

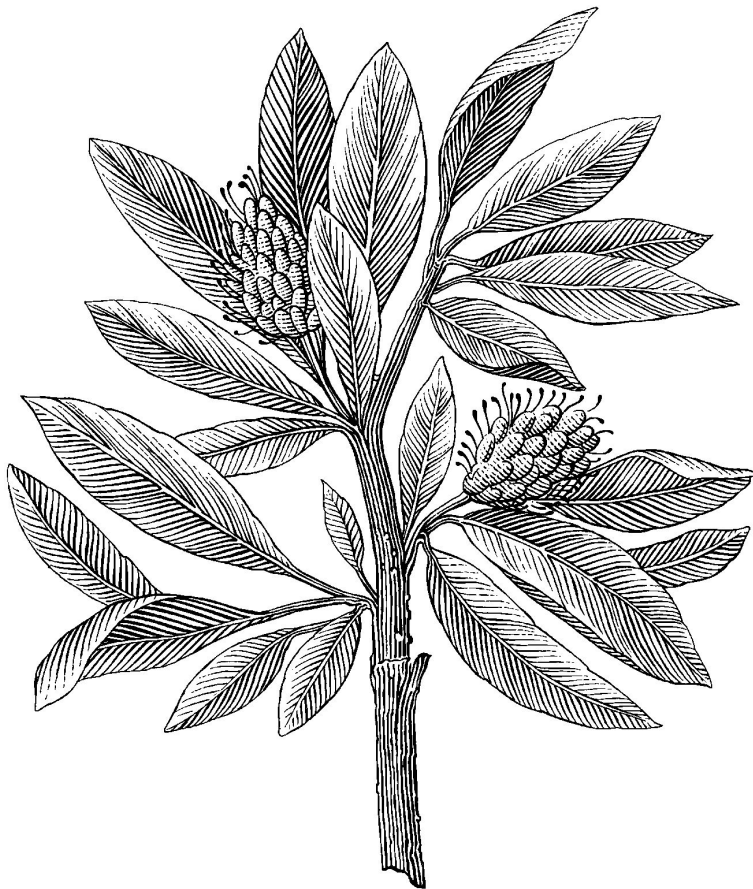


Linnéuniversitetet

Bachelor essay

Discrimination against foreign named students

- a study based on a Swedish University



**Authors: Jasmin Huskanovic
Bellal Adem Nur
Supervisor: Joakim Jansson
Examiner: Mats Hammarstedt
Date: Spring 2020
Subject: Economics
Level: Bachelor
Course code: 2NA11E/2NA12E**

Abstract

In this paper we investigate how anonymous exams impact the grades for foreign-named students and if there are differences in this treatment effect against them when considering the gender of the graders. By obtaining data from Stockholm University, we used a difference-in-difference model to estimate our results. Our results show that there was no negative treatment effect against foreign-named students when considering their test scores prior and after the reform of writing exams anonymously. The results are interpreted through the fact that foreign-named students received even lower test scores after the reform, meaning that the foreign-named students did not gain from the reform. Moreover it can be said that these results can be related to “reverse-bias”, which indicates that the grader is more lenient towards foreign-named students.

Further, when estimating the relationship between the gender of the teacher and the treatment effect, our results showed evidence of female graders treating foreign-named students more negatively compared to male graders. This, due to the results which implied that male graders showed a strong positive treatment effect against foreign-named students, but hence a negative treatment against Swedish named students. Moreover the female graders showed a positive treatment effect against Swedish-named students and the contrary against foreign-named students.

Keywords

Discrimination, Education, Sweden, Gender, Anonymous grading

Acknowledgments

First of all, we would like to address our gratitude to our supervisor Joakim Jansson for his guidance, support and for providing us with data.

We would also like to express our gratefulness to Mats Hammarstedt for his helpful and constructive comments which contributed to the improvements of our bachelor essay.

Finally, we appreciate the helpful advices we received from our discussants, Nelly Sand and Sofia Fritzson.

Table of contents

Abstract	2
1. Introduction	5
1.2 Immigration in Sweden	6
1.3 Purpose of our study	7
1.4 The effects of immigration	9
1.5 Structure of the essay	10
2. Theory	10
3. Material and methods	12
3.1 The introduction of the reform in 2009	12
3.2 Design of the data	12
3.3 The introductory Macroeconomics sample and the grading system	13
3.4 Design of the model	15
3.4.1 Design of our model	16
4. Results & Analysis	18
4.1 The treatment effect of foreign-named students	18
4.2 The relationship between the graders gender and the treatment effect	20
5. Conclusion	25
6. Reference list	27

1.Introduction

Discrimination is an interesting topic when considering education and especially on a university level. Most universities have traditionally diversified environment among the students where they have different foreign backgrounds and also that it occurs a lot of exchange of students between countries.

The aim with our study is to estimate the treatment effects of foreign named students when being graded both anonymously and non-anonymously compared to Swedish named students. Further we want to investigate if the treatment differs between the students test scores from the exams when the gender of the assigned professor is taken in consideration.

To be able to fulfill our aims we used data from Stockholm University where a reform was introduced by the principles of Stockholms university in 2009.

The reform consisted of a decision of removing the identity of the individuals that were taking the standard exams. The result of the reform meant that further on the exams should be written anonymously, this to assure the students of a more secure safety when the exams are corrected. (Stockholm University, 2010)

We used a difference-in-difference (DID) method due to the fact that this method is useful for us since we can examine how the treatment and control group will be affected by a certain event. By this method we found that there was no negative treatment effect against foreign-named students when considering their test scores, meaning that the foreign-named students did not gain from the reform by the fact that they received even lower test scores. Moreover an explanation can be related to a “reverse-bias”, which indicates that the grader is more lenient towards foreign-named students.

In addition when we looked into the relationship between the graders gender and the students test scores we found that female graders treated foreign-named students more negatively than male graders.

Grades are an important tool for future references when entering the labour market. It can give an indication of educational inequality regarding the results we get. Many studies has shown that discrimination that occurs when grades are set for the students can lead to durable consequences for future earnings and continuous educations.

We can relate this to a study by Diamond and Perssons where they investigate how the long-term consequences of teacher discretion when grading high-stakes tests. The authors found that higher grades will eventually lead to higher educational attainment and earnings benefit. However they conclude that the importance of grades is affecting the human capital with merge of higher efforts will in turn lead to substantial gains for outcomes in schooling and the labour market. (Diamond and Persson, 2016). Therefore the treatment effect of foreign named students is of major importance due to the outcomes affected of grading.

1.2 Immigration in Sweden

Since the introduction of the Swedish Immigration board in the late 1960's, the immigration flow to Sweden has increased drastically for several reasons.

The most reasonable explanation for the beginning of migration to Sweden is due to conflicts and wars in several countries such as Germany, Greece, Italy and Yugoslavia.

(Migrationsverket, 2020)

The population deposition has in Sweden, as in other European countries, had big changes in the last 40-50 years. By the year 1990, Sweden had a migrant population of 788.767 which stood for 9.22% of the total population and in 2015 the total population in Sweden consisted of 16.77% immigrants. (Sweden Immigration Statistics 1960-2020, 2020)

The immigrant groups that we will focus on in our study are the following: Middle Easterns, Yugoslavians, South-Americans, Albanians, Turkish, Greeks and also Finnish. The reason for the choice of the immigrant groups is due to the diversity in geographical locations for the immigrants but also these groups are the largest immigrant flows that settled in Sweden.

Each group had their different reasons for moving to Sweden, the reasons could be due to war, labor market opportunities etc. Middle Eastern immigrants, mainly from Iraq, Iran and Palestine, moved to Sweden due to repressive regimes and war in their countries of origin.

The biggest Middle-Eastern wave came during the Iran-Iraq war between the years 1980-1990. Further, another wave from Iraq immigrated after the US invasion in 2003 which led to many Iraqis moving to Sweden for better living standards. The Yugoslavs is another immigrant group that migrated to Sweden due to a series of military conflicts in Balkan

between the years 1991 and 1999. These years brought a huge immigration wave from the Balkan region during the war with many people arriving from Bosnia and Kosovo-Albanians. Furthermore, many South-Americans came to Sweden for the similar reasons, especially people from Chile. During Augusto Pinochet's dictatorship, during 1973-1990, many Chileans fled to Sweden and as a result, 45.000 immigrants with Chilean background are today Swedish residents. Finnish people on the other hand moved to Sweden for a different reason, Sweden had throughout the 1960s, low labor support and had difficulties in the economy growth whereas Finland had a situation that was reversed. Hence, many Finnish people moved to Sweden for better labor market opportunities. (Sweden and migration | sweden.se, 2020) After the World War II, immigration to Sweden was seen as attractive for many because of the labor market opportunities that was available. This was people mainly from Italy, Greece and Turkey which came through labor market authority recruitments. (Migrationsverket, 2020)

1.3 Purpose of our study

As mentioned the aim for this paper is to investigate the treatment effect of foreign named students in a Swedish university by comparing the test scores before and after the reform that lead to anonymously written exams.

The treatment effect can be referred to as if foreign named students have been discriminated against when being graded.

Our study is closely related to "Discrimination against students with foreign backgrounds: evidence from grading in Swedish public high schools" by Tyrefors et al. The authors tested for discrimination against students with foreign backgrounds in high schools in Sweden. They conducted an analysis based on random samples on national tests which was graded blindly by teachers that have graded national tests before without their identification numbers and non-blindly by their own teachers with the students identification information.

The authors also used a difference-in-difference analysis, as they wrote "The major advantage of such an analysis is that unobserved quality of a test can be controlled for, and since the regrader does not know the students foreign backgrounds". The main regression

results showed signs of discrimination against foreign students by concluding that the discrimination effect was sizeable, students with foreign backgrounds received about 10% lower test scores compared to Swedish students. (Hinnerich, Höglin and Johannesson, 2014) Another study that is similar to ours is one done by (Jansson and Tyrefors, 2018) where they tested for the gender grading bias at university level. In this study the aim was to find how the in-group bias mechanism could explain how the treatment effect was driven by different departments where male teachers constituted as a majority. To analyse this the authors estimated difference-in-difference estimations and the results implied evidence of ‘an overall bias against female students’. The study we will conduct is based on filling the gap by particularly these studies based in Sweden. (Jansson and Tyrefors, 2019)

Firstly we conduct our study on the university level compared to the study done by Tyrefors et al (2015) on the high school level. When referring to the other study done by (Jansson and Tyrefors, 2018) the authors used the same data from Stockholm university but they focused on the gender as mentioned before, and we will focus on the treatment differences between foreign named students and Swedish named students by also using a difference-in-difference model.

Other previous studies that we can relate to is Sprietsma (2013), where this paper aimed to study the effect of the teachers expectations on essay grading in Germany. The author randomly assigned Turkish or German first names to a range of essays, the design of the study was similar to those used when trying to find discrimination effects in hiring decisions on the labour market. In the mentioned methods identical application letters were sent out but they differed in names for the applicants, in a similar way the authors used their essay and sent them out to different teachers. Sprietsma found that Turkish students “obtained significantly worse grades and lower secondary school recommendations”.(Sprietsma, 2012) Similar test has been conducted in Brazil by Botelho et al (2015), the authors aim to investigate “ whether racial discrimination in the form of biased assessment of students is prevalent within Brazilian schools”. The study draws evidence from administrative data for eighth grade students and teachers. Moreover the authors find evidence in blindly-scored tests that blacks were getting lower grades in maths compared to their white classmates. (Botelho, Madeira and Rangel, 2015)

Further, previous research has shown that ethnic minority students have performed worse in school in comparison to ethnic majority students when the teachers are from the same ethnic

majority group. In van Ewijk study “Same work, lower grade? Student ethnicity and teachers’ subjective assessments”, the author investigates if ethnic majority teachers grade minority and majority students any different for the same work they complete.

By using an experiment, where different teachers graded the same 10 essays, Van Ewijk did not obtain any results that showed a direct bias against minority groups. Instead, he found indirect evidence of other explanations. Teachers have lower expectations on minority students and an negative attitude towards them. Therefore, the author suggests that these expectations and attitudes may have an indirect impact on why minorities perform worse. (Van Ewijk, 2011)

1.4 The effects of immigration

The effect of immigration in Sweden has led to discrimination in other markets as well. One example is evidence of immigrants that have been discriminated in the Swedish labour market. Moa Bursell conducted a field experiment on the hiring decision considering foreign named and Swedish named applicants where the applicants had identical CVs and the only thing that differed was the applicant’s names. The result implied discrimination against foreign named applicants, where a Swedish applicant received a callback after relatively 10 applications and a foreign named applicant received in general a callback after 18 applications (Bursell, 2014). An identical study was made by Bertrand and Mullainathan on the American labor market. By manipulating resumes through assigning them African-American or White-sounding names, they found that the later received a 50% more callbacks rate for interviews than African-Americans. The authors conclude that this could be an indication to why African-Americans perform worse in the labor market. (Bertrand and Mullainathan, 2004) Discrimination against immigrants has occurred in the housing market in Sweden as well. A study done by Ali M. Ahmed et al, has also shown results of discrimination against foreign named applicants. The study showed significant result of discrimination against the foreign named applicants on the rental housing market in Sweden.

The authors concluded that a foreign named rental applicants received fewer answers on their applications than a Swedish named applicant. (Ahmed, Andersson and Hammarstedt, 2010) Ahmed and Hammarstedt conducted another study on the housing market through a field experiment on the internet. By using made-up names that were ethnic and non-ethnic, they applied for rental apartments. Their findings showed similar results as the previous study, meaning that discrimination occurs on the housing market. (Ahmed and Hammarstedt, 2008)

1.5 Structure of the essay

This thesis has in total six sections. After the first section we will continue with the second section that provides some economic theories and studies that are relevant and can be connected to our research question. The third section includes our material and which method we used to analyze and obtain our results. Furthermore, the fourth provides our results which are presented in two different categories; treatment of foreign-named students, the relationship between the grader's gender and the treatment effect. This section will also include our discussion for the obtained results relating them to the theories in section two, and also connecting the results with previous studies. In the fifth section we will summarize our results in the thesis and afterwards draw conclusions that are related to our research question. Lastly, after the conclusion, we have another section for our bibliography.

2. Theory

Different theories in related studies suggest that discrimination can be due to teachers' expectations for different racial minority groups. In a study done by Harriet R. Tenenbaum and Ruck where evidence showed that teachers' expectations differed across ethnic groups in America. It was found that Asian American students had the highest expectations and that teachers also had higher expectations on European American students compared to Hispanics and African American students which also was statistically significant. (Tenenbaum and Ruck, 2007) This will not be investigated in our study, even though this theory can be relevant when considering the prejudice that professors may have during the grading process.

Tenenbaum and Rucks theory is closely related to Arrow and Phelps theories about Statistical discrimination.

In economics, several theories has been used to identify discrimination across the labour market. Statistical discrimination introduced by (Arrow 1973; Phelps 1972) is based on the prediction that employers generally think that minorities are less productive than majorities during the hiring process. This can also be expressed in the grading process, as when the professor has the prediction that a group is performing better or worse depending on the performance of previous results from the groups ethnic background. (Arrow, K. J. (1973). (Stoikov, Ashenfelter and Rees, 1975) and Phelps, E. S. (1972). However, this theory is developed through Becker's theory of taste-based discrimination. The theory explains how an individual's discrimination is based on their preference. One group can be chosen over another group upon the “taste” of the individual making the choice and it can be a potential mechanism for the observed effect we contain. (Becker, G. (1957). The professors benefit from choosing an individual/group over another is non-existing, hence we can make the assumption that the choice they make is upon their “taste” for a certain individual/group and could be an explanation for ethnic discrimination.

Another theory that can be related to our question of research is presented in a study done by Behaghel et.al (2015) where the authors evaluate an experimental program in France. The aim was to investigate how French firms hire minority and majority groups when they receive anonymized and non-anonymized applications. The study showed unexpected results where firms were less likely to interview and hire the minority applicants when the applications were received in a anonymous form. However, the theory that can explain these results can be referred to as a ‘reversed bias’ which can be explained as a compensation or a more lenient treatment for unprivileged groups or minorities. This theory is implemented in Behaghel et. al (2015) where the authors found that minority groups with long unemployment spells received a more positive probability of being employed and contrary when the applications were anonymous long unemployment spells were seen as a negative attribute from the employers view, and therefore the applicant had a lower probability of being employed. (Behaghel, Crépon and Le Barbanchon, 2015)

We have further chosen to consider the impact of the professors gender for the chosen course. The reason is to be able to extend our analysis. By using the gender of the professors we will be able to identify if there is any relationship between the gender of the professor and the

outcome of the students grades. We can relate this to Carlsson and Rooths study where they found that male recruiters tend to discriminate against Middle-Eastern applicants compared to female recruiters. However, this study is conducted on the labour market which differs from our field of focus.(Carlsson and Rooth, 2007) Due to the results in Carlsson and Rooths study we expect similar results when considering foreign named students.

3. Material and methods

3.1 The introduction of the reform in 2009

On March 5 2009, the principal of the Stockholms University introduced a reform considering the exams. The reform consisted of a decision of removing the identity of the individuals that were taking the standard exams. After the reform was tested during the fall of 2009 the university pronounced to continuously examine with anonymous grading on standard exams. Unfortunately during the implementation of the reform some problems occurred considering the IT systems, meaning that the identity for some students was revealed (Stockholm University, 2010). In our study we will not observe which students had been affected of this problem so we will not be able to control for this drawback.

3.2 Design of the data

We will use the same data as Jansson and Tyrefors (2019) use in their study. The focus will be solely on one course "The introductory macroeconomics course" which will be explained more in detail in the next section.

To fulfill the aim to measure the full effects of the treatment on foreign named students and how the gender of the professor is related to the treatment we would need a set of variables. The set of variables we will obtain from the data set is the points on each question the students had answered on the exam, the ethnicity, the students age and the professors gender. To be able to categorize our observations containing foreign named students and Swedish named students we collected a set of names. The set consisted of 2391 first names which were scattered in different regions such as Middle-eastern, Yugoslavian, South-American, Albanian, Turkish, Greek, Swedish names and also we obtained a list of surnames for Finnish observations due to the fact that their first names are similar to Swedish.

The names were collected through different internet based¹ data on the most common names for the different regions and countries and also through our own knowledge considering the fact that we have roots from Yugoslavia and Africa. Our main reason for obtaining these names in the chosen countries is due to the huge immigration waves that occurred in Sweden. With the chosen names we will have a large probability of collecting matches that our supervisor has in his data set from Stockholm University. The next step was to run the names through the data of the introductory Macroeconomic sample with the help of our supervisor, by doing this we were able to match the names and obtaining our observations and the rest of our variables needed.

3.3 The introductory Macroeconomics sample and the grading system

Our sample that we will use is from the introductory Macroeconomics exams. Through our supervisor Joakim Jansson, we gathered a more detailed information from one particular examination by hand. From spring 2008 to fall 2015, we obtain data from the Macroeconomics course at Stockholm University. The student output data was obtained from the course supervisor and the course coordinator. The key advantage of the introductory exam is that it consists of two multiple-choice questions and seven essay questions. We are going to observe the essay questions and their test scores. Every essay question was worth ten points before the fall term of 2013 and twelve points afterwards.

In our observations, we attained the average score for each exam in the scale 0-12.

The testscore in our summary statistics, as can be seen in Table 1, is on average 5.22 for all exams taken by students. However, to be able for us to make a comparison between the 10- and 12 points questions, we standardized the points for every set of questions.

By standardizing the test scores for the two different set ups of points we gathered the standard deviations for each test score to be able to obtain comparable variables. The standardized values will show us how far from the mean in standard deviations.

Each of the seven essay questions were revised by different teachers whom had each specific assigned questions by a poll. Hence this creates a non-intentional experiment. By using the course coordinator's forum we were able to collect the first names of the teachers and then typed on a program by hand. Lastly, these sets of information were combined together.

¹ Assessed from (Familyeducation, 2020), (Momjunction, 2020), (Fornamn, 2020), (Studentsoftheworld, 2020),

The main grading system that is relevant for our chosen time interval is the Bologna process within the European Union. The Bologna system had to be introduced at the latest by the fall of 2008, although it was used before that deadline in several departments and courses, and therefore the Bologna system is the only grading system that was in use during our time interval. The grades consists of A to F, where A is the highest grade whereas F and Fx is an indication for failure. Our data collection is from the administrative system Ladok.

For our data we find that a lower majority, almost 33% of the teachers were female regarding pre and post the reform. Since we collected the gender of the teachers and the names of the students (therefore also the gender), we can observe the relationship between the teachers gender and students ethnicity. The questions were handed randomly to the teachers and therefore there was no selection by gender or ethnicity, also by considering the other part that the students didn't either know the teachers gender until after the exam.

Further we can also conclude that the students were on average 23 years old. We also found that the majority of the students were Swedish before and after the reform standing for approximately 85% of the total population in our data set. To obtain the total number of exams that were written after the reform, we set a dummy which obtains a value of 1 if the test was taken after the reform, therefore we can observe that almost 82 % of the exams in our observations were written after the reform.

Table 1: Summary Statistics

Variable	Obs	Mean	Std.Dev.	Min	Max
Female-Teacher	39461	.327	.469	0	1
Test Score	39461	5.227	3.529	0	12
Age-students	39461	23.075	3.713	18	53
Swedish	39461	.85	.357	0	1
Foreign	39461	.15	.357	0	1
After	39461	.824	.38	0	1

3.4 Difference-in-difference method

Our main approach that we will use is the difference-in-difference (DID) method. This method is commonly used in econometrics and is often referred to as “controlled before-and -after study”.

The DID is used to estimate the causal relationship, where it compares the difference in outcome before and after some type of treatment for the affected group (treatment group) and unaffected group (control group). In general, the basic idea of the approach is that it captures the effects of policy interventions and policy changes that do not have an effect for everyone at the same time and same way. However, the main assumption that is very important to assure the legitimacy of the DID model, called parallel trend assumption. This assumption is the most difficult one to achieve. Parallel trend assumption demands in absence of treatment, that the groups difference is constant over time. This can't be tested through statistical test but by visual inspection when many observations is available. Therefore, this assumption is most useful when time-periods are short and a violation to this assumption may indicate that the causal effect is biased.

The DID approach is relatively flexible and very logical, if the assumptions are fulfilled. Hence, it gives an indication of causal effect through the observational data. As it focuses on improvement instead of absolute levels, the comparing groups will start at various levels. Another important aspect of the DID approach is that it reflects improvement due to causes that occurs, instead of just the procedure or intervention that is being examined. There are also some limitations for the DID-model.

To be able to use the DID- method both the baseline data and a non-intervention group must be included. The method is also unable to use when the amount of treatment is dependent on the baseline outcome and also when the pattern for the comparison groups differ. Therefore we study the treatment effect of foreign-named students relative to Swedish-named students when the grading reform is taken in to consideration.

(Difference-in-Difference Estimation | Columbia University Mailman School of Public Health, 2020) (Lechner, 2010) (Bertrand, Duflo and Mullainathan, 2004)

3.4.1 Design of our model

To be able to study the treatment effect of foreign-named students compared to Swedish named student we have created our difference-in-difference model marked with (1) below. With this model we are able to study how our treatment group (foreign-named students) and our control group (Swedish-named students) are being treated before and after the reform took place.

$$(1) \text{Testscore}_{it} = \alpha_0 + \alpha_1 \times FN_i + \alpha_2 \times \text{After}_t + \alpha_3 \times FN_i \times \text{After}_t + \epsilon_i$$

As mentioned we obtained data on the points the different students got on each question on the exam before and after the reform, this is our ‘‘Test Score’’ also referred to as our dependent variable. Our obtained observations will also include if they are foreign named or not.

If the case is that the observation/individual is foreign-named our dummy variable noted with ‘‘FN_{*i*}’’ will obtain a value of 1 and if not the dummy will obtain a value of 0.

Further if the students testscore is from after the reform the dummy noted ‘‘After_{*t*}’’ will get a value of 1 and if the testscore is from before the reform the this dummy will have a value of 0. The coefficients α_0 and α_1 is indicating the difference between our control and treatment group before the reform whereas α_2 will show the change across time for both of the groups due to the fact that this coefficient is interacting with the dummy After_{*t*}. The coefficient α_3 is of major importance due to the fact that this coefficient will show the causal effect for the treatment of foreign named students after the reform or in other words the difference in difference between the groups. Lastly we have our error term noted with ϵ_i indicating the uncertainty in the model.

Therefore the approach is based on first capturing the difference between the groups before the reform, then our second approach is to obtain the difference or the change including the reform for the foreign named students. Lastly we will obtain the difference across the time before and after the reform for the control group as well, when this is set we can obtain the difference-in-difference or the treatment effect for the foreign named students.

To estimate the treatment effect of foreign-named students when considering the graders gender we will limit our sample by female and male graders for the regression by using a dummy variable which will withhold a value of 1 if the grader is a female and 0 if the grader is a male.

By this method our aim is to capture if there is any relationship between the treatment effect of foreign named students and the graders gender.

As mentioned in the section above, our main assumption Parallel Trend has to be fulfilled. The basic idea of the approach is that it captures the effects of policy interventions and policy changes that do not have an effect for everyone at the same time and same way. In our case, for the test to be valid, the difference in test scores for foreign-named students and Swedish named students has to move in parallel before the reform is applied. However, after the reform, we need to estimate if there is any difference on both the treatment group and the control group to observe it in our DID-model.

4. Results & Analysis

4.1 The treatment effect of foreign-named students

In this section, our aim is to present the results of the treatment effect of foreign named students when compared to Swedish named students. Moreover, we will present the results of the treatment effect of foreign-named students when the grader is male or female to see if there are any differences in treatment. We also want to analyse the different results and explain them carefully. In line with the findings that we attained, we suppose that our main assumption parallel trend is fulfilled. In addition it can be noted that when computing our regressions we used robust regressions the reason is to control for heteroskedasticity and to obtain estimates that are not violating our assumptions .

Below we have table 2 which is the regression results of the test scores as our independent variable, and the rest of the coefficients as our dependent variables.

The aim is to obtain the treatment effect of foreign named students.

As observed in the table, row 2 column 2 shows the constant term and it has the value of 0.078 standard deviations. This term explains the test scores for the Swedish named students prior the reform and as mentioned in previous sections our test score was standardized meaning that the test scores were normalized to zero and therefore our first constant term shows that the test scores for Swedish students are 0.078 standard deviations above average. Hence, when comparing to the After variable (*row 3 ,column 2*) which shows the change in standard deviations after the reform for the Swedish students. This can further be

implemented in a way that the Swedish students obtained test scores 0.057 standard deviations lower than prior the reform, around 0.02.

We can therefore conclude that Swedish named students did not benefit from the anonymous examinations due to the lower test score results.

Further, if we examine the foreign-named students we first look at the variable noted Foreign (*row 4, column 2*) where this coefficient is related to the constant term. This coefficient has a value of -0.127, meaning that the test scores for the foreign-named students prior the reform were 0.127 below the average for our sample.

Lastly we have our Treatment effect (*row 5, column 2*) which is our main variable of interest and our α_3 in our model. However, this coefficient obtains a value of -0.103 standard deviations. As mentioned above the Foreign variable shows the test scores for foreign-named students before the reform, and when comparing this with our treatment effect which also is interacted with all the other variables we can conclude that the test scores for the foreign-named decreased with 0.103 standard deviations. Moreover this can be implemented that after the reform foreign-named students received even lower test scores compared to when the graders knew about their identity. Therefore this is showing evidence of a positive treatment effect on foreign named students prior the reform.

Both the Swedish-named students and the foreign named students were obtaining lower test scores after the reform. Furthermore this is indicating that foreign-named students were treated better when the graders knew their names. These surprising results compared to other studies can perhaps be explained by the fact that the Macroeconomic exam was written in Swedish. This is an important aspect because before the reform the graders may have specially treated the foreign-named students and overseen their language and writing ability and therefore graded them in a careful way. From a biased view the foreign-named students probably had a lower language ability compared to native Swedish students.

We can also relate these results to (Behaghel et. al 2015) where they found a reverse-bias effect of the treatment in the hiring process against minority groups in France. This reverse-bias can be introduced as being a sufficient explanation for our results where the graders had a more lenient treatment against the foreign named students due to a prejudice of them being less privileged than Swedish named students.

Further by looking at our table we can conclude that all our coefficients are statistically significant on the 95% confidence level.

Table 2: Treatment effect of foreign-named students

VARIABLES	(1) Standardized values of (Testscore)
Constant	0.078 *** (0.014)
After	-0.057*** (0.015)
Foreign	-0.127*** (0.034)
Treatment Effect	-0.103*** (0.037)
Observations	39,461
R-squared	0.006

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

4.2 The relationship between the graders gender and the treatment effect

In table 3, we estimated the treatment effect of foreign named compared to swedish named students but now instead only if the graders were males. The outline here is that we used the same regression as before with the same variables, but now we limit our sample by female and male graders for the respective table.

By interpreting the table 3, we can see that before the reform with the constant coefficient (*row 2, column 2*) 0.028, male teachers were grading Swedish named students more positively after the reform where the After coefficient (*row 3, column 2*) increased with 0.017 standard deviations. This further indicates that male graders gave better testscore results after the reform to swedish named students.

However, looking at the results for the foreign named students we can observe that before the reform the coefficient noted foreign (*row 4, column 2*) was -0.101 standard deviations below the testscore average whereas after the reform with interaction of all the variables they

received -0.143 standard deviations lower test scores. Hence, we can conclude that male graders were treating foreign-named students worse after the reform which is an indication of a positive treatment effect when knowing the identity of foreign students.

Simultaneously we can conclude that the Swedish named students were treated better after the reform and therefore we can say that the treatment effect against swedish-named students was negative.

Worth highlighting is that all the coefficients were statistically significant on the 95% confidence level.

Table 3: Treatment effect of foreign-named students when the grader is a male

VARIABLES	Standardized values of (Testscore)
Constant	0.028* (0.016)
After	0.017 (0.017)
Foreign	-0.101*** (0.038)
Treatment Effect	-0.143*** (0.043)
Observations	26,575
R-squared	0.006

Standard errors in parentheses
 *** p<0.01, ** p<0.05, * p<0.1

Table 4 shows the same regression as the last table but now instead the graders are females, it can also be noted that one third of the exams were graded by females.

When estimating the regression results we denoted the dummy variable with the value of 1 indicating that only tests graded by females were taken in consideration.

The results we achieved in table 4 differ substantially from the results in table 3. By looking at the table below, we find that before the reform for Swedish students, the test scores were 0.211 (*row 2, column 2*) standard deviations above the average sample test score.

The estimation after the reform (*row 3, column 2*) showed a decrease in standard deviations where we attained a negative value of -0.236. Therefore we can express that Swedish named students were treated significantly worse after reform which differs from substantially from the results that they attained when there was a male grading them.

Foreign named students received similar results as in table 3 where they obtained -0.193 (*row 4, column 2*) standard deviations below the average test score. Our findings show that female teachers were treating foreign named students negatively before the tests were taken anonymously. Furthermore after the reform, we found in our estimation that the test scores decreased with 0.009 standard deviations for the foreign-named students. It can also be noted that the coefficients "Constant" and "After" in this regression were statistically significant on the 95% confidence level.

To conclude table 4 it can be noted that the Swedish named students were treated positively before the reform by the female graders due to the decrease in the test scores after the reform. The foreign-named students on the other hand were treated negatively before the reform and slightly worse after the reform (-0.009).

By comparing the result in table 3 and 4 considering the gender of the graders we can in summation say that the female graders treated foreign named students more negatively than compared to the male graders. This is due to the fundamental decrease in the test scores after the reform, when the male graders were correcting the exams we saw a decrease of -0.143 standard deviations compared to a -0.009 standard deviation decrease when a female was grading. This is an indication of a strong positive treatment effect against foreign-named students when a male is grading.

Simultaneously we can say that the female graders treated the Swedish named students more positively when grading compared to the male graders. This result is due to the higher test scores above the average for the Swedish named students before the reform, and when comparing these test scores with the scores after the reform we obtain that male graders graded Swedish-named students even with better scores after the reform which is an

indication on a negative treatment effect on Swedish students. When female graders on the other hand were grading the Swedish students the students obtained worse test scores after the reform which instead is an indication of a positive treatment effect against Swedish named students when a female is grading. With these results considering the graders gender we can with in summation say that the treatment effect against foreign named students was more negative when the grader was a female. By some computations it can also be noted that male and female teachers grade foreign-named students more similar after the reform, since the coefficients have moved closer to each other.

These results however show the opposite compared to similar studies conducted on i.e the labour market. We can recall to Carlsson and Rooths study where they found that male recruiters tend to discriminate against Middle-Eastern applicants compared to female recruiters (Carlsson and Rooth, 2007).

Other explanations can be related to Jansson and Tyrefors study “*Gender Grading Bias at Stockholm University: Quasi-Experimental Evidence From an Anonymous Grading Reform 2018*” where the authors use a mechanism called “in-group bias”.

This in-group bias can be explained by a favoritism, basically if the graders characteristics are similar to the students characteristics. These different characteristics can affect the outcome of the test scores, i.e if the graders has the same gender or ethnicity as the students then the test scores can be biased. The in-group bias however, cannot be affiliated with our study since we have not obtained the ethnicity of the teachers or the gender of the students. Other explanations for the results that we obtained are in line with the explanations for table 2.

Table 4: Treatment effect of foreign-named students when the grader is a female

VARIABLES	(1) Standardized values of (Testscore)
Constant	0.211*** (0.026)
After	-0.236*** (0.028)
Foreign	-0.193*** (0.062)
Treatment Effect	-0.009 (0.067)
Observations	12,886
R-squared	0.013

Robust standard errors in parentheses

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

5. Conclusion

As mentioned in our introduction section, there are a range of studies that investigate treatment effects against minority groups in different sectors. We focused on the schooling sector since grades are of major importance for future references and studies but also when entering the labour market. As Diamond and Persson showed that there are long-term consequences when considering earnings and the human capital.

By using a difference-in-difference method on our obtained data from introductory macroeconomic sample, we are able to conclude that our results did not show any negative treatment effect against foreign-named students. These results are implemented through the fact that foreign-named students received even lower test scores after the reform, thus we can say that the graders treated the foreign-named students positively. When considering these results we can conclude that they show the opposite from previous similar studies conducted by Hinnerich, Höglin and Johannesson (2014), Sprietsma (2013), Botelho et al (2015) and (Van Ewijk, 2011). As mentioned in our introduction, discrimination is an interesting topic especially on a university level. Most universities have a diversified environment where students have different foreign backgrounds and also that it occurs a lot of exchange of students between countries. Furthermore we can conclude that the results deviate from the theory about Statistical discrimination and Becker's theory about Taste-based discrimination due to the lack of negative treatment effects on foreign-named students. However an explanation could instead be due to the theory about "reversed-bias" by Behaghel et.al (2015).

The positive treatment against foreign-named students could be due to a more compensating and a more lenient treatment for groups or minorities considered to be less privileged.

This "reversed-bias" can also be related to the fact that the macroeconomic exam was written in Swedish, thus this could imply that the graders overlooked the language and writing abilities and graded the foreign students in a more compliant way.

In addition our second aim for this study was to investigate if there is a relationship between the graders gender and the treatment effect of the students.

By extending our regression with a dummy for the graders gender, we conducted results showing that female graders treated foreign-named students more negatively compared to

male graders. The male graders showed results of a strong positive treatment effect against foreign-named students, but hence a negative treatment against Swedish named students. Moreover the female graders showed a positive treatment effect against Swedish-named students, and due to a slight decrease in test scores for foreign named students after the reform and fundamentally low test scores prior the reform therefore we came to our conclusion.

These results however show the opposite compared to the study by Carlsson and Rooth (2007) where they found that male recruiters on the Swedish labour market tended to discriminate foreign applicants compared to female recruiters.

Our explanations for these results can be related to the mechanism “in-group bias” where favoritism due to similar characteristics between the grader and the students could have had an impact. Thus this mechanism is beyond the scope of our paper, and we can therefore not conclusively say what factors contributed to these results considering the graders gender but this mechanism could be interesting to test in future studies.

In the essence of our research questions we can conclude that there was no negative treatment effect against foreign-named students, and that female graders showed a more negative treatment effect against foreign named students compared to the male graders.

6. Reference list

Internet

FamilyEducation (2020). *Finnish Last Names*.

Available at:

<https://www.familyeducation.com/baby-names/browse-origin/surname/finnish> [Accessed 12 April 2020].

Fornamn (2020). *Namn till flickor och pojkar – 40 000 st förnamn med ursprung och betydelse* Available at: <https://www.fornamn.se/> [Accessed 10 April 2020]

Macrotrends.net. (2020). *Sweden Immigration Statistics 1960-2020*. Available at:

<https://www.macrotrends.net/countries/SWE/sweden/immigration-statistics>
[Accessed 8 April 2020].

Columbia Public Health (2020). *Difference-In-Difference Estimation | Columbia University Mailman School Of Public Health*. Available at:

<https://www.mailman.columbia.edu/research/population-health-methods/difference-difference-estimation> [Accessed 10 April 2020].

Migrationsverket.se. (2020). *History*. Available at:

<https://www.migrationsverket.se/English/About-the-Migration-Agency/Migration-to-Sweden/History.html>
[Accessed 8 April 2020].

Momjunction (2020). *'100 Albanian Baby Names For Girls And Boys'* Available at:

https://www.momjunction.com/articles/albanian-baby-girl-names_00352752/?fbclid=IwAR3APc8KTKrPVii5XQuOw_lxwR8b7Nvsl0nwDk7r9IFHRID0wV-tH2ceOo
[Accessed 10 April 2020]

Stockholm University. (2010). *Införande av anonyma tentamina vid Stockholms Universitet*.

Available at:

http://www.su.se/polopoly_fs/1.26344.1320939789!/Beslut_om_anonyma_tentamina_vid_Stockholms_universitet_Dnr_SU_459_2690_08.pdf. [Accessed 3 April 2020]

Studentsoftheworld (2020). *Spanish names - Top 100 first names in Spain & statistics*

Available at:

<http://www.studentsoftheworld.info/penpals/stats.php?Pays=ESP&fbclid=IwAR3bi05reQ-pLsTy0EB5mcz89MSes9dw1MswuWMAU3-4glMH75hFkXcyo-g> [Accessed 10 April 2020]

Sweden.se. (2020). *Sweden And Migration* Available at: <https://sweden.se/migration/#2000> [Accessed 7 May 2020].

Literature

Ahmed, A., Andersson, L. and Hammarstedt, M., 2010. *Can Discrimination in the Housing Market Be Reduced by Increasing the Information about the Applicants?. Land Economics*, 86(1), pp.79-90.

Ahmed, A. and Hammarstedt, M., 2008. *Discrimination in the rental housing market: A field experiment on the Internet. Journal of Urban Economics*, 64(2), pp.362-372.

Becker, G. (1957). *(The economics of discrimination (2nd ed.)*. Chicago: University of Chicago Press

Behaghel, L., Crépon, B. and Le Barbanchon, T., 2015. *Unintended Effects of Anonymous Résumés. American Economic Journal: Applied Economics*, 7(3), pp.1-27.

Bertrand, M., Duflo, E. and Mullainathan, S., 2004. *How Much Should We Trust Differences-In-Differences Estimates?. The Quarterly Journal of Economics*, 119(1), pp.249-275.

Bertrand, M. and Mullainathan, S., 2004. *Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination*. *American Economic Review*, 94(4), pp.991-1013.

Botelho, F., Madeira, R. and Rangel, M., 2015. *Racial Discrimination in Grading: Evidence from Brazil*. *American Economic Journal: Applied Economics*, 7(4), pp.37-52.

Bursell, M., 2014. *The Multiple Burdens of Foreign-Named Men--Evidence from a Field Experiment on Gendered Ethnic Hiring Discrimination in Sweden*. *European Sociological Review*, 30(3), pp.399-409.

Carlsson, M. and Rooth, D., 2007. *Evidence of ethnic discrimination in the Swedish labor market using experimental data*. *Labour Economics*, 14(4), pp.716-729.

Diamond, R. and Persson, P., 2016. *The long-term consequences of teacher discretion in grading of high-stakes tests*. *National Bureau of economic research, Nber working paper series*.

Hinnerich, B., Höglin, E. and Johannesson, M., 2014. *Discrimination against students with foreign backgrounds: evidence from grading in Swedish public high schools*. *Education Economics*, 23(6), pp.660-676.

Jansson, J. and Tyrefors, B., 2019. *Gender Grading Bias at Stockholm University: Quasi-Experimental Evidence From an Anonymous Grading Reform*. *SSRN Electronic Journal*,.

Lechner, M., 2010. *The Estimation of Causal Effects by Difference-in-Difference Methods Estimation of Spatial Panels*. *Foundations and Trends® in Econometrics*, 4(3), pp.165-224.

Phelps, E. S. (1972). *The statistical theory of racism and sexism*. *The American Economic Review*, 62(4), 659– 661)

Sprietsma, M., 2012. Discrimination in grading: experimental evidence from primary school teachers. Empirical Economics, 45(1), pp.523-538.

Stoikov, V., Ashenfelter, O. and Rees, A., 1975. Discrimination in Labor Markets. Industrial and Labor Relations Review, 29(1), p.140.

Tenenbaum, H. and Ruck, M., 2007. Are teachers' expectations different for racial minority than for European American students? A meta-analysis. Journal of Educational Psychology, 99(2), pp.253-273.

van Ewijk, R., 2011. Same work, lower grade? Student ethnicity and teachers' subjective assessments. Economics of Education Review, 30(5), pp.1045-1058.