contrary, the presence of the treatment was measurable in the sum of participation if I used the CACE setup. The average sum of participation was 0.27, while in Treatment 1 it was higher with 0.1, so it is a 40% increase in the number of participation among the treated. The total number of participation during the research period shows that there is no direct trade-off between the short- and long-run effectiveness of the personalisation.

The above introduced results are in line with the previous findings of the literature, even so the universality of the outcomes is questionable. As a conclusion, I think that despite the unique sample (talented Hungarian secondary school students), my thesis contributes to the broader policy implementation of personalised mails.

Based on these results some further research areas and topics could arise. First of all, the follow-up of the recent cohort of students is a reasonable research strategy. Measuring the differences in the participation during the following year, or number of completed semesters until the graduation can deliver more interesting long run effects. The possibly high level of attrition would cause problems, but long-lasting effect of personalised messages are underpinned in political mobilization literature.

Doing the experiment on a new cohort with a somewhat different approach would be fruitful. This recent experiment made clear that treatment is not vanishing from one month to the other. One possible strategy is to let the effect “wash out”: introducing a no treatment round after the treatments may cause a longer lasting positive effect on the total amount of participation. This approach is widespread in medical literature.[28]

Since the combination of the “goal reminder” and the “reciprocity” theories was not as successful as I wished, it may worth to change to the “social pressure” approach. Sentences, like “90% of students from your county participated in more events like you” may increase the participation more, than the pure name added personalisation. **Figure 2** showed that the lower the number of finished semesters was, the lower the effect of participation resulted. Addressing different mailing strategies to different groups can be considered, by reason of making letters more self-relevant, which causes higher level of anticipation by the students. Using “social pressure” messages, based on the finished semesters, can be one way to do this.

Another possible field to improve is the efficiency of the mail tracking. As I previously described, the fact, whether students actually opens the e-mail or not is measurable, changes the structure of the latent groups. SSP program is recently developing a device which allows the managers to track whether the students read the letters at all, and send reminders to those who have not. This new source of information may increase the precision of the estimation, since I would be able to exclude more students as never-takers if they have not read the mails.

In their seminal paper, Gerber et al. (2010) suggest that placebo treatment increases the credibility of the results, especially, if the share of compliers is small in the dataset. The idea is originated again from the classical medical literature. To translate it into my research environment, asking students to act in accordance with the mail (like sign up to newsletter) can help to make difference between the effect of the mail on other actions, and on the effect on participation in events.

[28] Gerber and Green (2012) section 8.4

BIBLIOGRAPHY

Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996). Identification of Causal Effects Using Instrumental Variables. Journal of the American Statistical Association, 444-455.

Bargh, J. A. (1982). Attention and Automaticity in the Processing of Self-Relevant Information. Journal of Personality and Social Psychology, 425-36.

Bertrand, M., Karlan, D. S., Mullainathan, S., Shafir, E., and Zinman, J. (2005). \What's Psychology Worth? A Field Experiment in the Consumer Credit. *Economic Growth Center, Working Papers 918*.

Cadena, X., and Schoar, A. (2011). Remembering to Pay? Reminders vs. Financial Incentives for Loan Payments. *NBER Working Paper Series*, no. 17020.

Cahill, K., and Perera, R. (2008). Quit and Win contests for smoking cessation. *Cochrane Database of Systematic Reviews, 4*. Forrás: http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD004986.pub3/abstract

Cherry, C. E. (1953). Some Experiments on the Recognition of Speech, with One and with Two Ears. *Journal of the Acoustical Society of America*, 975-979.

Dale, A., and Strauss, A. (2009). Don't forget to Vote: Text Message Reminders as a Mobilization Tool. *American Journal of Political Science*, 787-804.

Dijkstra, A. (2005). Working Mechanisms of Computer-Tailored Health Education: Evidence from Smoking Cessation. *Health Education Research*, 527-539.

Duflo, E., and Saez, E. (2003). The role of Information and Social Interactions in Retirement Plan Decisions: Evidence from a Randomized Experiment. *Quarterly Journal of Economics*, 815-42.

Duflo, E., Geinnerster, R., and Kremer, M. (2007). Using randomisation in development economics research: a toolkit.

Falk, A., and Fischbacher, U. (2006). A theory of reciprocity. *Games and Economic Behavior*, 293-315.

Free, C., Knight, R., Robertson, S., Whittaker, R., Edwards, P., Zhou, W., . . . Roberts, I. (2011). Smoking Cessation Support Delivered Via Mobile Phone Text Messaging (Txt2stop): A Single-Blind, Randomized Trial". *Lancet*, 49-55.

Gerber, A. S., and Green, D. P. (2000). The Effect of Canvassing, Telephone Calls, and Direct Mail on Voter Tournout: A Field Experiment. *The American Political Science Review*, 653-63.

Gerber, A. S., and Green, P. D. (2012). *Field experiments.* London, New York: W. W. Norton & Company.

Gerber, A. S., Green, D. P., Kaplan, E. H., and Kern, H. L. (2010). Baseline, Placebo, and Treatment: Efficient Estimation. *Political Analysis*, 297–315.

Gleerup, M., Larsen, A., Leth-Petersen, S., and Togeby, M. (2010). The Effect of Feedback by Text Message (SMS) and email on Household Electricity Consumption: Experimental Evidence. *Energy Journal*.

Haynes, L., Green, D. P., Gallagher, R., John, P., and Torgerson, D. J. (2013). Collection of Delinquent Fines: An Adaptive Randomized Trial to Assess the Effectiveness of Alternative Text Messages. *Journal of Policy Analysis and Management*, 718-730.

Karlan, D., McConnell, M., Mullainathan, S., and Zinman, J. (2010). Getting to the Top of Mind: How Reminders Increase Saving. *NBER Working Paper Series*, No. 16205.

Kreuter, M. W., Strecher, V. J., and Glassman, B. (1999). One size does not fit all: case for tailoring print materials. *Annals of Behavioral Medicine*, 276-283.

Moray, N. (1959). Attention in dichotic listening: Affective cues and the influance of instructions. *Quarterly Journal of Experimental Psychology*, 56-60.

Morten, M., Karlan, D., and Zinman, J. (2012). A personal touch: text messaging for loan repayment. *NBER Working Paper Series*, no. 17952.

Oliver, A. (2012. 12). A Nudge Too Far? A Nudge at All? On Paying People to Be Healthy. *Healthcare Papers*, *12*, 8-16.

Rubin, D. B. (1974). Estimating Casual Effects of Treatments in Randomized and Nonrandomized Studies. *Journal of Educational Psychology*, 688-701.

Sanders, M., and Kirkman, E. (2014). I've booked you a place. Good luck: a field experiment applying behavioural science to improve attendace at high-impact recruitment events. *The Centre for Market and Public Organisation Working Paper Series* , No. 14/334.

Simester, D., Hu, Y. J., Brynjolfsson, E., and Anderson, E. (2009). Dynamics of Retail Advertising: Evidence from a Field Experiment. *Economic Inquiry*, 482-499.

Wooldridge, J. M. (2008). *Introductory Econometrics: A Modern Approach.* Cengage Learning.

APPENDIX

Appendix Table 1: Descriptive statistics of the used variables and their distribution among the different groups of treatment

	CONTROL	TREATMENT 1	TREATMENT 2	TOTAL
STUDENT #	386	392	412	1190
BUDAPEST	83	96	86	265
%	21,5%	24,5%	20,9%	22,3%
BUDAPEST REGION	68	81	79	228
%	17,6%	20,7%	19,2%	19,2%
COUNTRYSIDE	235	215	247	697
%	60,9%	54,8%	60,0%	58,6%
GIRL #	272	283	270	825
%	70,5%	72,2%	65,5%	69,3%
GRADE				
9	54	45	59	158
%	14%	11%	14%	13%
10	103	116	105	324
%	27%	30%	25%	27%
11	145	116	144	405
%	38%	30%	35%	34%
12	75	104	91	270
%	19%	27%	22%	23%
13	9	11	13	33
%	2%	3%	3%	3%
FINISHED SEMESTERS				
0	101	100	123	324
%	26%	26%	30%	27%
1	71	72	80	223
%	18%	18%	19%	19%
2	131	131	122	384
%	34%	33%	30%	32%
3	46	38	40	124
%	12%	10%	10%	10%
4	19	28	18	65
%	5%	7%	4%	5%
5 OR MORE	18	23	29	70
%	5%	6%	7%	6%

Appendix Table 2: Summary of one-tailed *p* and *z* values

	Column (1)	Column (2)	Column (3)	Column (4)
Table 7				
Treatment 1	0.10	0.01	0.63	0.09
Treatment 2	0.10	0.04	0.19	0.34
Table 9				
Treatment 1	0.01	0.01	0.01	0.28
Treatment 2	0.06	0.10	0.11	0.42
Table 10				
Treatment 1	0.06	0.04	0.02	0.02
Treatment 2	0.08	0.08	0.16	0.14

Appendix Table 3: Marginal effect estimations
in Model 1in different months

		Point	Down	Up
February	Control	0.12	0.09	0.07
	Treatment 1	0.16	0.12	0.07
	Treatment 2	0.16	0.12	0.07
March	Control	0.08	0.06	0.06
	Treatment 1	0.13	0.10	0.07
	Treatment 2	0.12	0.09	0.06
April	Control	0.07	0.05	0.05
	Treatment 1	0.07	0.04	0.05
	Treatment 2	0.09	0.06	0.06
May	Control	0.07	0.05	0.05
	Treatment 1	0.10	0.07	0.06
	Treatment 2	0.08	0.05	0.05

Appendix Table 4: Marginal effect estimations in Model 2
at different levels of the *semester* variable

		Point	Down	Up	Obs.
Semester = 0	Control	0.17	0.10	0.24	217
	Treatment 1	0.29	0.21	0.38	
	Treatment 2	0.23	0.15	0.30	
Semester = 1	Control	0.13	0.06	0.19	116
	Treatment 1	0.23	0.15	0.32	
	Treatment 2	0.17	0.11	0.24	
Semester = 2	Control	0.07	0.04	0.11	191
	Treatment 1	0.15	0.09	0.21	
	Treatment 2	0.11	0.06	0.16	
Semester = 3	Control	0.05	0.01	0.09	77
	Treatment 1	0.11	0.05	0.18	
	Treatment 2	0.08	0.03	0.13	
Semester ≥ 4	Control	0.01	0.00	0.03	50
	Treatment 1	0.03	0.00	0.07	
	Treatment 2	0.02	0.00	0.04	
AVG	Control	0.09	0.05	0.14	651
	Treatment 1	0.18	0.12	0.24	
	Treatment 2	0.14	0.09	0.18	

Appendix Table 5: Translation of the invitation letters

<table>
<tr><td colspan="3">Dear <<NAME>>,

It is our pleasure to invite you to next SSP Saturday event of the Spring semester at 2016/17 term, which is going to take place at <<DATE>>. The participation is free for the members of SSP, but you have to register beforehand.

<<Reminder>>

You can find the registration link here: (link).

For further details please visit the webpage of MCC.

The date of the event: <<DATE>>.

The event would take place at the building of MCC at Somlói Street 51, Budapest.

<<Kind farewell>>

<<Name of the manager>></td></tr>
<tr><td>Variable</td><td>Description</td><td>In experiment</td></tr>
<tr><td><<NAME>></td><td>First name and surname</td><td>It is one of the curtail variable of the letters. It takes “Student” in Control, and the name of the student otherwise.</td></tr>
<tr><td><<DATE>></td><td>date of the SSP event</td><td></td></tr>
<tr><td><<Reminder>></td><td>reminder, why it is important to come</td><td>“We would like to kindly remind you that your semester would be successfully finished only if participate in at least two SSP Saturdays.” It is part of the letter, if student belonged to Treatment 2.</td></tr>
<tr><td><<Kind farewell>></td><td>some kind closing remarks</td><td>“It would make us very happy if we can enjoy your company at our next SSP Saturday. We look forward to it.”</td></tr>
</table>