Assigned seating in students' exams: A field experiment

ERAMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Department of Economics

Supervisor: Prof. Dr. Robert Dur

Second reader: Dr. Sacha Kapoor

Name: Felix Jansen

Exam number: 350346

E-mail address: felixcasperjansen@gmail.com

Abstract

I set up an experiment to examine the effects of assigned seating versus free seating on the multiple-choice part of exam scores during students' bachelor final exams. Using a simple multiple regression, I find that students in the assigned seating group perform significantly better than students in the free seating group. Students sitting in the back of an examination block generally perform worse; this effect disappears when seats are assigned. I furthermore find that giving different students different versions of the exam does not alter their scores. The assigned seating system is easier to use according to the proctors supervising.

Table of Contents

Introduction

Students and universities have different and similar interest concerning exams. The need for a quiet, orderly conduct of exams in the exam hall and the wish to score as high as possible a grade on the exams are similar interest between both parties. Because of the overlap in these interests it is obvious that students are very quiet during exams and they try their best to scores as high as possible.

But interests diverge concerning the methods which students use to achieve these high grades. When they simply study hard, attend lessons and practise well this aligns with the interest of university. But there is a proportion of students who cheat on exams (McCabe, 2005). These students might use different methods, like copying answers from others students with or without their knowledge, using cheat sheets or using a digital method of as unauthorized aid.

In this research, I use an experimental set-up to determine the effects of a cheap, simple way to bring down the likelihood of answer copying between students who know each other. By seating students according to a predetermined plan not known in advance to these students you might eliminate students copying answers from other students with or without their knowledge. Students might favour copying answers from students who they know and prepared together with because they can arrange seating and answer sheet placement so they can copy easily and they know the skill level of the student they are copying from. Both of which are more difficult to execute if you are copying from someone who is unaware.

The aim of this research is to determine whether imposing different placement rules during exams has an effect on students and to determine the size of this effect. Using an experimental setup, I will try to determine whether imposing seating arrangements changes exam results.

How does imposing seating arrangements affect students' exam results?

Having an answer to this question can help the university in determining the feasibility of streamlining the placement process in the exam rooms. It might be easier for the proctors if the students sit in predetermined seats. It will be easier to detect students who failed to register for an exam, rather than sorting this out when the exam is already ongoing. This experiment is also a way of finding out how difficult it will be for the Erasmus University to shift over to a system of assigned seating.

This experiment encompassed a large set of students who took their exams on one of three different economics courses at the Erasmus School of Economics (ESE). The experiment was partly executed to determine if introducing assigned seating would be implementable without being harmful to students' performance.

Theoretical Framework

This thesis falls within the field of research regarding academic performance of students. There is some work done by economists on this topic (Levitt & Lin, 2015), but usually this work is done by educational researchers or psychologist interested in academic performance indicators and stress (Ramirez & Beilock, 2011; Frattaroli, Thomas, & Lyubomirsky, 2011; Oaten & Cheng, 2005). We try to focus on the economic aspects of students' exam behaviour. From the perspective of the university cheating is a clear problem. When it is possible to cheat at a university and students are able to do so, it directly affects the value of the university's degree. This is both detrimental to universities and students, but students won't likely take into account the negative externalities all other students suffer when they cheat and the value of their common degree goes down. University suffer the full consequences of students cheating. Therefore, to protect the integrity of the degree universities institute harsh individual punishments to deter cheating. A student caught cheating can in extreme cases be terminated from enrolment in the programme they are following. At the very least the result for the course they cheated on will be declared invalid. (ssc OO&S and USC/Marketing & Communicatie, 2015)

When considering cheating in an economic framework, usually the monitoring game is named. This game has a principal and an agent. The principal first decides to monitor the agent. The agent then decides to cheat or be honest. Cheating is considered to increase payoff of the agent and reduce payoff of the principal. Monitoring is costly to the principal, but also to the agent. Students need to determine whether cheating is feasible, to do this they would weigh the expected benefits versus the expected costs of cheating.

$$\text{Expected benefits of cheating} = \Pr(Cheating\ successfully) * gains\ of\ cheating$$

$$\text{Expected loss of cheating} = \Pr(\textit{Getting caught cheating}) * \textit{Punishment when caught cheating}$$

Students cheat because they believe that cheating is beneficial to their grade. If cheating would not be beneficial students will not exercise cheating behaviour. To determine whether to cheat a student weighs the expected gain of cheating: the chance of cheating successfully multiplied by the benefits of cheating. This is then weighed against the disadvantages of cheating: the chance of getting caught multiplied by the penalty when getting caught.

To deter students from cheating, the university may change different aspects of this equation, usually focused on the side of expected losses to the student, because the chance of getting caught and the penalties are directly dependent on university policies. The Erasmus University uses different practical methods to lower the chances of cheating successfully, like banning mobile phones or watches during exams. They also carefully monitor toilet visitation for possible sources of cheating. The university can also make it harder to directly copy answers during an exam. The two main methods used to deter direct copying of answers are the assignment of different versions to students who are seated next to each other and the assignment of one proctor to every group of 48 students.

Basis of this experiment

The experiment carried out drew heavily on earlier work done in changing situational factors (assigned seating, different versions) to study students' exam performance (Houston, 1986; Levitt & Lin, 2015). Houston used a metric denoting similar answers between students to determine a metric of cheating. He found that free-seating arrangements lead to students cheating more than those in assigned seating conditions. Levitt and Lin had the opportunity to use data from 200 students taking a midterm and a final exam in one course. They used the same metric of cheating. While this research draws on those experiments, the main difference is the large simple size and diversity in courses tested and the fact that I did not use the same metric due to time constraints.

Testing students during an exam is difficult because exams are important moment during the students' career, therefore not much testing in exam circumstances has been done. It is difficult to carry out experiments using students during their exams, because they must not be intrusive. The only intrusive measures a university can usually take which influence exam proceedings are anti-cheating measures. As these are done to deter cheating, the university has more possibilities to carry these out. The students in this exam are not subject to a measure which is proven to deter cheating. But, the experiment carried out was not intrusive to students.

Hypotheses

Hypothesis one: The relation between assigned seating and exam score

Psychological research usually focuses on the timeframes around exams and is often related to stress (Oaten & Cheng, 2005). Writing about stress just before an exam will increase exam performance (Ramirez & Beilock, 2011). It seems to be a form of coping with exam/academic stress. In the current experiment, with alphabetical seating assignment, the fellow students you would normally prefer to sit next to and talk with before the exam might now be seated far away from you. It is possible not being able to talk to those students that you know well might impair your ability to cope with stress and lower exam results.

Other research focuses on the examinations of students, usually about cheating behaviour during tests (Whitley, 1998; McCabe, 2005). Logically, this is interesting to universities because they would want to minimize cheating during exams. The majority of experiments are done on survey basis. While it proves no problem for researchers to get people to talk about cheating in general, it is more difficult to derive the true extent of cheating during exams. These usually try to identify the main predictors of cheating behaviour. This is summarized very well by Whitley (1998) in his literature review of cheating behaviour in college students.

Whitley (1998) split up cheating factors into five categories. Students' characteristics, students' attitudes towards cheating, general students' personality variables, situational/environmental variables and other variables. Regarding the situational variables, because those are mostly the influences during the experiment in this paper, there is a large body of work by John P. Houston. He has written a series of papers in the 1970's and 1980's concerning situational and attitude factors in students' cheating.

Assigning seats to students during exams reduces the likelihood of cheating (Houston, 1986). Cheating behaviour is also positively related to acquaintanceship; the better you know the student next to you, the more likely it is that you cheat. Assigning seats therefore might impair the ability to cheat. Given these finding in stress- and cheating related literature this is the first hypothesis:

Hypothesis one: Students' exam performance goes down when in an assigned seating setting.

Hypothesis two: The relation between seating location, assigned seating and exam score.

One of the effects of assigned seating by definition is the disability of students to choose their seating location. When choosing a seat, students might have different considerations. If a student wants to cheat, they will want multiple students sitting in front of him. It is even more important who the student has sitting in diagonally in front of him/her, as the student directly in front of them has a different version to him/her. This is common knowledge to students at the Erasmus University.

If a student is looking to cheat, they might want to sit as far away or as far to the back from the proctor as possible. It is known that having more proctors influences cheating behaviour (Kerkvliet & Sigmund, 1999)

Apart from cheating, students might also pick seats for other reasons. Students choosing a spot at the front might be more confident. Students might want to sit at the edge of a block to avoid being in the noisier middle of the block. While there is little theoretical evidence to suggest so, I think that students are likely to differ their seating location according to different unknown factors mentioned above.

Hypothesis two: Students in different seating locations perform differently depending on whether they are in an assigned seating block or not.

The last thing of importance regarding all these hypotheses is the fact that there is no research to be found linking cheating behaviour to a higher grade. Even though the general population assumes that it does. To my knowledge this has not been tested in a real environment experiment. Conventional wisdom would suggest that it is stupid to cheat if you wouldn't get a higher grade out of it. But there could be something said for the fact that it might be better to focus on your exams and not spend valuable time and resources on trying to cheat.

Hypothesis three: The relation between exam form and exam score.

While some do not find that introducing different versions of exams reduced cheating (Kerkvliet & Sigmund, 1999) other find that giving different students different versions of an exam reduced cheating, but only when both questions and answers were shuffled. (Houston, 1983). As shuffling questions, but not answers, is the way in which all three bachelor courses in our experiment set up their second test form I cannot test whether there are differences concerning this. Taking this into account, if scores between the different versions are compared, there should be no difference between them.

Hypothesis three: Students' performance is not influenced by the different versions of the same exam.

Method

Experiment

All courses of the bachelor Economics and Business Economics, given at the Erasmus School of Economics (ESE), are taught both in English and Dutch, because there are a Dutch and an international version of the bachelor. The international bachelor in economics and business economics (IBEB) is open to Dutch speaking students, as long as they can speak and write English. In practise, these bachelors are the same in every aspect apart from the used language and the characteristics of the student populations.

The course load of a typical course in a bachelor of Economics and Business Economics is either four or eight ECTS (one ECTS averages about 25-30 hours of work), two of these make twelve ECTS over eight weeks, and five bimesters make 60 ECTS in a year. The total course load of a bachelor's degree spans 3 years and 180 ECTS.

To determine what the effect of forced seating on students is the following experimental setup was used. The experiment was designed so that students can proceed as they normally would, so that their exams do not get disturbed, but with the added benefit that they would not be notified of the fact they are in an actual academic experiment. Three bachelor final exams were selected, one from each year of the bachelor to run this experiment. All three courses are taught in the last period of the academic year at the Erasmus University. All three courses are taught via the Erasmus School of Economics. I will first give a short overview of each course and the normal course of event during exams. After this I will discuss the experimental setup.

Bachelor Course 1

The selected bachelor course from the first year was called ‘Organisation & Strategy’. Its goals were: ‘*To identify, both theoretically and empirically, the boundaries of the firm’* and *‘To identify and analyse the strategic reactions of firms to changes in the external environment*.’ The course had a course load of 8 ECTS and classes were taught both in its Dutch and IBEB form by the same lecturer.

Bachelor Course 2

The selected bachelor course from the second year was called ‘Intermediate Accounting’. Its goals were: ‘*To develop understanding of the theory and practice of providing internal information (Management Accounting) and external information (Financial Accounting), necessary for decision making purposes by stakeholders of an organization. This course builds on the Bachelor I introduction.*’ and ‘*To develop an understanding of the basic terminology, methods and techniques, both from a theoretical and empirical point of view.*’ The course had a course load of 8 ECTS and the Dutch and the IBEB version were taught by **different lecturers**.

Bachelor Course 3

The last selected bachelor course from the third year was called ‘Philosophy of Economics’. Its goals were: ‘*Students have basic knowledge of the significant views, ideas and lines of arguments in the philosophy of economics.*’, ‘*Students have basic knowledge of the significant views, ideas and lines of arguments in the methodology of economics.*’ And ‘*Students have basic knowledge of the significant views, ideas and lines of arguments in the areas in which economics and ethics overlap.*’ The course had a course load of 4 ECTS and the Dutch and English version were taught by different lecturers. This was also the only course of the three where the exams differ apart from language.

Final exams for all 3 courses were taken in week 25 of 2016, more precisely on 21-06-2016 (Ba1 and Ba2) and 22-06-2016 (Ba3). Students could re-take these exams in July of the same year, but a student can **only re-take** three exams per academic year.

Normal Exam setup

At the Erasmus School of Economics (ESE) students sit exams at the end of each bimester. Exams are taken in a hall, in which there is space for a maximum of 22 blocks of 48 chair-and-table combinations. Each table can be used by one student making an exam. Students apply for an exam and are then distributed over the blocks in the exam hall according to student number. Each block will be assigned a range of student numbers e.g. 123456 to 145678, in the subsequent block students in the next range will be seated up until all block were filled up.

The time limit set for exams at the end of a course is usually 180 minutes. Students can enter the exam hall from 15 minutes before the exam. They look up their assigned block and take place at any of the 48 seats they want. Students are allowed to leave when they have finished their exams, but not before 60 minutes have passed or in the last 15 minutes to prevent interrupting any people who are still working. Exams are distributed in different versions. The official requirement for each course is that there are at least two different versions of an exam to prevent cheating. For these selected courses, there were two versions of each exam. Usually teachers generate a second version of their exam by alternating different blocks of questions, so for instance that for version two questions 1 through 15 equal question 16 through 30 on version one and vice versa. In all three selected courses version 2 was adapted from version 1 using a simple conversion key.

Each examination block is supervised by a proctor who hands out the exam papers, checks attendance and generally makes sure proceedings in a single block are going well. The proctor hands out exam papers in a pre-determined way. The tables are laid out according to a standard table plan (Table 1). A proctor will walk a route according to the table numbering. They will hand out exams according to this route at the start of the exam. They will usually hand out the different versions according to this route. Table one receives version one, table two receives version, table three receives version one again, and so on. This is done in order to make sure tables next to each other and tables behind each other do not receive the same version. When this is done, proctors walk past each table with a clipboard to check whether the student has registered for the exam. If students did not register for an exam, they will still be allowed to finish their exam, but the proctor will mark their exam with large red letters reading '**OWN RISK**' and students will be told they had to register with the exam administration in order to receive their marks.

Table 1. Table plan.

7	8	21	22	34	35	48
6	9	20	23	33	36	47
5	10	19	24	32	37	46
4	11	18	25	31	38	45
3	12	17	26	30	39	44
2	13	16	27	29	40	43
1	14	15	Extra table	28	41	42
			Proctor Table			

Note: Students sit at the north end of the tables. Treatment groups were numbered with small numbers written on the tables in the top right corner

Experimental setup

The blocks were paired starting at block one and per pair one block was randomly assigned to the control group and one to the test group at random (Table 2). The result of this was that for this experiment the treatment-control groups were spread over the blocks (Table 3). The students whose blocks were assigned to be control blocks were allowed to sit wherever they want within their block. The students whose blocks were assigned to be treatment blocks had tables marked which were assigned to them. The students in the assigned blocks were matched to tables in alphabetical order, because this was the best balance between ease of execution and true randomness. Tables were numbered in accordance with the normal route in which proctors walked past the tables (Table 1). In the first few minutes after the start of each exam a proctor walked past the tables in this route. The purpose of this was to verify that the correct student was sitting at the correct marked table in the treatment group and for both the treatment and control group to make sure that students were registered to make the exam. Students were identified by their student cards, which contain their name, student number and picture.

Table 2. Exam blocks randomization.

Exam hall block	Treatment
1	1
2	0
3	0
4	1
5	0
6	1
7	1
8	0
9	1
10	0
11	1
12	0
13	0
14	1
15	1
16	0
17	0
18	1
19	1
20	0
21	0

Note: Treatment=1 is a treatment block. Blocks were randomized in pairs.

Table 3. Distribution of treatment blocks.

Course	Ba1	Ba1	Ba2	Ba2	Ba3	Ba3
Language	NL	IBEB	NL	IBEB	NL	IBEB
Number of blocks	9	4	9	5	9	7
Exam Block 1	T	T	T	T	T	
Exam Block 2						T
Exam Block 3						T
Exam Block 4	T	T	T	T	T	
Exam Block 5				T		
Exam Block 6	T		T		T	T
Exam Block 7	T		T		T	T
Exam Block 8						
Exam Block 9	T		T		T	

Note: T denotes a treatment block. The rest of the blocks are control blocks.

In each of the blocks that were assigned to the treatment groups a note was placed on each table (Appendix 2) explaining the new method of assigning seats to them, without mentioning this was an experiment. The lists informing students of their assigned table number (Appendix 1) were placed on a small sign standing there to identify the blocks. In the control blocks, there were no special measures taken and students just proceeded as normal. They did not receive any notes or other information. All exams were no different from normal.

The proctors were instructed to conduct exams in as normal a way as possible, given the experiment, to prevent them having any effect on the outcome of the experiment.

When students had forgotten to register for an exam, usually they would still take a seat in the block in which they were assigned, with the option to register for the exam in a short time-span after the end of the exam. This was not possible in the blocks which were assigned to the treatment group. When all tables are assigned to individual students and an unregistered student shows up early and took the place of the actual student assigned to that table that would cause problems. Therefore, students were directed by proctors to report to the head-proctor if their name did not show up on the list of assigned tables. Then they would be asked to seat themselves in a block specially reserved for them (these were dubbed overflow blocks). There they could make the exam as they normally would, apart from having to register afterwards like other students who did not register.

The experiment worked as intended. There occurred no issues or problems worth mentioning. The main difficulty for students and proctors was the fact that unregistered student would notice that they weren't registered before the exam started instead of during the exam. This was handled well and quickly by the proctors.

Data

After the three exams were completed the seating data was matched with the given answers on the exams for each student. The correct answers for the multiple-choice parts of each exam were received from the individual course supervisors (e.g. the (assistant-) professors). The amount of multiple choice questions differed for each exam, but always constituted the majority of total questions (Table 4). The IBEB exam for the Ba3 course also included a few open-answer questions as opposed to the NL version which consisted of 100% multiple-choice questions. The dataset also included information on the version of each exam students had.

A score was then computed concerning the multiple-choice answers. This score is a percentage of correct answers out of the total number of possible correct answers for that exam. Afterwards, the resulting dataset had a sample size of 2271. This does not represent 2271 different students, because some were doing two or even all three of the exams selected for this experiment.

Table 4. Percentage of Multiple-choice questions

Exam		% of questions are MC	% of pages MC
Ba1	NL	94%	84%
	IBEB	94%	84%
Ba2	NL	100%	100%
	IBEB	100%	100%
Ba3	NL	100%	100%
	IBEB	97%	82%

Note: In the third column, the number of multiple-choice question as a percentage of total questions are given. In the fourth column, the number of pages containing multiple-choice questions as a percentage of total number of pages containing questions are given

Data excluded from the analysis

Due to a randomizing mistake the block were not assigned to treatment-control groups at random. The block dubbed number one at the front of the exam hall was always a control block due to the method used. As all Dutch exams started at block one in the actual exam hall, the assignment of blocks between treatment and control did not differ for the Dutch exams. Block one was therefore always a treatment block. When looking at the effects of treatment on score, I had to include block controls. This is necessary because students are distributed in blocks according to student numbers, a higher student numbers means a student is seated in a block with a higher number. As student numbers are not assigned randomly, but chronologically according to starting date at university, I have to control for hidden characteristics associated with being a student who started earlier at university. Due to the assignment of treatment blocks with the Dutch exams, this could not be done and I had to leave the Dutch students out of the analysis in this paper.

There were also a few other sets of students left out of the analysis. This was done because of two reasons. Firstly, the overflow blocks were not full. They contained students which were unregistered and moved there, and sometimes contained only a few students. Because of this they were not suitable for this analysis. Secondly there was a group of students registered for the exam, which took their exams in a separate room because they receive extra time or extra materials due to dyslexia. These students were also left out of the analysis

In the treatment groups, another set of students was specifically left out of the analysis. Not registered as 'present' in their normally assigned block, these students were moved to the overflow blocks. Blocks can only contain 48 people normally, when 5 unregistered students showed up, they would usually still take a seat in that block and be reminded they need to register after the exam. They might not even know they failed to register, because in control blocks students only know where to sit because blocks are assigned a range of student numbers (e.g. 123456 to 145678). Usually this not an issue, because people who have registered will also fail to show up, leaving enough room. Once in a while this would been an issue because 53 students try to seat themselves in blocks where there is only room for 48. Because in the experiment these students were moved to the overflow block, but they sometimes showed up double in the data. Seeing as these students were not seated where the data might suggest they were, they were left out.

When all these groups were eliminated, there was a sample size of 641 left over from the original sample size of 2271. This means about 72% of the original dataset was eliminated. This was mostly due to the mistake with the randomization of the Dutch blocks described above. In the next part I will describe the data gathered in more detail.

Data

Table 5 shows the mean, standard deviation, and sample size for scores amongst a number of categories. This data is examined briefly in this section. The dataset is examined with respect to two factors, versions and seating location (both between and within blocks). These analyses are done because they give some indication why I would include these variables as controls in the regression analysis. A significance threshold of .10 is used for the entire thesis unless otherwise stated.

Courses

There was no statistically significant difference between scores for Ba1, Ba2 and Ba3. This was determined by a one-way ANOVA ($F(2, 638) = 1.08$, p-value = .34). When the dataset was split into treatment and control groups, means could be compared again. There was a statistically significant difference between scores for Ba1, Ba2 and Ba3 in the control group subset (F (2,

324) = 2.64, p-value = .07) but no statistically significant difference in the treatment group subset (F (2, 311) = 0.12, p-value = .88).

Blocks

Using a one-way ANOVA test it was determined that there was no significant difference between the five different exam blocks (F (4, 636) = 1.91, p-value = .11). There was no significant difference between blocks in the control subset (F (4, 322) = 1.541), p-value = .19) nor between the blocks in the treatment subset (F (4, 309) = 0.49), p-value = 0.74).

Versions

There was no statistically significant difference in means between individuals in groups with different versions (t (639) = 0.16, p-value = .87). There were also no significant differences in means between groups with different versions in the control subset (t (325) = 0.09, p-value = .93) and the treatment subset (t (312) = 0.26, p-value = .80)

Seating location within blocks

There was no significant difference between the different sets of rows (F (2, 634) = 1.33, p-value = .24. There was a significant difference between the different sets of rows for the subsets control (F (6, 320) = 1.862, p-value = .09) but no significant difference between rows in the treatment blocks (F (6, 307) = 0.397, p-value = .88).

Variance.

The variance of the control group (M=69.31%, SD= 12.23%) was not significantly different from that of the treatment group (M=69.80%, SD= 11.59%) as determined by a Levene's test for equality of variances (F=0.94. p-value = 0.33). There was no significantly larger spread of score with the control group as compared to the treatment group.

Table 5. Mean scores, standard deviation and sample sizes of multiple-choice parts on exams. Split amongst predictor variables split by treatment.

Variable name		Control groups			Treatment Groups	
	M	SD	N	M	SD	N
Overall	71.33%	12.06%	327	73.10%	11.23%	314
Ba1	69.13%	14.19%	96	73.59%	12.34%	91
Ba2	73.07%	14.76%	92	72.95%	12.50%	132
Ba3	71.69%	7.48%	139	73.21%	12.41%	91
Version 1	71.39%	11.93%	170	73.27%	11.12%	151
Version 2	71.27%	12.25%	157	72.94%	11.36%	163
Block 1	69.04%	7.31%	47	72.32%	13.67%	85
Block 2	69.78%	14.40%	93	72.34%	8.07%	45
Block 3	72.31%	14.70%	95	73.31%	7.32%	46
Block 4	72.27%	7.40%	47	73.12%	11.31%	91
Block 5	73.87%	7.03%	45	75.00%	12.11%	47
Row 1	75.97%	10.17%	42	72.89%	11.14%	38
Row 2	71.23%	11.28%	49	73.84%	10.99%	48
Row 3	73.17%	9.04%	47	72.78%	11.15%	46
Row 4	68.75%	13.88%	49	70.89%	12.48%	43
Row 5	69.95%	12.63%	48	73.32%	11.85%	46
Row 6	69.74%	13.37%	46	74.04%	9.51%	47
Row 7	71.10%	12.41%	46	73.73%	11.83%	46

Analysis

Functional form

A functional model to estimate the effect of our independent variable on our outcome variable is found. The continuous outcome variable, exam multiple-choice score is denoted by score in a percentage. There were 6 predictor variables. 1 variable was ordinal (students' location), 2 variables were dichotomous (treatment and version), and 3 were categorical variables (course and block location). Our main predictor variable, treatment, is a dichotomous variable. Therefore, a simple multiple regression analysis was used to estimate the effects of the treatment on exam scores. The functional form of the formula of the estimation is as follows:

$$\gamma_i = \boldsymbol{\chi}_i \beta + \eta_i$$

Where γ_i represents the dependent variable, which is the score in the multiple-choice part of the exam. $\boldsymbol{\chi}_i$ is a vector containing the various predictor variables, the variables of interest herein are treatment, version, course, block, row and row*treatment dummy's. β denotes the coefficients associated with these variables. Finally, η_i is the unobserved error term. This is assumed to be drawn from a normal distribution with mean zero. It is also assumed to be independent from χ_i. The dataset and setup described above were used to estimate a series of regressions using IBM SPSS statistics 23.

Results

Assumptions

For a multiple regression to give solid predictions about the effect size, the following four assumptions needed to be met. The error term needed to have constant variance. This assumption is checked by plotting a scatter of the regression standardized residuals on the y-axis versus the standardized predicted values on the x-axis (Figure 1). The errors appeared to be homogenous and the data seemed to have a linear relation.

It needed to be checked whether the errors are random normally distributed. To check for this, a histogram was plotted with residuals on the x-axis and the frequency of the dependent variable (score) on the y-axis (Figure 2). For the errors to be normally distributed this histogram had to follow a normal distribution. This was the case for our regression analysis. Alternatively, the normal P-P plot (Figure 3) could be checked. This plotted the standardized residual with expected cumulative probability on the y-axis and observed cumulative probability on the x-axis. For the errors to be normally distributed, the dotted line needed to run along the diagonal. This was also the case. These two methods confirmed each other's findings of normally distributed errors.

There were two methods to check for outliers. One was to plot a scatter of the dependent variable (y-axis) and the residuals (x-axis) (Figure 4). There were outliers to the left bottom corner. One outlier was seemingly about 5 times the standard deviation (SD) of the dependent variable. The second method to identify this, was to check the outliers straight from the regression diagnostics. There was one particular student on one exam with a score (15%) -4.60 SD from his or her predicted score. This outlier was more than the usually used limit of 3 SD's from the predicted score, but it was not unexpected (some people might choke on exams). I chose not to delete this. The rest of the outliers were all about 3 SD's or less from their respective predicted scores. It was not necessary to delete these from the dataset.

Finally, the regression needed to be checked for multicollinearity. This was done by looking at the VIF statistics on the regression estimation results. When these did not exceed 10 multicollinearities could be ruled out. This was not the case; the highest statistic is 2.19. Therefore, the conclusion was that there is no multicollinearity.

Regression analysis

In order to test the hypotheses formulated in the first part of this thesis, models one through three were used to estimate hypothesis one. Models four and five were used to test hypothesis three. Model six was used to formulate an answer to hypothesis two. All six estimations can be found in Table 6.

Table 6. Regression estimations with predictors of exam multiple-choice score as the dependent variable.

	1	2	3	4	5	6
Treatment	1.773	1.696	1.718	1.697	-3.010	-3.012
	(0.922)	(0.934)	(0.981)	(0.979)	(2.625)	(2.628)
p-value	0.06	0.07	0.08	0.08	0.25	0.25
Version						0.031
						(0.925)
p-value						0.94
Course Controls	NO	YES	YES	YES	YES	YES
Block Controls	NO	NO	YES	YES	YES	YES
Location controls	NO	NO	NO	YES	YES	YES
Location*treatment interactions	NO	NO	NO	NO	YES	YES
# of predictors	1	3	7	13	19	20
N	641	641	641	641	641	641
R2	0.006	0.009	0.018	0.031	0.040	0.040
R2 adj.	0.004	0.004	0.007	0.011	0.011	0.009
F	3.703	1.823	1.69	1.549	1.366	1.296
p-value	0.055	0.142	0.108	0.095	0.136	0.174
F Change (From)		0.884	1.585	1.377	0.972	0.001
p-value		0.414	0.177	0.221	0.443	0.973

Note: Standard errors reported in parentheses.

Testing hypothesis one: The relation between assigned seating and exam score

Model one

The effects of treatment on score were estimated using our regression setup. In model 1 only *Score* was regressed against *Treatment*. A regression with significant joint prediction power (F (1, 639) = 3.703, p = 0.06) was found with a R^2 of 0.01 and an Adjusted R^2 of 0.00. The predicted effect of treatment was 1.77%.

There were differences between the mean scores of students when they were split along treatment groups. This might be because some courses are more difficult compared to others (4 vs. 8 ECTS), they had different lecturers with different skills or by luck one teacher had made a difficult exam and the other teacher had made an easier exam that year. These differences had to be accounted for to better understand the effect of treatment.

Model two

The effects of treatment on score were estimated again, but now the dummies for the courses were included, leaving one out to prevent multicollinearity. A regression with non-significant joint predictive power (F (3, 637) = 1.82, p = .14) was found, with a R^2 of 0.01 and an Adjusted R^2 of 0.00. There was a non-significant increase in predictive power (F (2, 637) = 0.88, p = .41). The predicted effect of treatment was 1.70% and this was significant (t (637) = 1.82, p = .07).

While there were no significant differences between blocks, they are included in model three, because I have to account for them to prevent any hidden students' characteristics influencing the results.

Model three

To estimate the effect of sitting in a block with assigned seating when correcting for blocks, 4 dummy variables were added (for 5 blocks, one left out) and the regression was estimated for a third time. This estimated a regression with non-significant joint predictive power (F (7, 633) = 1.69, p-value = .11) and a R^2 of 0.02 and an adjusted R^2 of 0.01. There was a significant increase in predictive power (F (4, 633) = 1.59, p-value = .18). The predicted effect of treatment was 1.72% and it was significant (t (633) = 1.75, p-value = .08). From the control predictors values I see that sitting in a block with a higher number corresponds with higher grades.[1]

Testing hypothesis two: The relation between Seating location, treatment and exam score

Model four

To determine the effect of sitting in a different location within a block, 6 dummies were included denoting the seven rows in a block, with the first row left out. The predictors here, in model four, had significant joint predictive power (F (13, 627) = 1.55, p-value = .10), with a R^2 of 0.03 and with an adjusted R^2 of 0.01. There was no significant increase (Compared to model 3) in predictive power (F (6, 627) = 1.38, p-value = .22). The predicted effect of treatment was 1.70% and it was significant (t (627) = 1.73, p-value = .08). From the control predictors values you see that students sitting towards the middle and back of the block score lower on average given whether they are in a treatment block. [2]

The location predictors were interacted with the treatment predictor and these dummies (leaving the first row*treatment interaction out as a base to prevent multicollinearity) were added. Because of this it was possible to compare the effect of being more towards the back of a treatment block versus being more towards the bock of a control block more closely.

Model five

Using these new dummies model five was estimated. This regression did not have significant joint predictive power (F (19, 621) = 1.37, p-value = .14), with a R^2 of 0.04 and with an adjusted R^2 of 0.01. There was no significant increase in predictive power (F (6, 621) = 0.97, p-value = .44.

---

[1] The predictor values for the block controls had the following values: Block 2 0.01%, Block 3 2.02%, Block 4 1.65% and block 5 3.24%. Only the value for block 5 was significantly different from 0 at the 0.10 level. These values stayed constant through models three, four, five and six.

[2] The predictor values for the row controls had the following values: Row 2 -2.07, Row 3 -1.62, Row 4 -4.84, Row 5 -2.99, Row 6 -2.65, Row 7 -2.22. The values for row 4 and row 5 are significantly different from 0 at the 0.10 level.

When looking at the effect sizes; the predicted effect of treatment was now -3.01% and it was non-significant (t (621) = -1.15, p-value = .25). Looking at the control values[3] students towards the back score lower on average when they are free to choose their seat. (Table 7)

Table 7. Score difference between students with free vs. assigned seating (Means taken from block 1)

	Control	Treatment	Difference (Control – Treatment)
Row 1	0.00%	-3.01%	3.01%
Row 2	-4.74%	-2.16%	-2.58%
Row 3	-2.82%	-3.22%	0.41%
Row 4	-7.22%	-5.21%	-2.01%
Row 5	-5.99%	-2.72%	-3.27%
Row 6	-6.16%	-1.93%	-4.23%
Row 7	-4.86%	-2.34%	-2.52%

Note: Mean difference is -1.60%

Testing hypothesis three: The relation between exam form and exam score

In order to test hypothesis two, it had to be checked whether having a different version of the exam had an effect on a student´s grade. Version two was the exam adapted from the original in all three courses. To check if there is a disadvantage of having questions in a different (illogical) order one dummy was added to the model.

Model six

To correct for different versions, one dummy was added which takes the value of one if you have version number two and estimate the regression for a sixth time. This regression did not have significant joint predictive power (F (20, 620) = 1.30, p-value = .17), with a R^2 of 0.04 and an adjusted R^2 of 0.01. There was no significant increase in predictive power (F (1, 620) = 0.00, p-value = .97). The predicted effect of treatment was -3.01%, and non-significant (t (620) = -1.15, p-value = .25). The predicted effect of having version 2 was 0.03% and non-significant (t (620) = 0.03, p-value = .97)

[3] The predictor values for the row and row*treatment controls had the following values: Row 2 -4.74 Row 2 interaction 5.59, Row 3 -2.82 Row 3 interaction 2.60, Row 4 -7.22 Row 4 interaction 5.02, Row 5 -5.99 Row 5 interaction 6.28, Row 6 -6.16 Row 6 interaction 7.24, Row 7 -4.86 interaction 5.53. Only the value for Row 3 is not significantly different from 0 at the 0.10 level. Only the value for rows 5 and 6 interacted are significantly different from 0 at the 0.10 level.

Discussion and Conclusion

Hypothesis one: The relation between assigned seating and exam score

There is a relation between assigned seating and exam score. It seems that those students who were sitting in a block with assigned seating on average performed better than those sitting in free seating block. Controlling for course (difficulty, course size, amount of MC questions in the exam), and some hidden students' characteristics (students are sometimes grouped together in blocks based on starting time at the university), there is a positive relation between being in a treatment group and scoring higher on the multiple-choice part of the exam. Overall it can be concluded there is a positive relation between exam score and assigned seating. Students in treatment blocks score on average about 1,6% higher than students in the control blocks. This would contradict hypothesis one (*Students performance will go down when in an assigned seating setting*).

A possible explanation for this is the fact that students have some time to focus on their exam before started, not being distracted socially. While for some students being distracted socially might relieve stress, for other this might not be the case.

Hypothesis two: The relation between seating location, treatment and exam score.

When looking at the differences for students as a result of their seating choice there is a stark difference between students who can choose to sit wherever they want and those who can't. In a treatment block, people sitting at the back have a higher score, on average, than people sitting in the back rows of a control group. This contradicts the second hypothesis (S*tudents who are seated in two different locations in the same block perform the same, regardless if they are in the free or assigned seating blocks.)* These effects are only partially significant.

One reason could be that when students are confident in their ability they gravitate towards the front of the room and vice-versa. It could also be that students who are planning on cheating (because of a lack of confidence) gravitate towards the back, further away from the proctor (and doing worse on their test). This would suggest there are more effects from the experimental setup to be found within blocks.

Hypothesis three: The relation between exam form and exam score.

From the result of model 5 it appears that someone who receives version two has, on average, an equal score to someone who has received version one. This confirms hypothesis two (*There is no difference between students making the same exam in a different form i.e. between different versions).* While this may seem to state the obvious, it gives some insights in the practise of using different versions in exams. We can safely conclude that in these exams, having version 2 gave students no disadvantage over students who received version one.

Limitations study

One of the limitations of this research was the availability of personal data. The only method of identification was a randomized number which was explicitly not the student number. There was no information regarding the students other than what they did in the exam room and where they sat. There was not information regarding age, gender, seniority, average academic performance or other personal factors. When there is information on these aspects this could be used to investigate the personal relations with sitting in an assigned seating block.

Experimenting during exams

When executing this experiment, it was important to be careful to make the experiment as inconspicuous as possible, because I wanted students to focus on their normal exams. It cannot be ruled out that the students in the treatment blocks were influenced by the changes, but as the influence of being in a treatment block was mostly positive, it is not necessary to worry about negative effects.

Practical implications

Proctors did experience a positive effect of knowing exactly which student was seated where. They said they made fewer mistakes checking the lists of students. Making it easier for proctors to find students on this list left more time to check their block for eventual cheating. It was also easier to filter out those students who did not register, so they could be assigned to one block. Every proctor spoken to has confirmed that the new system would make their work easier and more effective. This came both types of proctors, those using the old and the new system.

Recommendations for future research

If there was a chance to re-do the experiment, I would take more care randomizing the data. The dataset is quite large, but the number of blocks isn't as large. It is therefore important that the assignment of the blocks is done purely random to ensure each block is present in both the control and the treatment groups. Due to a mistake on my part, this was not the case. A true randomization would be to make a new coin flip for every pair of blocks you try to assign. The assignment of blocks would be truly random, and there would be little correlation between treatment and block number. This would be more suited to the regression estimation and this might have made the eventual findings more significant.

The amount of individual data gathered is large, there are still possibilities of analysing students' exam behaviour. Especially in the area of potential fraud there are research options where you would be able to use this dataset. The overall results and conclusions from this experiment point towards aspects of interest on the within-block level. A chunk of the magnitude of the predictor for treatment shifted to location effects when they were included in model 4 and 5. Using the methods from Houston (1986) or, more recently Levitt and Lin (2015) one could calculate a measure of relative performance amongst the different students in this dataset. You would calculate the number of matching incorrect answers. So two students both incorrectly answering D while the correct answer is A, would be highly suspicious. Due to difficulty and time constraints I was not able to calculate this measure and use it. This could be used to determine if any cheating had been going on during the exams. The analysis would be to compare student A with student B sitting next to him and some random student C sitting anywhere else. Using this measure you can check for cheating and see if assigned seats reduced cheating.

Conclusion

When taking these results into account you can see that there is a significant positive effect of putting students in assigned seats. There is not only a small effect on their score, but also an effect with the general ease of use of the system. I would recommend on the basis of this data to implement the system of assigned seating, as I've found no evidence of negative effects and it eases the work of the proctors. There is more to be found in this data, but for now this research has shown that there are certainly possibilities of using actual students' exams for doing interesting research into the conduct of examinations, as long as these experiments are non-intrusive.

References

Frattaroli, J., Thomas, M., & Lyubomirsky, S. (2011). Opening up in the classroom: Effects of expressive writing on graduate school entrance exam performance. *Emotion, 11*(3), 691-696.

Houston, J. (1976). The assessment and prevention of answer copying on undergaduate multiple-choice examinations. *Research in higher education, 5*, 301-311.

Houston, J. (1983). Alternate test forms as a means of reducing multiple-choice answer copying in the classroom. *Journal of Educational Psychology, 75*(4), 572-575.

Houston, J. (1986). Classroom answer copying: Roles of acquaintanceship and free versus assigned seating. *Journal of Educational Psychology, 78*(3), 230-232. doi:10.1037//0022-0663.78.3.230

Kerkvliet, J., & Sigmund, C. J. (1999). Can We Control Cheating in the Classroom. *The Journal of Economic Education*, 331-343.

Levitt, S. D., & Lin, M.-J. (2015). Catching cheating students. *National Bureau of Economics Research*.

McCabe, D. L. (2005). Cheating among college and university students: A North American perspective. *International Journal for Educational Integrity, 1*(1).

Oaten, M., & Cheng, K. (2005). Academic examination stress impairs self-control. *Journal of Social and Clinical Psychology, 24*(2), 254-279.

Ramirez, G., & Beilock, S. (2011). Writing about testing worries boosts exam performance in the classroom. *Science*, 211-213.

ssc OO&S and USC/Marketing & Communicatie. (2015, August 1). *Cheating and plagiarism in assignments, theses and exams.* Retrieved March 7, 2017, from Erasmus University

Rotterdam: https://www.eur.nl/fileadmin/ASSETS/ieb/corp-publicaties/pdf/webbrochure2015_Cheating-plagiarism_EN.pdf

Whitley, B. E. (1998). Factors associated with cheating among college students: A review. *Research in higher education, 39*(3), 235-274.

Appendix

Appendix 1. Example of a list present at the front of each treatment block informing students of their assigned table.

FEB12002X Philosophy of Economics			**22-6-2016** 18:30 - 21:30
Studentnr.	**Naam**	**Tafel**	
123123	John Doe	1	**Block 1**
123124	Jane Doe	2	
123125	John Doe	3	
123126	Jane Doe	4	
123127	John Doe	5	
123128	Jane Doe	6	
123129	John Doe	7	
123130	Jane Doe	8	
123131	John Doe	9	
123132	Jane Doe	10	
123133	John Doe	11	
123134	Jane Doe	12	
123135	John Doe	13	
123136	Jane Doe	14	
123137	John Doe	15	
123138	Jane Doe	16	
123139	John Doe	17	
123140	Jane Doe	18	
123141	John Doe	19	
123142	Jane Doe	20	
123143	John Doe	21	
123144	Jane Doe	22	
123145	John Doe	23	
123146	Jane Doe	24	
123147	John Doe	25	
123148	Jane Doe	26	
123149	John Doe	27	
123150	Jane Doe	28	
123151	John Doe	29	

123152	Jane Doe	30
123153	John Doe	31
123154	Jane Doe	32
123155	John Doe	33
123156	Jane Doe	34
123157	John Doe	35
123158	Jane Doe	36
123159	John Doe	37
123160	Jane Doe	38
123161	John Doe	39
123162	Jane Doe	40
123163	John Doe	41
123164	Jane Doe	42
123165	John Doe	43
123166	Jane Doe	44
123167	John Doe	45
123168	Jane Doe	46
123169	John Doe	47
123170	Jane Doe	48

Appendix 2. Note to students informing them of the different exam setup. Dutch students in treatment groups got a Dutch version of this note, which was an exact translation.

Dear Student,

For this exam, the tables are numbered and allocated in order of student number. At the head of your block there is a list of table numbers. Check whether you are seated at the correct table.

Figures

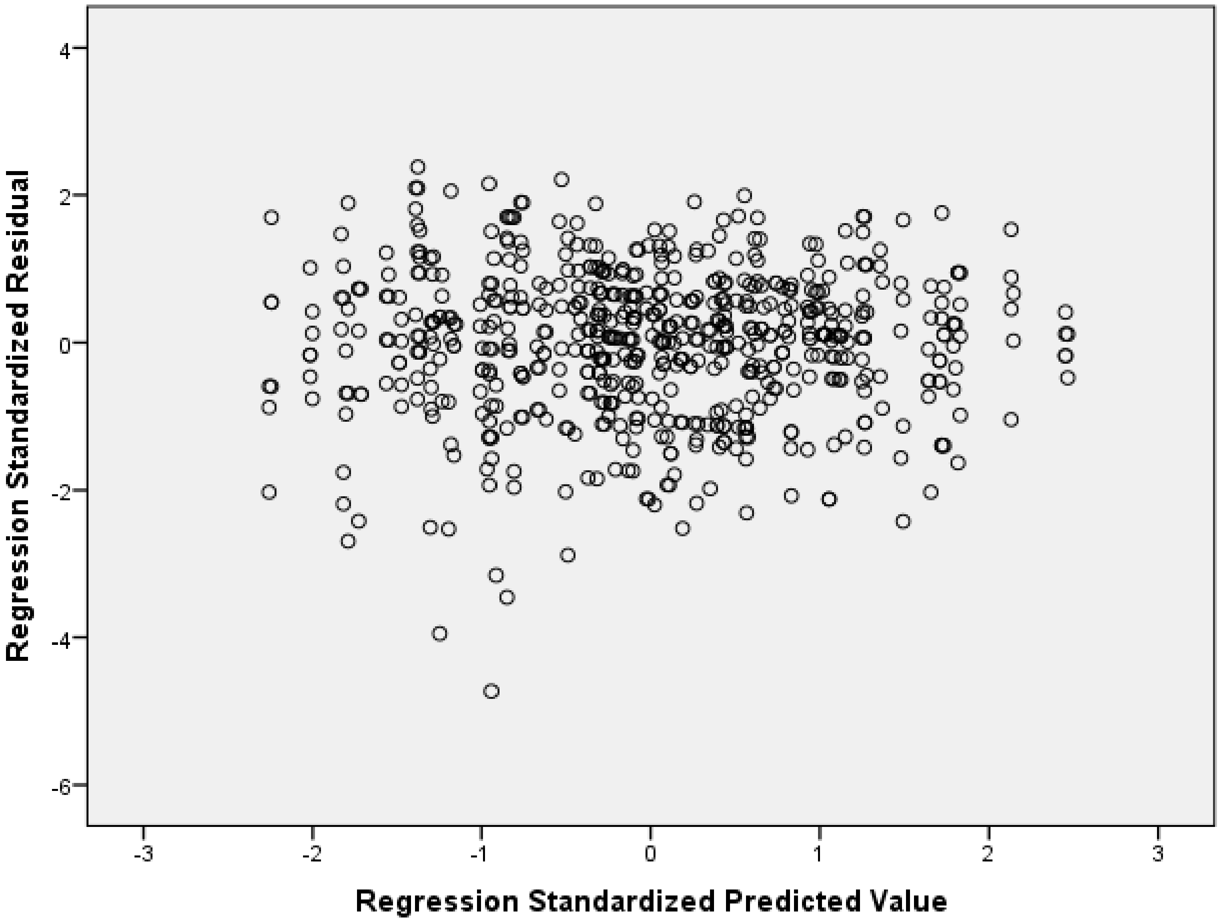

Figure 1. Scatter of standardized residual (y-axis) plotted versus standardized predicted values. (Dependent Variable: Score)

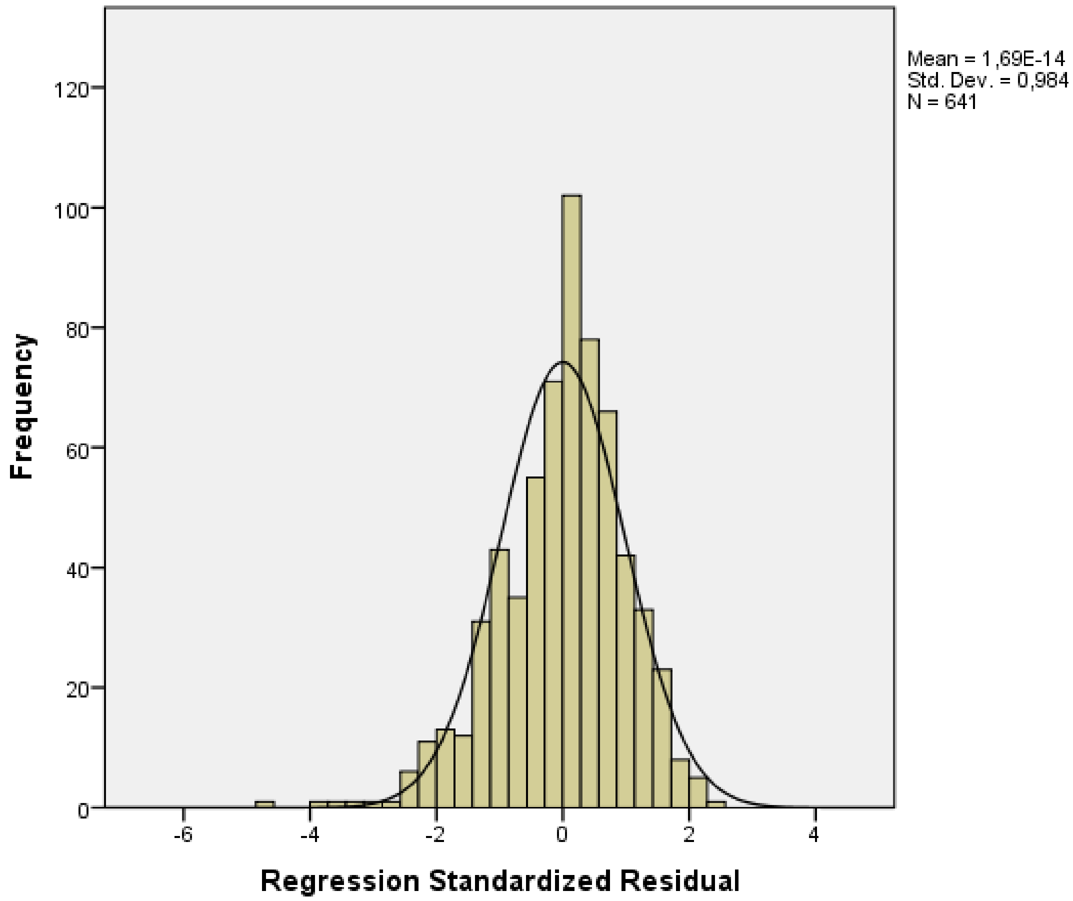

Figure 2. Histogram of standardized residual (x-axis) and the frequency of the dependent variable (score) (y-axis)

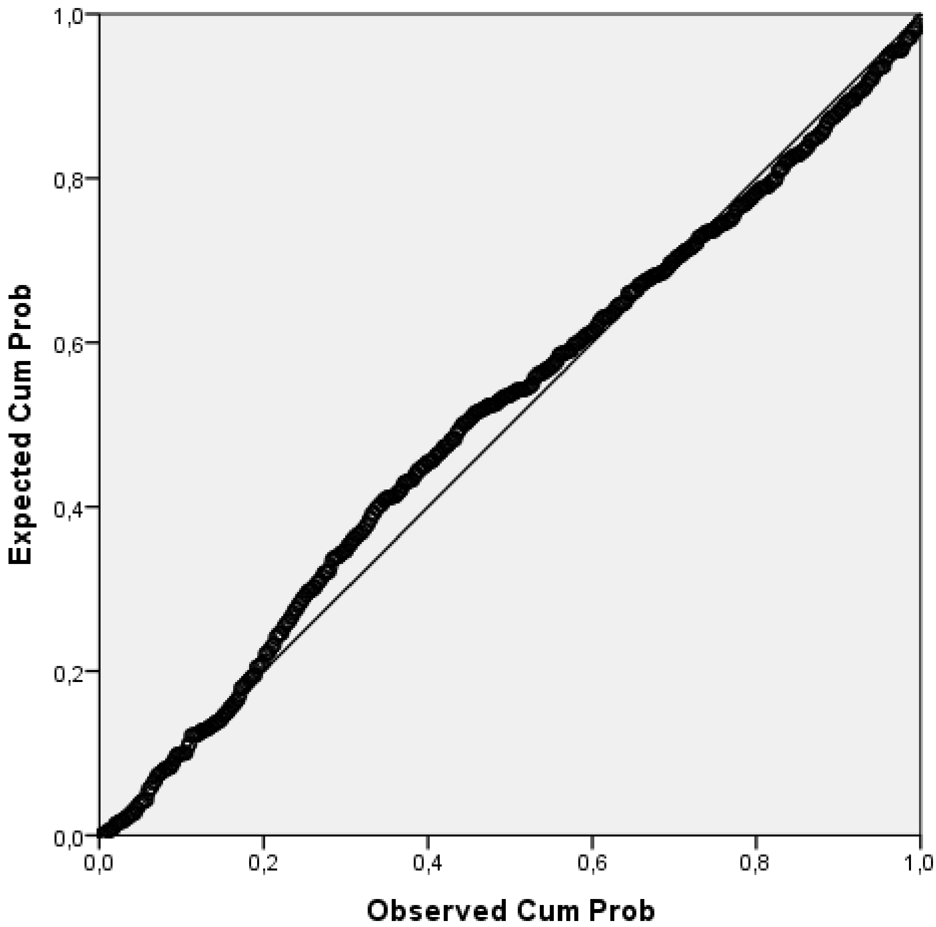

Figure 3. Normal P-P plot of the standardized residual with Expected cumulative probability is on the y-axis and observed cumulative probability on the x-axis.

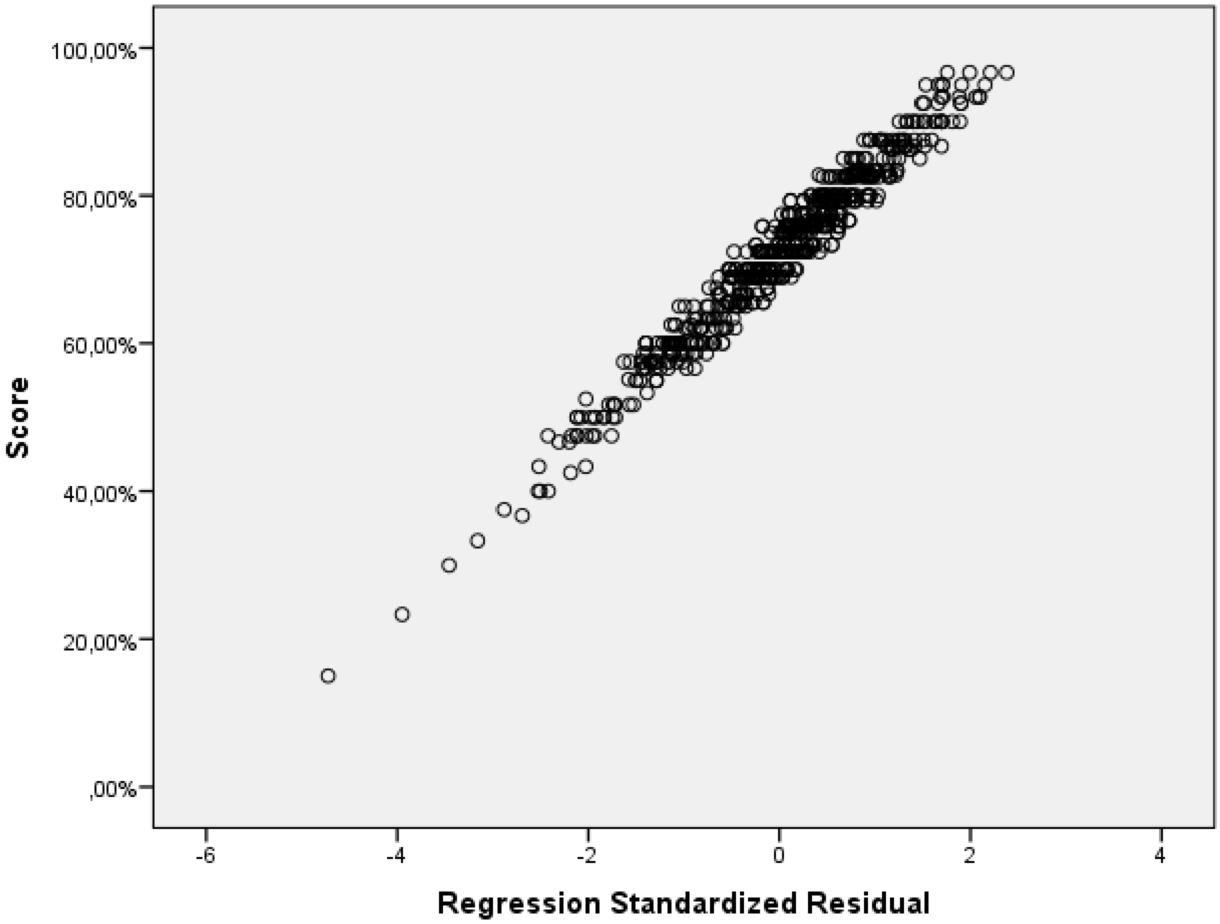

Figure 4. Scatterplot of the dependent variable on the y-axis and the standardized residual on the x-axis.