Faculty of *Business, Economics & Law,* School of Economics

THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

# The good, the bad and the ugly of high school: measuring superstar effects in public schools in Greece

by
*Raul Mathias Leon*

Supervisor: Dr. Rigissa Megalokonomou

28 October 2022

An Honours Thesis submitted to the School of Economics, The University of Queensland, in partial fulfillment of the requirements for the degree of Bachelor of Economics (Honours)

*Approximate word count: 20 000*

# Acknowledgments

First and foremost, I would like to express my sincere gratitude to my supervisor Dr. Rigissa Megalokonomou for her support, guidance, and patience throughout the entire Honours program.

To my parents, thank you for your unending and unconditional love and support. It is what has kept me strong and motivated throughout my live.

Finally, I would like to dedicate this thesis to my grandfather, Raúl León Barúa, who died one month ago without having read this work. He was not only my role model and beloved family member but an avid researcher whose legacy I wish to preserve.

# Declaration

I declare that the work presented in this Honours thesis is, to the best of my knowledge and belief, original and my own work, except as acknowledged in the text, and that material has not been submitted, either in whole or in part, for a degree at this or any other university.

Raúl M León P

Raul Mathias Leon

# Abstract

The literature has shown that female high-achieving students mostly create positive externalities in a classroom setting. Will these peer effects change with the use of stronger definitions to classify students with extremely high performance as superstars? I study the impact of being exposed to superstars in early high school on the remaining students' scholastic outcomes and university choices. To do so, I exploit the quasi-random assignment of students and teachers to classrooms in Greece. I find negative effects on remaining students from being exposed to superstars, possibly suggesting that remaining students—including other high achievers— may reduce effort. Using novel data from a representative sample of schools in Greece, I find that students exposed to superstars' test scores in geometry decreased by 0.031 standard deviations in their end-of-the-year exam. Their test scores in modern Greek dropped by 0.047 standard deviations. Interestingly, these superstar effects are mostly driven by the effects that female superstars have on other female students. I also find evidence for a perverse effect of heterogenous abilities in effort by showing that the superstar effects are substantially smaller for the second best, top 3 and top 5 students of the classroom. Finally, I find that male and female STEM superstars do not influence the beliefs of women STEM related gender roles and preferences. I find that exposure to female non-STEM superstars increases a female student probability of being admitted to a STEM university program as well as choosing a STEM track choice in high school by approximately 0.02 and 0.05 percentage points, respectively.

# Table of Contents

## Chapter 4: Empirical Framework



## Chapter 5: Empirical Results

# Chapter 6: Conclusion

# Lists of figures and tables

# Chapter 1

# Introduction

## 1.1 Contribution

The first study that found evidence that peers and parents can have a bigger impact than teachers or school characteristics on educational attainment is the Coleman Report (Coleman et al., 1996). Subsequent studies have found similar evidence and have made this view the conventional wisdom in the economics of education (Dearden et al., 2002; Carrell et al., 2009).

Studies have later been concentrated around the interactions between genders in the classroom (Lavy & Schlosser, 2011; Goulas et al., 2018). While others have tried to enrich the literature by finding some evidence that peer effects are more prevalent inside racial groups (Hoxby, C., 2000). Recently, the literature has tried study the impact of the interaction of the class composition and high achievers on why women tend to stop studying STEM degrees (Fischer, S., 2017) or classroom peer effects of high and low ability individuals (Lavy et., 2008; Busso & Frisancho, 2021).

On the other hand, Brown (2011) found that there is “superstar effect” that reduces effort level in players when the ability gap is large in golf tournaments (perverse effect of heterogenous abilities in effort). Obviously, the sign and channels the “superstar effect” does its magic could substantially differ in a classroom environment. Mainly, since it is not just a place of competition but of cooperation and gender roles formation. I will try to identify if the perverse effect of heterogenous abilities in effort occurs in a classroom environment by checking the superstar effect on the second-best student in the classroom and its impact on the top 3 and 5 students as a whole.

Finally, I will follow this tradition and expand the peer effects literature in a school environment by linking superstar, gender and STEM scholastic outcomes. For this, I will classify high achieving students into female (supergirl), male (superboy), STEM and non-STEM superstars to study the presence of a type of superstar on the rest of the classrooms short- and long-term scholastic outcomes.

## 1.2 Motivation

Many policies used by institutions in the education sector have the ability to drastically change the compositions of cohorts (Zimmerman, 2003). For example, affirmative action can increase racial diversity while selective admissions can make the cohort more intellectually diverse. These policies, as well as the status quo, can have an impact on the students' educational achievement, choice of study and matriculation rates. Conceptually, there could exist pareto-efficient allocation that depend on the extent of peer-group externalities and complementarity of students.

As mentioned above, the Coleman Report showed that peers and parents can have a bigger impact than teachers or school characteristics on educational attainment. Some, most famously Card and Krueger (1992), have tried to challenge this conventional wisdom by finding results that school quality, measured through basic standards (i.e. pupil/teacher ratio, average term length, and relative teacher pay), does affect the rate of return to education. Whereas Carrell, Sacerdote & West (2013) have used an affirmative argument by which they accept the predominant role of peer effects and proceeded to estimate them to construct a treatment group of classes with the objective of bettering the academic achievements of the students with the lowest ability. Nevertheless, they obtained adverse results due to the endogenous responses in peer group formation. The endogenous quality of the subject of study appears to make the effect of policy changes very difficult to predict, adding skepticism in the application of fine-tuned peer effects policies. Thus, restricting the application of our results as a policy to better education.

In a sense, the current policy debate revolves around the Coleman Report vs. Card & Krueger (1992) (Dearden et al., 2002). If the latter are right the problem of education could be fixed by increasing expenditure. But if they are wrong, the expenditure could go to waste or at the very least be inefficient if peer effects are not considered.

Furthermore, there has been extensive research as well as policy directives done on woman's underrepresentation in STEM. Nonetheless, there hasn't been much change in gender inequalities in STEM as measured by baccalaureates awarded in recent decades. The picture gets worse when considering academic positions (Casad et al., 2021). In the last years, new dynamics in test scores across developed countries indicate that sex differences are not purely biological. Illustrated by the fact that the gap between female and male math ability is narrowing and has even reversed in some countries. Painting the argument in favor of the existence of strong environmental influences (Ceci et al., 2014).

Lastly, teachers' gender role attitudes seem to influence the gender gap in a variety of outcomes (Lavy & Megalokonomou, 2019). Influenced by this paper, this thesis will try to explore how being exposed to a (fe)male STEM or non-STEM superstar, instead of a teachers' gender role attitudes, can change environmental influences inside the classroom i.e. having an effect on the student's formation of gender roles. A plausible hypothesis could consist that female students exposed to a male STEM superstar will be discouraged from pursuing further studies in STEM. Intuitively, the channel might consist of women comparing themselves with the superstar and conclude that they are not apt or reinforce the idea that STEM is male dominated area of study. Which is very plausible since Murphy & Weinhardt (2020) found that females are more negatively affected by having a low rank. Or, a female student could observe a female STEM superstar and generate a role model belief that women can also achieve success in STEM. I will try to find some initial evidence for this type of questions.

## 1.3 Outline

The rest of this paper is organized as follows. Chapter 2 reviews the existing literature of peer effects in the educational setting, as well as providing some limitations and common methodologies usually used in applied work. Chapter 3 covers the institutional setting and the dataset used in this empirical model. In Chapter 4, the identification strategy and the econometric specifications are discussed in detail. Chapter 5 presents the initial estimation results as well as a variety of heterogeneity and robustness exercises. To conclude, there is an analysis of the study's findings and its implication for policy and future research in Chapter 6.

# Chapter 2

# Literature Review

## 2.1 Introduction

As discussed in Chapter 1, the gap between female and male math ability is narrowing and has even reversed in some countries. Nevertheless, woman participation in STEM degrees at the university level has not had a similar trajectory. In addition, there could potentially exist a considerable superstar effect in classrooms and this could affect not only short-term high school performance, but the process on how the students gender roles and preferences are constructed over the years bringing about undesired STEM long-term outcomes. For example, the crowding hypothesis (Bergman B., 1974) establishes that the concentration of women in non-STEM jobs could be generating an excess supply which reduces salaries and intensifies the wage gap between men and women. This chapter will review existing literature and will present the limitations to our results as well as opportunities for further research.

## 2.2 Why do girls and boys make different STEM related choices?

A study found that women in a high achieving chemistry introductory classes are less likely to finish the degree, while men are unaffected by it (Fischer, S., 2017). This an example of perverse effect of heterogenous abilities in effort in the education sector. Initially, this idea of perverse effect of heterogenous abilities in effort is taken from Brown (2011). The emblematic paper proposes a simple economic model in which the benefits of internal competition vary according to workers' relative abilities. In simple, the model indicates that a large difference in ability reduces effort levels in the rest of competitors.

Additionally, Brown (2011) provided some empirical evidence to the model by studying a panel data from professional golfers that consisted of 363 PGA tournaments during the years 1999-2016. His findings suggest that there exists a substantial superstar effect on the performance of the other players (Tiger Woods' being the superstar). These results were robust to streak and slump indicators. Specifically, a compelling pattern in the data was presented that allowed for the superstar effect to increase when the superstar was playing above average, and the effect was reduced when the superstar was playing worse than average.

In any account, the current study differences from the previous work (Fischer, S., 2017; Cools et al., 2019) since it will try to answer: why do women are more likely to select a non-STEM field of study? Contrary to: why they stop studying STEM at a university level? Thus, considering the superstar effect in the classroom environment and checking if it is influencing short as well as long-term preferences and gender roles seems to be a sensible thing to do.

As mentioned before, there exists some evidence that a higher proportion of females in a neighborhood or classroom improves not only the academic achievement but the university matriculation rates and can impact their choice of study across both genders (Goulas et al., 2018). Thus, the hypothesis that female superstars could potentially affect STEM preferences in the long-run is a plausible hypothesis. The paper will also try to differentiate between standard, female STEM or non-STEM superstars as to check which one has a negative or positive effect on STEM application and enrollment at the university level.

Another important channel of peer effect is the rank concern. Murphy & Weinhardt (2020) found that men are more positively affected by having a better rank. In contrast, females are more negatively affected by having a low rank. Thus, rank is an important factor for the education production function. Rank position within primary school has significant effects on secondary school achievement and the likelihood of completing STEM subjects. As a result, it would be natural to expect that the superstar effects could be potentially be positive or insignificant in boys and negative in girls. There even exists some literature oriented to develop a theoretical model where students care about rank and not just of learning (Tincani, M., 2017).

Furthermore, there is also evidence that absolute STEM advantage (defined as the ratio of average performance in STEM and non-STEM courses) and comparative STEM advantage (constructed by using within-classroom rank of students' absolute STEM advantage) have different implications (Goulas et al., 2020). They found that comparative STEM advantage has long term effects, especially for women. To make the result more concrete, if a student is assigned to a classroom that increases the student's comparative STEM advantage then he is more likely to apply for a STEM degree at the university. Thus, rank peer effects seem to be a potential, if not the most important factor contributing to the STEM gap. This is more evidence that relative ability to others does have some implications on preferences and gender roles formation where the superstar effect could be playing an important role.

## 2.3 High achieving students peer effects

There has been a recent study that found asymmetric effects of high-performing male and female students on their corresponding peers' academic achievement separately by gender (Busso & Frisancho, 2021). Specifically, they find that the effect of high performing students is positive and considerable higher on same-gender students, especially within females. Furthermore, they find evidence that high performing female students influence male students while high performing males do not have an impact on their opposite-gender peers.

Another paper has found that the exposure to high achieving student, independently of gender, negatively affects the preferences for schools with a higher admission standard (Modena et al., 2022). They report the fact that high achieving boys create a negative externality on other high performing girls in the classroom by reducing admission scores and affecting their preferences for school with higher admission standards. Once again, the literature identifies that peer effects are very sensible to gender.

Moreover, there exists some interesting effects of exposure to a female and male high achiever on long-run scholastic outcomes (Cools et al., 2019). Interestingly, they found that a female that is exposed to a male high achiever is less likely to complete a bachelor degree and that her performance in STEM courses is substantially reduced. In the long-term they find that the male high achiever effect has implications for an increase in fertility and a decrease in labor force participation. These effects are of a bigger magnitude for low achieving students. Conversely, Cools et al. (2019) found that being exposed to a female high achiever increased the likeliness that a female student completes a bachelor degree. There were no peer effects of high achievers on males, even after controlling separately for gender.

Another interesting study has estimated peer effects in the classroom by defining high and low ability individuals as those being one year ahead of their own cohort and those that fell behind, skippers and repeaters, respectively (Lavy et al., 2008). A higher proportion of skippers don't affect the academic outcome in the middle of the distribution of regular students. Nonetheless, they have a positive impact in the upper distribution of the regular students. On the contrary, a higher number of repeaters inside the classroom has a negative effect, predominantly in the lower part of the distribution. The study will try to find other relevant and especially stronger definitions of "high performing students" and do heterogeneity checks across definitions to see if the superstar effects vary in the upper distribution.

Lastly, Mouganie & Wang (2020) found that having high-achieving female peers in mathematics increases the prospect of a woman following a STEM degree, and male peers have the diametrical opposite effect. Men are not affected. The logic behind is that a student compares themselves with other classmates and are encouraged to pursue the STEM tracks (affirmation effect). The gender stereotypes, ability beliefs, and preferences are influenced by the presence of high-performing female role models in the class.

## 2.4 Common identification strategies for measuring peer effects

Most papers in the literature of peer effects use a quasi-random assignment of some sort (Fischer, 2017; Duflo et al., 2011; Carrell et al., 2013; Goulas et al., 2018; Goulas et al., 2020; Carrell et al., 2009; Feld & Zölitz, 2017). For example, Lu & Anderson (2015) uses an especially creative random assignment of seating in classrooms in a middle-school to study cooperative learning behavior among students of the same or complementary gender. In contrast, Zimmerman, D. J. (2003) uses data from Williams College where first year roommates are assigned randomly. He argued that the preferences that would create more bias (those that are correlated with academic ability i.e. prior academic performance, ethnicity, etc.) in the housing preferences form are not used during the allocation process of rooms. Finally, Fischer, S. (2017) identification strategy comes from the random assignment of freshman to introductory chemistry classes.

In contrast, and to the best of my knowledge, not many papers have used instrumental variables to cure the selection bias of students selecting their classes and classmates (Dearden et al., 2002; Betts et al., 2000). Specifically, Dearden et al. (2002) used instrumental variable methodology to try to differentiate direct effect of peers in academic achievement from values of peer effects. An example of values is that of better-behaved individuals, value education more highly, or have parents that take part in their educational success. Betts et al. (2000) instead used instrumental variable estimation to check for peer effects of grouping high, average and low-achieving students.

Nevertheless, even if its obviated in much empirical research, an exclusion restriction is needed when using the instrumental variable approach due to the reflection problem (Manski, 1993). A non-identification issue that arises when one tries to predict the behavior of an individual using the behavior of the group. These restrictions are usually very difficult to proof or even motivate in much of empirical research.  Recently, Lin & Tang (2022) have offered an innovating solution to the problem by showing that no exclusion restriction is really necessary when using exogenous instruments to take care of the bias created by endogenous covariates. They also do an empirical exercise on peer effects on math test scores in grade 3. They use the self-reported motivation scores of students as instruments for grade 2 math scores. Due to this innovating result, I expect that the use of instruments will become more common in the literature in the near future since the results will be even more convincing.

## 2.5 Well established facts in the literature

Card & Payne (2021) found that females usually have about the same math and sciences grades as males but higher grades in the other courses that determine university rankings. This absolute non-STEM advantage explained most of the gender difference in the probability of pursuing a STEM major, conditional on being STEM-ready at the end of high school. As another example, Breda & Napp (2019) show that female students who are good at math are much more likely than male students to be even better in reading. Once again, this absolute advantage in math and reading abilities explained as much as 80% of the gender gap in pursuing math related studies and careers.

In addition, Card & Payne (2021) also indicate that entry to STEM programs depends on STEM readiness or end-of-high school courses in math and science. This is a very important fact to consider in our setting since students will select a track choice of study in grade 11 and 12 before applying to university.

Furthermore, Zimmerman, D. J. (2003) found that peer effects are almost always linked to verbal scores and that they are predominantly significant in the middle of the distribution (i.e. middle 70%) but are not large. I expect that peer effects in modern Greek, a language course, to have the biggest superstar effects. In accordance with this, Betts & Shkolnik (2000) found that grouping students doesn’t have an impact in math grades. No significant effect of grouping high-achieving, average or low-achieving students. They also found that lower achieving classes are usually smaller and that high-achieving classes have better teachers.

In comparison, Lavy & Schlosser (2011) centered their findings around class size. They found that class size reduction and a higher proportion of female peers are two instruments to reduce classroom disruptions and violence. They recommend that cohorts with a lot of boy should try to accommodate boys in smaller classes. Due to these famous results I will be controlling for class size and proportion of females in the upcoming regressions.

There are some laboratory experiments that show that there are no gender differences in performance when solving a task (Niederle & Vesterlund, 2007). Nonetheless, men select the competitive scenario twice as much as women when choosing their compensation tournament incentive scheme. This is evidence that men have a tendency of being overconfident and of the existence of gender differences in preferences after controlling for performance. Peer effects, especially superstar effects, could then vary greatly across genres. For example, the overconfidence in men could, in their heads, shorten the ability gap between the superstar and the male student. Hence, men could potentially not be affected by the existence of a superstar or at the very least, the magnitude could be considerably lower.

Another interesting field of study is the interaction of teacher and peer effects. This thesis won't be able to study or control for teacher effects. Nevertheless, it seems that peer effects work indirectly by influencing teacher effort and choice of target teaching level (Duflo et al., 2011). In addition, teaching experience raises scores especially in reading subjects and there is a large difference in teacher's quality within schools and teacher quality has an impact on scores (Rockoff, 2004).

Literature has also been done around the incentive effects from both financial and non-financial incentives on test scores. Non-financial incentives are considerably more cost-effective than financial incentives for young students, but were less effective with the older students (Levitt et al., 2016). With no incentives, students tend to have low effort on standardized tests. Thus, potentially biasing measures of student ability, teacher value added, school quality, and achievement gaps. This incentive could be changing across school characteristics. Thus, outcome variables such as university matriculation exam, other than presenting the important quality of being homogenous across all schools, could be more "bias free" since they are the means by which you access university studies. Clearly, a better performance in the national exam is a non-financial incentive of some sort since students are more likely to study their career of choice if they get a higher performance. It could even entail a long-term financial incentive if students believe that their future income depends dramatically on the university they attend.

Another innovating area of peer effect literature that has recently opened up tries to study the effects of the interaction of peer effects and competition over the individual creativity and not just over educational attainment (Gross, 2020). Preliminary results show that increasing competition results in more original/innovative ideas, especially in the high ability group. Nevertheless, when the competition is fierce the results are negative. Hence, it is another example of the perverse effect of heterogenous abilities in effort in the education sector. Due to data limitations the study won't be able to evaluate the effect on having a superstar in the classroom on creativity outcomes.

In general, the different empirical evidence of peer effects provided in this chapter can be given a theoretical foundation and easily rationalized by observing that the education production function is characterized by having a customer input i.e. student/parent quality (Zimmerman, D. J., 2003). Something that doesn't occur in many markets.

# Chapter 3

# Data description

In this section, the data and institutional setting of high schools in Greece will be explained in detail.

## 3.1 Institutional Setting

In Greece the high school academic year consists of 2 semesters: namely, semester 1 that starts at the beginning of September in the current year and semester 2 which starts in the beginning of February of the following year. A semester lasts a total of 16 weeks. Students are assigned to a physical classroom where all the lectures transpire.

This project will exploit a unique system that exploits quasi-random assignment of students to classrooms in the Greek setting. This comes from the institutional setting of assigning students to classroom lexicographically or alphabetically at the start of the $10^{th}$ grade. It is important to notice that this is a requirement set by law. There will be use of descriptive statistics and some evidence to show that the assignment of class is indeed random later on in the text. Moreover, law dictates that teachers must also be assigned randomly to public schools.

As a direct consequence of this law, students with a last name starting with a letter at the beginning of the alphabet are assigned to a smaller classroom number than students with a last name starting with a letter further down in the alphabet. Luckily, after this class ordering has been constructed students cannot change classrooms.

It is important to notice that students are assigned to public schools based on their residential address and its proximity to a selected school. Hence, most students are not able to change high schools after it is centrally assigned. I will not be able to test for this since the data identifies the student at the school level. As a result, students cannot be tracked across schools. Clearly, students in private schools are exempted from this type of rule. But, this will not affect as significantly since more than 92% of students in the Greek educational system pursue studies in a public school (Goulas et al., 2018).

Furthermore, students at grade $11^{th}$ must choose a track choice based on the field of study they wish to pursue at university. They must select one of the following 3 tracks, namely Science, Information & Technology (IT) and Classics. Hereafter, for analysis purposes, the Science and IT tracks are treated jointly as part of a unified STEM track choice. Evidently, in each track choice each student takes different elective courses. In grade $12^{th}$ students have the option of changing track if they do desire. Having said that, it is a very unlikely phenomenon.

Also, it is important to notice that grade $12^{th}$ is the last year of high school and hence, the most relevant for university admission. During this intense year, students do not only have to take school-level exams but a national exam which consists on track choice courses and an elective exam. Track related subjects are given a bigger weight in the construction of the admission grade. Exams are administered by the Ministry of Education during the period beginning at the end of May and the start of June. As most centralized exams, this exam is blind and graded centrally. The overall national exam performance or admission grade is the only criterion used for universities admissions.

Once the results of the national exam have been disclosed to students they must construct a list with ranked choices of the degrees they wish to pursue in a certain university. Afterwards, the Ministry of Education assigns the top-ranking students to their preferred choice. One can easily follow the logic of how this allocation continues for the lower-ranking individuals subject to space availability.

## 3.2 Data description

I use a panel data that consist of a sample of 134 high schools in Greece that runs across the years 2001-2009. The sample corresponds to approximately 10% of public schools in Greece and 70 000 students (Goulas et al., 2020). In addition, Goulas and Megalokonomou (2015) has shown that the panel data used in this study is a representative sample of the whole Greek secondary educational system with regards to a myriad of important variables such as female share and track choice selection rates.

Each student has an individual identifier at the school level. Furthermore, there is information for all test scores across all high school grades (i.e. $10^{th}$-$12^{th}$grade), detailed information on the student as well as some demographic information: age at $10^{th}$ grade, gender, national exam performance, high school information (private vs. public), neighborhood income, STEM track choice, university admission, STEM application to university, etc.

There are five core or compulsory courses which are taken by all students in all grades. From these three are STEM-related: Algebra, Geometry and Physics. While Modern Greek and History are treated as non-STEM. The test scores from the previously mentioned courses are observed throughout high school. The baseline control that measures the prior performance of a student before exposure to superstar will be constructed by using the $1^{st}$ exam in semester 1 of grade 10.

A limitation of the use of the baseline test scores is that it is was not taken the first few days of class before any type of interaction. Instead, it was taken inside a two-month timeframe. Its exact nature was a mid-semester exam in the $1^{st}$ semester of grade 10. Nevertheless, one can still consider the student performance in these early exams as a benchmark since they take place shortly after randomization. In addition, it is the measurement with the least interaction between peers and teachers at my disposal.

In Greece, test scores go from 0-20 but all grades used throughout the analysis will be standardized accordingly for external validity and comparative purposes between subjects. These grades will be used to construct most of the consequent control variables. As well as for classifying students as superstars or not. For example, a student will be classified as a standard superstar if he has obtained a test score over 18 in each of the core courses in the baseline exam. While a student will be classified as a STEM superstar in the scenario that he has obtained a test score above 19 in the three STEM subjects, viz., algebra, geometry and physics. Similarly, a student will be classified as a non-STEM superstar if he achieved a test score above 19 in history and modern Greek.

As mentioned before, I also have information regarding university admission, university application and STEM application. The data defines a STEM degree as a degree offered at Science or Engineering and Technology departments. As such, degrees as Economics or even Medicine are not considered. For the sake of preciseness, Science or Engineering and Technology departments include mathematics, engineering, physics, computer science, biology, chemistry, pharmacy, veterinary. Non- STEM university departments are those in liberal arts, literature, psychology, journalism, philosophy, education, Greek language, history, foreign languages, home economics and law, economics, statistics, business and management, accounting, political science and European studies.

## 3.3 Defining superstar

For starters, I will define the standard superstar and built on it. For this, it's important to remember that test scores go from 0 - 20 and that the STEM core courses are algebra, geometry and physics. While the non-STEM courses are modern Greek and history. As mentioned before, the standard superstar variable in the analysis will be defined as an indicator function that takes the value of 1 if a student has obtained a test score over 19 across all core courses in the baseline exam and 0 otherwise. Namely, observation i will be classified as a superstar if the following indicator variable takes the value of 1:

$$standard\ superstar_i = \mathbb{I}\{i\ test\ scores\ are\ over\ 18\}$$

It will be worthwhile to subdivide the standard superstar population separately by gender. Namely, observation i will be classified as a female superstar (supergirl) and male superstar (superboy) if the following indicator variable takes the value of 1, respectively:

$$supergirl_i = \mathbb{I}\{\ i\ is\ female\ and\ her\ test\ scores\ are\ over\ 18\}$$

$$superboy_i = \mathbb{I}\{\ i\ is\ male\ and\ his\ test\ scores\ are\ over\ 18\}$$

Finally, observation i will be classified as a STEM superstar and non-STEM superstar if the following indicator variables take the value of 1, respectively:

$$STEM\ superstar_i = \mathbb{I}\{i\ test\ scores\ in\ the\ STEM\ core\ subjects\ are\ over\ 19\}$$

$$non-STEM\ superstar_i = \mathbb{I}\{i\ test\ scores\ in\ the\ non-STEM\ core\ subjects\ \ are\ over\ 19\}$$

In a similar fashion I will also define:

$$STEM\ supergirl_i = \mathbb{I}\{i\ is\ female\ and\ is\ a\ STEM\ superstar\}$$

$$non-STEM\ supergirl_i = \mathbb{I}\{i\ is\ female\ and\ is\ a\ non-STEM\ superstar\}$$

$$STEM\ superboy_i = \mathbb{I}\{i\ is\ male\ and\ is\ a\ STEM\ superstar\}$$

## 3.4 Descriptive statistics

Figure 1 shows the dynamics of type of superstar across the years of interest. One can clearly observe that there are more female than male superstar. Furthermore, it is also interesting to note that the counterintuitive fact that there are more STEM superstars than non-STEM superstars in every year even though the STEM requirements are stronger. In particular, the definition required that the tests scores threshold had to be met across three subjects. Whereas the weaker non-STEM condition was done across two subjects. It would seem that the effect produced by the higher variability in STEM test scores recorded in Table 1 and 2 is dominating the inorganic effect produced by my ad-hoc superstar definitions.

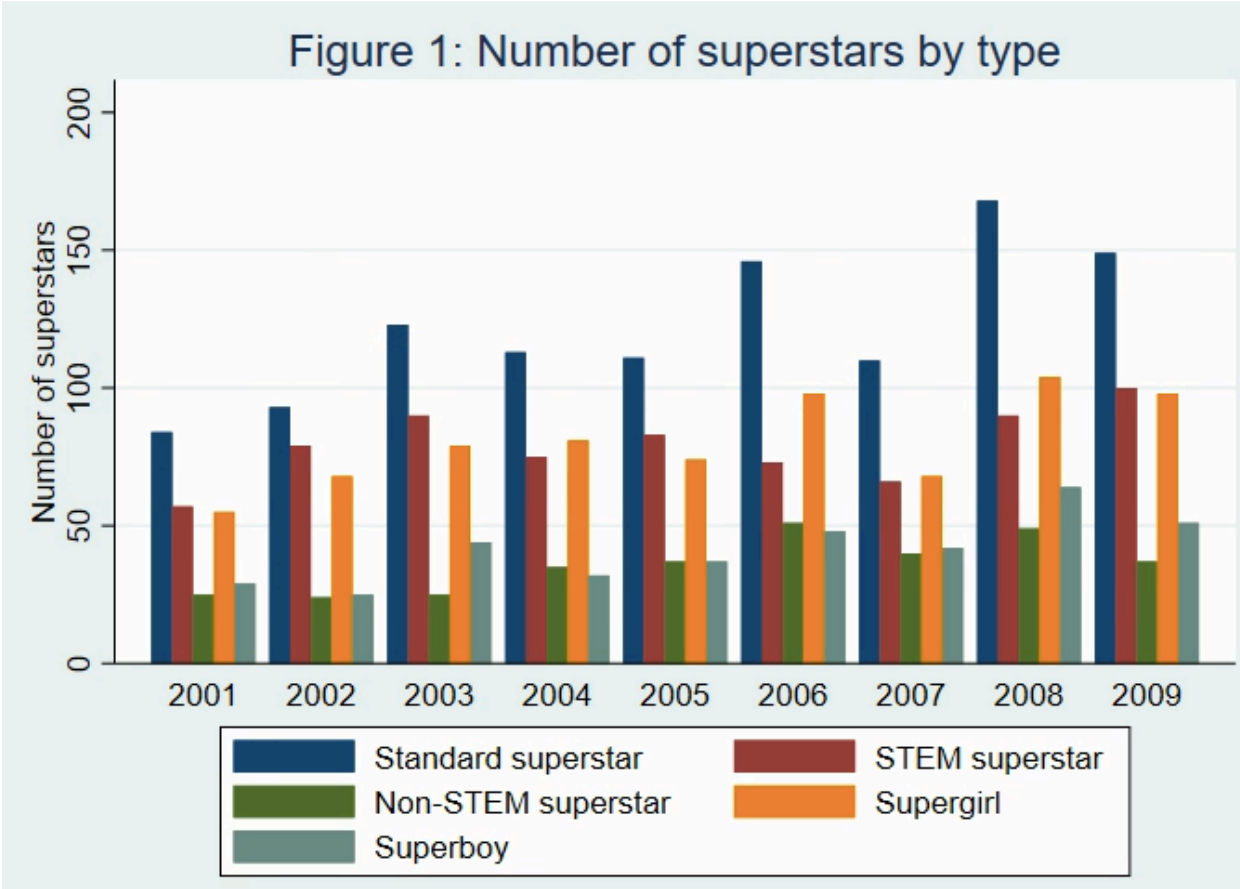


Table 1 allows for the visualization of the average performance by gender and subject. Clearly, it shows that there is a significant gender difference that favors girls in the baseline exam performance in most of the core subjects as well as in the overall performance. The exception is history, where there is no significant difference.

It is also important to note that the difference regarding female over performance in STEM courses (difference > 0.2) is considerably higher than in non-STEM courses (difference < 0.127).

Table 1: Gender differences in average performance in the baseline exam (grade 10)

| | Number of males | Number of females | Mean score of males | Mean score of females | Diff. | t-value | p-value |
|---|---|---|---|---|---|---|---|
| Algebra test score (baseline exam) | 28598 | 36093 | 14.662 | 14.866 | -.204 | -8.55 | 0 |
| Geometry test score (baseline exam) | 28598 | 36093 | 14.367 | 15.329 | -.962 | -52.7 | 0 |
| History test score (baseline exam) | 28598 | 36093 | 14.892 | 14.891 | 0 | 0 | .995 |
| Modern Greek test score (baseline exam) | 28598 | 36093 | 14.675 | 14.802 | -.127 | -5.25 | 0 |
| Physics test score (baseline exam) | 28598 | 36093 | 15.410 | 16.107 | -.697 | -30.55 | 0 |
| GPA of all courses (baseline exam) | 28598 | 36093 | 14.801 | 15.199 | -.398 | -21 | 0 |

Notes: Reports the gender differences in performance in the first exam of grade 10 for the five core subjects and the GPA score. The raw scores are out of 20.

Firstly, Panel 1 of Table 2 describes some basic statistics of the classification of students as a type of superstar. In proportion to all the students, there are approximately two times as many STEM superstar (supergirls) than non-STEM superstar (superboys).

Secondly, Panel 2 of Table 2 shows the same basic statistics for the controls used in the regressions of this study and some variables at the school and class level. As can be seen in the relevant panel, I will control for student-level controls, student predetermined characteristics and class-by-year characteristics. Similar to what was mentioned previously, the courses in the baseline exam with a higher mean are the two non-mathematical subjects, history and modern Greek. Most of the educational system in Greece is public since just 2.8% of schools are private.

**Table 2: Descriptive statistics**

Panel 1: By type of superstar

| VARIABLES | (1) Mean | (2) SD | (3) Max. | (4) Min. |
|---|---|---|---|---|
| Superstar | 0.017 | 0.129 | 1.000 | 0.000 |
| STEM superstar | 0.011 | 0.104 | 1.000 | 0.000 |
| Non-STEM superstar | 0.005 | 0.070 | 1.000 | 0.000 |
| Supergirl | 0.011 | 0.105 | 1.000 | 0.000 |
| Superboy | 0.006 | 0.076 | 1.000 | 0.000 |

Panel 2: Controls used in regressions

| | | | | |
|---|---|---|---|---|
| Female | 0.558 | 0.497 | 1.000 | 0.000 |
| Age at grade 10 | 15.926 | 0.443 | 25.000 | 15.000 |
| Private school | 0.028 | 0.165 | 1.000 | 0.000 |
| Morning students | 0.122 | 0.327 | 1.000 | 0.000 |
| Algebra test score (baseline exam) | 14.776 | 3.008 | 20.000 | 6.000 |
| Geometry test score (baseline exam) | 14.746 | 3.056 | 20.000 | 6.000 |
| History test score (baseline exam) | 15.799 | 2.898 | 20.000 | 7.000 |
| Modern Greek test score (baseline exam) | 14.904 | 2.356 | 20.000 | 7.000 |
| Physics test score (baseline exam) | 14.891 | 3.019 | 20.000 | 6.000 |
| GPA of all courses (baseline exam) | 15.023 | 2.400 | 20.000 | 8.200 |
| Class av. test score in Modern Greek | 14.904 | 1.095 | 19.211 | 11.077 |
| Class av. test score in Physics | 14.891 | 1.284 | 19.739 | 10.333 |
| Class av. test score in History | 15.799 | 1.286 | 19.526 | 10.850 |
| Class av. test score in Algebra | 14.776 | 1.250 | 18.941 | 9.824 |
| Class av. test score in Geometry | 14.746 | 1.277 | 19.733 | 9.368 |
| Class size | 20.315 | 4.243 | 33.000 | 10.000 |
| Proportion of females | 0.558 | 0.132 | 1.000 | 0.063 |

Notes: This table shows summary statistics for the classification of students as a type of superstar. A superstar is a student that obtains a grade over 18 points in every core course in the baseline exam. The STEM superstar and non-STEM superstar are defined by obtaining more than 19 points in STEM courses and non-STEM courses, respectively. Lastly, I have defined girl and boy superstar using the standard definition of a superstar with the addition of classifying it by gender. Scores go from 0-20 points.

Table 3 presents the summary statistics of short and long-term outcome variables. In essence, Panel 1 shows the final exams test scores in semester 2 of grade 10. History and modern Greek still possess the highest mean scores as well as the least variability.

Similarly, Panel 2 gives the summary statistics for mid-term variables such as the overall performance in grade 11 by core course and selected track choice. As well as long-term variables such as national exam performance and other university related outcomes. For example, 63% of students select one of the two pathways in the STEM track. There also seems to be a big gap between STEM applications and being admitted to a STEM program. Over the years of interest of this study, 83% of students have been admitted to university and no student has managed to score a clean score in the national exam.

**Table 3: Further descriptive statistics**

Panel 1: Short term outcome variables

| VARIABLES | (1) Mean | (2) SD | (3) Max. | (4) Min. |
|---|---|---|---|---|
| End of grade 10 test score in Algebra | 10.48 | 5.53 | 20.00 | 0.00 |
| End of grade 10 test score in Geometry | 10.70 | 5.67 | 20.00 | 0.00 |
| End of grade 10 test score in History | 12.87 | 4.69 | 20.00 | 0.00 |
| End of grade 10 test score in Modern Greek | 14.24 | 3.06 | 20.00 | 0.00 |
| End of grade 10 test score in Physics | 11.12 | 5.06 | 20.00 | 0.00 |

Panel 2: Long-term outcome variables

| | | | | |
|---|---|---|---|---|
| National exam performance | 13.06 | 3.93 | 19.84 | 1.17 |
| Admitted to a university | 0.83 | 0.38 | 1.00 | 0.00 |
| STEM Application | 0.62 | 0.49 | 1.00 | 0.00 |
| STEM Admitted | 0.11 | 0.31 | 1.00 | 0.00 |
| Classic track in grade 11 | 0.37 | 0.48 | 1.00 | 0.00 |
| STEM track in grade 11 | 0.63 | 0.48 | 1.00 | 0.00 |
| Algebra test score in year 11 | 13.49 | 3.89 | 20.00 | 3.77 |
| Geometry test score in year 11 | 13.28 | 3.93 | 20.00 | 2.33 |
| History test score in year 11 | 14.49 | 3.30 | 20.00 | 4.87 |
| Modern Greek test score in year 11 | 14.83 | 2.41 | 20.00 | 6.33 |
| Physics test score in year 11 | 13.81 | 3.68 | 20.00 | 4.67 |

Notes: This table shows summary statistics for the output variables used in the following regression analysis. Grades go from 0-20 points.

# Chapter 4

# Empirical Framework

This chapter presents the methodology that allows to give some evidence for the identification strategy that students are randomly assigned to classrooms as well as the econometric methods employed to answer the questions presented in Chapter 1 using the data and institutional setting explained in Chapter 3.

## 4.1 Identification strategy

Measuring peer effects is a difficult enterprise since student's educational achievements depend on numerous factors other than the characteristics of the relevant peers. This can be aggravated and become quite problematic when students select the peers and classrooms they would like to attend with (selection bias). Specifically, in the case that students are not randomly assigned, if there exists an individual or school characteristic that is omitted but affects the student's achievement and if its correlated with the student or class characteristics the estimate of the peer effect will be biased. Nevertheless, if students are randomly assigned to classrooms then the estimate of interest will be unbiased.

To give some evidence for the pseudo-random assignment of students to classrooms I will construct a dummy variable that takes the value of 1 if the class c is the first class in the school. Similarly, I will create a total of 5 dummy variable for each class number since schools have from 1 to 6 classes.

Let $X_c$ be a class level control. Then, estimating the following regression using fixed effects and clustered standard errors at the school by year level

$$X_c = \beta_0 + classnumber1 * \beta_1 + classnumber2 * \beta_2 + classnumber3 * \beta_3 + classnumber4 * \beta_4 + classnumber5 * \beta_5 + \varepsilon_{cst} \quad (1)$$

and testing for the following null hypothesis $H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$ will provide the required evidence.

## 4.2 Econometric specification

The description of the problem suggests the following initial econometric specification:

$$y_{icst} = \beta_{s,t} + X_i^T\delta + S_{cst}^T\pi + \omega D_{cst} + \varepsilon_{icst} \qquad (2)$$

where $y_{icst}$ is the achievement measure for student i, in class c, in school s, in time t; $\beta_{s,t}$ is a school by year fixed effect; $X_i$ is a matrix of student level controls; $S_{cst}^T$ is a matrix of class level controls; $D_{cst}$ is a dummy variable where 1 represents the case that the class c in school s in cohort t has a superstar; $\varepsilon_{icst}$ is the error term that allows for any type of correlation within observations of the same school, time, individuals and class. Finally, the coefficient of interest is $\omega$ (omega) since it will capture the effect of being exposed to a superstar on the students' achievement measure. Clearly, it should be estimated using fixed effects.

As mentioned above, for the purpose of this study I derived a classroom that has been exposed to a superstar dummy variable for each type of superstar. These indicators variable will take the value of 1 if the student i's class has been exposed to a certain superstar and 0 otherwise. It is also important to note that superstar observations should be dropped before running the regressions since the problem in question is interested in studying the impact of exposure to superstars.

Finally, Abadie et al. (2017), recommends clustering at a higher level of aggregation than that of the randomization (in this case class-level) when restricted to finite samples. Following this criteria, standard errors in regressions have been clustered at the school by year level. Allowing for heteroskedasticity and serial correlation among students within each school.

## 4.3 Econometric specification extensions

Another question of interest was the existence of the perverse effect of heterogenous abilities in effort in a classroom environment. Thus, it is required to check the effect of a superstar on the second-best student in the classroom as well as its impact on the top 3 and 5 students as a whole. For clarification, the second-best student has been defined as either the best student after the superstar in a classroom with superstar or the best student in a classroom that hasn't been exposed. The top 3 and 5 students have been identified using the same logic.

For this I will require two different methodologies. To measure the impact on the top 3 and 5 students as a whole I just need to adapt the equation (2) by adding a top 3 and top 5 student dummy variable and its interaction with $D_{cst}$, accordingly. As a result (2) is modified into:

$$y_{icst} = \beta_{s,t} + X_i^T \delta + S_{cst}^T \pi + \omega D_{cst} + \pi T_{icst} + \rho D_{cst} * T_{icst} + \varepsilon_{icst} \tag{3}$$

where $T_{icst}$ is a dummy variable that takes the value of 1 if the student i is a top student in class c in school s in time t. It is important to notice that $\omega + \rho$ captures the effect of interest i.e. the superstar effect on a top student. In contrast, $\pi$ just captures the average difference in achievement between a top student and non-top student. Finally, $\rho$ just captures the difference in the superstar effect between a top and a non-top student.

To measure the exposure of a superstar on the second best I can just take (2) and run a regression conditioned on the data points being the second best in the classroom. I didn't run the regressions with the dummy and interaction term specification because in each class only one individual is defined as the second-best.

# Chapter 5

# Empirical Results

## 5.1 Introduction

First of all, all regressions contained in this chapter have been estimated using the areg command in Stata. Secondly, Section 5.2 presents evidence that the random assignment of students to classrooms actually holds while section 5.3 presents the initial estimate for the standard superstar effects on short- and long-term scholastic outcomes as well as the limitations to its veracity. Thirdly, section 5.4 and 5.5 present the superstar effect by type of superstar and how the effect differs across genders. Section 5.6 and presents some heterogeneity exercises. Furthermore, section 5.7 presents the STEM and non-STEM supergirl/superboy effects separately by gender on long-term academic outcomes. Section 5.8 studies how the superstar effects affect high achieving students. Finally, section 5.9 will provide some robustness exercises by using other definitions to classify students as superstars.

## 5.2 Evidence for the identification strategy

As was previously discussed, the estimates of the superstar effect presented later on this chapter will only be unbiased when students are randomly assigned to classrooms. To give some evidence that this identification strategy holds in practice, I have employed a series of regressions following the specification in equation (1) to test for balance across several outcomes. The treatments consist of students being lexicographically assigned to a classroom number.

To do this, I have regressed the classroom number with a dummy variable that takes the value of 1 if the classroom has a superstar (column 1), classroom average in Modern Greek (column 2), classroom average grade in Physics (column 3), classroom average in History (column 4), classroom average in Algebra (column 5), classroom average in Geometry (column 6) on the dummy variables representing classroom numbers. All classrooms average scores have been constructed using the results in the baseline exam. As mentioned in the previous chapter, classroom is the unit of observation.

It can immediately be seen that Table 4 shows evidence of random assignment of students into classrooms at the start of grade 10 by noticing that all the coefficients are not significant. As a result, the presence of a superstar and class average in every core subject don't differ across class numbers. The p-values presented are for the test of joint significance. Most importantly, the p-value in column 1 suggest that the classroom number is not associated with differences in classroom-level exposure to a superstar. The only p-value not significant at the 5% level is in column 4. As a result, I can infer that at least one of the coefficients from that column is not significant at the 5% level of significance. Nonetheless, it is barely significant at the mentioned level (p-value=$0.0498$) and obviously not significant at the 10% level.

As a result of this discussion, the results presented in the table are precisely what is expected to happen by chance. It would seem to indicate that randomization has succeeded in generating balance across treatments.

Table 4: Evidence of random assignment of students into classrooms in grade 10

| VARIABLES | (1)<br>Classroom with a Superstar | (2)<br>Class Avg. Mod. Greek | (3)<br>Class Avg. Physics | (4)<br>Class Avg. History | (5)<br>Class Avg. Algebra | (6)<br>Class Avg. Geometry |
|---|---|---|---|---|---|---|
| classnumber=1 | -0.139 | -0.039 | 0.073 | -0.177 | -0.023 | -0.124 |
| | (0.101) | (0.153) | (0.192) | (0.220) | (0.197) | (0.213) |
| classnumber=2 | -0.134 | -0.104 | 0.048 | -0.212 | -0.110 | -0.139 |
| | (0.102) | (0.154) | (0.189) | (0.221) | (0.198) | (0.214) |
| classnumber=3 | -0.151 | -0.096 | 0.059 | -0.077 | -0.026 | -0.085 |
| | (0.102) | (0.156) | (0.192) | (0.222) | (0.200) | (0.216) |
| classnumber=4 | -0.130 | -0.035 | 0.028 | 0.002 | 0.030 | -0.014 |
| | (0.102) | (0.157) | (0.193) | (0.217) | (0.197) | (0.210) |
| classnumber=5 | -0.120 | -0.015 | 0.096 | -0.006 | 0.101 | 0.038 |
| | (0.104) | (0.157) | (0.205) | (0.230) | (0.206) | (0.218) |
| Observations | 3,357 | 3,357 | 3,357 | 3,357 | 3,357 | 3,357 |
| R-squared | 0.441 | 0.519 | 0.620 | 0.608 | 0.589 | 0.568 |
| Adj. R-squared | 0.194 | 0.305 | 0.451 | 0.434 | 0.407 | 0.377 |
| F-statistic | 0.559 | 0.571 | 0.154 | 2.225 | 1.393 | 0.792 |
| P-value | 0.731 | 0.723 | 0.979 | 0.0498** | 0.224 | 0.555 |

Notes: The table shows results of the estimated effects of the classroom number on a variety of outcomes. The outcome variables are reported in the column. In particular, I regress the classroom number with a dummy variable that takes the value of 1 if the classroom has a superstar (column 1), classroom average in Modern Greek (column 2), classroom average grade in Physics (column 3), classroom average in History (column 4), classroom average in Algebra (column 5), classroom average in Geometry (column 6). All scores are from the baseline exam. Classroom is the unit of observation. *, ** and *** denote significance at the 10%, 5% and 1% level, respectively. Robust standard errors clustered at the school by year level are reported in parentheses. Lastly, the p-values presented are for the test of joint significance. The p-value in column 1 suggest that the classroom number is not associated with differences in classroom-level exposure to a superstar.

## 5.3 Standard superstar effect

Table 5 reports results from regressions estimates for the standard superstar effect on the average student across all core subjects in the final exam of semester 2 of grade 10. I run the model specification described in equation (2) while using a myriad of outcomes. These include the standardized student performance in algebra, physics, history, modern Greek and geometry in the final exam of grade 10 in columns (1), (2), (3), (4), and (5), respectively. The regression controls for student characteristics such as gender, age, GPA and prior performance in the same course of the outcome variable; and, class by year characteristics such as class size, proportion of females and avg. class GPA across all subjects. I will reuse this specification for the following heterogeneity exercises contained in this chapter unless it is duly noted.

The main variable on interest is the estimate for the superstar effect or the effect of being exposed to a superstar, but I also report the estimated effects for the remaining covariates, as they are interesting per se. For instance, the superstar effect under this initial specification is just significant in modern Greek and geometry which are of moderate magnitude. In particular, the scores are reduced by 0.047 and 0.031 standard deviations when a student is exposed to a superstar, accordingly.

Without trying to be pedantic, it is important to have in mind that the measurement of this impact on students exposed to a superstar compared with students who are not exposed to a superstar is conditional on school by year fixed effects. It is crucial to include school by year fixed effects to control for the most obvious potential confounding factor—the endogenous sorting of students across schools in a given year and make sure I compare comparable students.

Interestingly, proportion of females doesn't seem to be affecting the performance of a student in any of the subjects as I had expected due to do the literature. There is also another unexpected result. In average, an additional student in the classroom raises the student grade by 0.006 standard deviations across subjects.

On average, a female student test scores in algebra, physics and geometry are more than 0.1 standard deviation below the scores achieved by male peers. It would seem that in Greece there is still a substantial STEM gender gap. In contrast, on average, a female student test scores in modern Greek are 0.104 standard deviation above the scores achieved by male peers. As a result, the absolute non-STEM advantage for women (i.e. within-individual relative academic strengths and weakness in non-STEM subjects compared to STEM ones), could be aggravating women decisions of not taking pro STEM decisions.

Finally, I would like to mention that my definition of superstar is quite ad hoc at first sight. Having said that, I will give an argument to motivate the specific threshold in the next paragraph. For now, I would like to mention that due to the difficulty of motivating my definition, I have done some robustness exercises with some other extreme definitions and obtained similar superstar effects in modern Greek and geometry plus a negative and positive superstar effect for algebra and physics, respectively (see Table 24 and 25 in section 5.7).

For illustration purposes I have lowered the threshold in the superstar definition from 18 to 17.8 in each of the core courses. The results change drastically since the superstar almost quadrupled! To be exact, the number of superstars went from 1,097 to 4,017. Table 6 presents the mentioned sensibility. The most striking fact is that superstars' effect for physics and algebra become positive. In specific, exposure to said superstar increases, in average, algebra (physics) scores by 0.036 (0.025) standard deviations.

Nevertheless, this is not so necessarily invalidating my results. For example, literature shows that high achievers (i.e. where the definition is weaker) generally create a positive effect on classroom scholastic outcomes (Busso & Frisancho, 2021; Lavy et al., 2008; Cools et al., 2019). In contrast, even if not presented for obvious reasons that will follow, when I expanded the threshold to 18.2-18.5, the results didn't change at all since the number of superstars identified stayed the same throughout the different specifications, viz., after the threshold nobody becomes a superstar and before it a lot of people become a superstar. Hence, I am correctly identifying an ability gap, i.e. I am differentiating superstars from high achievers.

Table 5: Estimates of the exposure of a superstar by performance in the final exam of the core course in grade10

| VARIABLES | (1)<br>Algebra | (2)<br>Physics | (3)<br>History | (4)<br>Modern Greek | (5)<br>Geometry |
|---|---|---|---|---|---|
| Exposure to a superstar | -0.020 | 0.011 | -0.019 | -0.047** | -0.031** |
| | (0.013) | (0.013) | (0.015) | (0.019) | (0.014) |
| Female | -0.105*** | -0.175*** | -0.043*** | 0.104*** | -0.195*** |
| | (0.006) | (0.006) | (0.006) | (0.006) | (0.006) |
| Class size | 0.005** | 0.005** | 0.006** | 0.007** | 0.007*** |
| | (0.002) | (0.002) | (0.002) | (0.003) | (0.002) |
| Proportion of females | 0.031 | 0.014 | -0.029 | 0.051 | 0.057 |
| | (0.039) | (0.039) | (0.044) | (0.052) | (0.043) |
| GPA (baseline exam) | 0.467*** | 0.538*** | 0.379*** | 0.255*** | 0.448*** |
| | (0.008) | (0.007) | (0.007) | (0.006) | (0.008) |
| Age at grade 10 | -0.076*** | -0.033*** | 0.000 | -0.055*** | -0.021** |
| | (0.009) | (0.009) | (0.010) | (0.010) | (0.009) |
| Morning students | -0.048*** | -0.010 | -0.003 | -0.065*** | 0.006 |
| | (0.012) | (0.012) | (0.013) | (0.012) | (0.012) |
| Avg. class GPA (baseline exam) | -0.244*** | -0.235*** | -0.187*** | -0.105*** | -0.278*** |
| | (0.024) | (0.023) | (0.026) | (0.030) | (0.025) |
| Observations | 63,594 | 63,594 | 63,594 | 63,594 | 63,594 |
| R-squared | 0.603 | 0.622 | 0.562 | 0.619 | 0.589 |
| School by year FE | YES | YES | YES | YES | YES |
| Control | YES | YES | YES | YES | YES |
| Adj. R-squared | 0.597 | 0.616 | 0.555 | 0.613 | 0.582 |

Notes: Each column variable represents the grade achieved by the student in the final exam of Year 10 of the relevant course. Regressions control for school by year fixed effects; student-level controls; student predetermined characteristics; and class-by-year characteristics. Each estimate is generated from a different regression. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively. Robust standard errors clustered at the school by year level are reported in parentheses. All grades presented have been standardized. Control indicates that prior performance in the baseline exam in that specific course is being controlled for. Nevertheless, due to presentation purposes it is not showed.

Table 6: Estimates of the exposure of a superstar by performance in the final exam of the core course in grade 10 by decreasing the superstar definition threshold by 0.2 points

| VARIABLES | (1)<br>Algebra | (2)<br>Physics | (3)<br>History | (4)<br>Modern Greek | (5)<br>Geometry |
|---|---|---|---|---|---|
| Exposure to a superstar | 0.036*** | 0.025** | -0.005 | -0.020 | 0.011 |
| | (0.012) | (0.012) | (0.015) | (0.016) | (0.013) |
| Female | -0.110*** | -0.181*** | -0.048*** | 0.105*** | -0.202*** |
| | (0.006) | (0.006) | (0.006) | (0.006) | (0.006) |
| Class size | 0.005** | 0.005** | 0.006** | 0.007** | 0.007*** |
| | (0.002) | (0.002) | (0.002) | (0.003) | (0.002) |
| Proportion of females | 0.026 | 0.007 | -0.032 | 0.058 | 0.061 |
| | (0.041) | (0.041) | (0.046) | (0.053) | (0.044) |
| GPA (baseline exam) | 0.452*** | 0.528*** | 0.367*** | 0.253*** | 0.437*** |
| | (0.008) | (0.007) | (0.007) | (0.006) | (0.008) |
| Age at grade 10 | -0.075*** | -0.033*** | 0.000 | -0.055*** | -0.020** |
| | (0.009) | (0.009) | (0.010) | (0.010) | (0.009) |
| Morning students | -0.046*** | -0.008 | -0.000 | -0.064*** | 0.007 |
| | (0.012) | (0.012) | (0.013) | (0.013) | (0.013) |
| Avg. class GPA (baseline exam) | -0.282*** | -0.253*** | -0.204*** | -0.114*** | -0.306*** |
| | (0.024) | (0.025) | (0.027) | (0.031) | (0.026) |
| Observations | 60,674 | 60,674 | 60,674 | 60,674 | 60,674 |
| R-squared | 0.571 | 0.590 | 0.530 | 0.590 | 0.559 |
| School by year FE | YES | YES | YES | YES | YES |
| Control | YES | YES | YES | YES | YES |
| Adj. R-squared | 0.563 | 0.583 | 0.522 | 0.583 | 0.551 |

Notes: Each column variable represents the grade achieved by the student in the final exam of grade 10 of the relevant course. Regressions control for school by year fixed effects; student-level controls; student predetermined characteristics; and class-by-year characteristics. Each estimate is generated from a different regression. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively. Robust standard errors clustered at the school by year level are reported in parentheses. All grades presented have been standardized. Control indicates that prior performance in the baseline exam in that specific course is being controlled for. Nevertheless, due to presentation purposes it is not showed.

Likewise, Table 7 reports results from regressions estimates for the standard superstar effect on the average student across long-term outcome variables such as core courses academic achievement throughout grade 11, national exam performance, and other outcomes concerned with university admission and degree preferences.

This time around, the superstar effect is significant and of similar reasonable magnitude and direction in modern Greek, geometry as well as algebra and physics during grade 11. Specifically, the scores in algebra and physics in grade 11 are reduced by 0.031 and 0.022 standard deviations respectively when a student is exposed to a superstar in grade 10.

Furthermore, there are no superstar effects with respect to longer term outcomes such as higher admission, STEM admission or STEM applications rates. But, there is a negative superstar effect with respect to national exam performance- decreases by 0.019 standard deviations. At first sight, this indicates that superstars are not affecting gender roles nor preference during high school. Since national exam performance is the only criterion used for universities admissions I have run regressions controlling for national exam performance that are not presented here that have similar coefficients and R-squared. It seems that this national exam performance is already captured GPA (baseline exam). For consistency purposes I have decided to just control for the GPA (baseline exam) hereafter. It is important to mention the regressions in column (7) – (9) have a pretty low R-squared indicating that our model doesn't explain much of the variability of the outcome variables. Nevertheless, the low p-values of some coefficients still indicate a real relationship between the predictors and relevant outcome variable.

In addition, one can observe that female students achieve a superior performance in the national exam than males. That is, if a student is a female, in average, she will be score 0.016 standard deviations higher than a male student in the national exam. This can be observed in column (6) of Table 7.

It also reports the well-established fact that women are less likely to apply and to be admitted to a STEM degree than men. The magnitude of the first effect is much higher as can be seen in columns (7) and (9) of Table 7. Expressly, being a female student decreases the probability of applying to a STEM degree by 0.346 percentage points compared to a male student. In Table 9 one can see a similar effect on the STEM track choice in grade 11 outcome variable. In contrast, being female decreases a student probability of being admitted to a STEM degree by 0.103 percentage points.

Once again, proportion of females does not seem to be affecting any of the outcomes of interest and the magnitude of the class size coefficient remains somewhat consistent throughout long-term outcome variables.

Table 7: Estimates of the effect of a superstar by long-term outcome variables

| VARIABLES | (1) Algebra grade 11 | (2) Physics grade 11 | (3) History grade 11 | (4) Modern Greek grade 11 | (5) Geometry grade 11 | (6) Performance in National Exam | (7) Admitted to STEM | (8) Admitted to university | (9) Applied to STEM |
|---|---|---|---|---|---|---|---|---|---|
| Exposure to a superstar | -0.031** | -0.022* | -0.008 | -0.035* | -0.031** | -0.019* | 0.002 | -0.001 | 0.001 |
| | (0.013) | (0.013) | (0.016) | (0.019) | (0.014) | (0.011) | (0.005) | (0.005) | (0.007) |
| Female | -0.164*** | -0.174*** | 0.127*** | 0.250*** | -0.142*** | 0.016** | -0.103*** | -0.037*** | -0.346*** |
| | (0.006) | (0.006) | (0.007) | (0.006) | (0.006) | (0.007) | (0.003) | (0.003) | (0.005) |
| Class size | 0.004* | 0.004** | 0.004* | 0.005* | 0.006*** | 0.007*** | 0.001* | 0.001* | -0.001 |
| | (0.002) | (0.002) | (0.002) | (0.003) | (0.002) | (0.002) | (0.001) | (0.001) | (0.001) |
| Proportion of females | 0.021 | 0.008 | -0.010 | 0.045 | 0.044 | 0.006 | 0.011 | 0.017 | 0.012 |
| | (0.040) | (0.036) | (0.044) | (0.050) | (0.041) | (0.036) | (0.013) | (0.015) | (0.019) |
| GPA (baseline exam) | 0.583*** | 0.624*** | 0.423*** | 0.443*** | 0.627*** | 0.769*** | 0.059*** | 0.183*** | 0.060*** |
| | (0.007) | (0.007) | (0.007) | (0.007) | (0.007) | (0.004) | (0.002) | (0.003) | (0.003) |
| Age at grade 10 | -0.067*** | -0.047*** | -0.022** | -0.120*** | -0.035*** | -0.139*** | -0.011* | -0.066*** | 0.001 |
| | (0.009) | (0.009) | (0.009) | (0.010) | (0.009) | (0.013) | (0.006) | (0.007) | (0.009) |
| Morning students | -0.009 | -0.010 | -0.039*** | -0.124*** | 0.016 | -0.096*** | -0.004 | -0.055*** | 0.015 |
| | (0.011) | (0.012) | (0.013) | (0.012) | (0.012) | (0.016) | (0.007) | (0.009) | (0.011) |
| Avg. class GPA (baseline exam) | -0.297*** | -0.255*** | -0.236*** | -0.354*** | -0.315*** | -0.301*** | -0.033*** | -0.076*** | -0.029*** |
| | (0.024) | (0.022) | (0.027) | (0.030) | (0.024) | (0.023) | (0.008) | (0.009) | (0.011) |
| Observations | 63,594 | 63,594 | 63,594 | 63,594 | 63,594 | 55,389 | 45,716 | 55,389 | 55,389 |
| R-squared | 0.635 | 0.619 | 0.580 | 0.623 | 0.635 | 0.592 | 0.082 | 0.273 | 0.172 |
| School by year FE | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Control | YES | YES | YES | YES | YES | NO | NO | NO | NO |
| Adj. R-squared | 0.629 | 0.613 | 0.573 | 0.617 | 0.629 | 0.585 | 0.0606 | 0.259 | 0.156 |

Notes: Columns (1) to (5) present estimates from school by year fixed effects regression of the grades obtained throughout the grade 11 in each course. Column (6) - (9) present estimates from school by year fixed effects regression of long-term outcomes. Regressions control for school by year fixed effects; student-level controls; student predetermined characteristics; and class-by-year characteristics. Each estimate is generated from a different regression. *, ** and *** denote significance at the 10%, 5% and 1% level, respectively. Robust standard errors clustered at the school by year level are reported in parentheses. All grades presented have been standardized. Control indicates that prior performance in that specific course is being controlled for. Nevertheless, due to presentation purposes it is not showed.

## 5.4 Superstar effect by type of superstar

In this section, I will discuss the results presented in the Tables 8, 9 and 10. In the regressions I use the same specifications used in the previous section.

A priori, there isn't a reason that there cannot exist differences in the effects of a standard superstar and subject specialized or gender specific type of superstar. For illustration purposes, I have chosen two courses, namely a standard math (algebra) and language (modern Greek) courses, as they are the most widely used subjects in the literature. Later on, there are also going to be some interesting dynamics with the superstar effect in Algebra when considering the female superstar exposure. Remember, initially, that there was no superstar effect in algebra (Table 5). But, this won't present itself as a tautological exercise.

To answer this type of question I have defined other types of superstar (see Section 3.3). Table 8 presents the estimates of the superstar effect in the final exam of algebra and modern Greek in semester 2 of grade 10 by STEM and non-STEM superstar. The STEM and non-STEM superstar effect is nonexistent across both courses of interest. Thus, the main effect in modern Greek comes from just the standard superstars that are categorized as STEM and non-STEM superstars.

Table 9 presents the same information but regarding long-term outcomes which follows a dynamic of just affecting the outcome variable of the same specialization. Specifically, being exposed to a non-STEM superstar decreases the student's probability of choosing the Classic track choice by 0.020 percentage points. Since track choice determines students' degrees at the university level the effect is proportional but inversely related for the case of STEM applications, as expected.

Once again, I would like to mention that since the definition of a STEM and non-STEM superstar is quite ad hoc, but I have played with some other definitions and obtained similar results. For example, I defined as a STEM superstar using all the grades of all the courses but with a weighted average allocating a higher weight to the stem courses. Instead of just using the stem courses to define it as I initially have done.

Finally, Table 10 presents the estimates of the female and male superstar effects in the final exam of semester 2 of grade 10 in the subject of algebra and modern Greek. The male superstar effect estimates are insignificant. Nevertheless, the female superstar estimates are significant in both courses! Hence, in average, being exposed to a female superstar decreases a student's algebra (modern Greek) score by 0.029 (0.050) standard deviations. Thus, the main conclusion of this table is that most of the negative superstar effect is being dragged by the female superstars. The effect being larger in the language subject is concurrent with the literature (Zimmerman, 2003). This time around the female coefficients are constant. Literature indicates that the proportion of females should have a positive effect on scholastic outcomes but our initial covariates show, once again, that this is not happening in practice (Lavy & Schlosser, 2011, Goulas et al., 2018).

Table 8: Estimates of the effect of a superstar in the final exam of Algebra and Modern Greek in grade 10 by STEM and non-STEM superstar

| | (1) #Standard Superstar | | (2) #STEM Superstar | | (3) #Non - STEM Superstar | |
|---|---|---|---|---|---|---|
| VARIABLES | Algebra | Modern Greek | Algebra | Modern Greek | Algebra | Modern Greek |
| Exposure to a superstar# | -0.020 | -0.047** | -0.017 | 0.014 | 0.009 | -0.043 |
| | (0.013) | (0.019) | (0.015) | (0.020) | (0.021) | (0.033) |
| Female | -0.105*** | 0.104*** | -0.104*** | 0.104*** | -0.104*** | 0.104*** |
| | (0.006) | (0.006) | (0.006) | (0.006) | (0.006) | (0.006) |
| Class size | 0.005** | 0.007** | 0.005** | 0.007** | 0.005** | 0.007** |
| | (0.002) | (0.003) | (0.002) | (0.003) | (0.002) | (0.003) |
| Proportion of females | 0.031 | 0.051 | 0.026 | 0.050 | 0.024 | 0.053 |
| | (0.039) | (0.052) | (0.039) | (0.052) | (0.039) | (0.051) |
| GPA (baseline exam) | 0.467*** | 0.255*** | 0.470*** | 0.256*** | 0.471*** | 0.256*** |
| | (0.008) | (0.006) | (0.008) | (0.006) | (0.008) | (0.006) |
| Age at grade 10 | -0.076*** | -0.055*** | -0.078*** | -0.056*** | -0.078*** | -0.056*** |
| | (0.009) | (0.010) | (0.009) | (0.010) | (0.009) | (0.010) |
| Morning students | -0.048*** | -0.065*** | -0.051*** | -0.066*** | -0.050*** | -0.065*** |
| | (0.012) | (0.012) | (0.012) | (0.012) | (0.012) | (0.012) |
| Avg. class GPA (baseline exam) | -0.244*** | -0.105*** | -0.248*** | -0.129*** | -0.252*** | -0.120*** |
| | (0.024) | (0.030) | (0.023) | (0.030) | (0.023) | (0.030) |
| Observations | 63,594 | 63,594 | 63,978 | 63,978 | 64,368 | 64,368 |
| R-squared | 0.603 | 0.619 | 0.607 | 0.625 | 0.614 | 0.627 |
| School by year FE | YES | YES | YES | YES | YES | YES |
| Control | YES | YES | YES | YES | YES | YES |
| Adj. R-squared | 0.597 | 0.613 | 0.601 | 0.619 | 0.607 | 0.621 |

Notes: Each column variable represents the grade achieved by the student in the final exam of grade 10 of the relevant course. Presents estimates from school by year fixed effects regression of short-term outcomes. Regressions in (1) use a standard superstar (obtaining a grade over 18/20 in all core courses) exposure, while (2) and (3) use a STEM and non-STEM superstar (obtaining a grade over 19/20 in specialized subjects) exposure, accordingly. Regressions control for school by year fixed effects; student-level controls; student predetermined characteristics; and class-by-year characteristics. Each estimate is generated from a different regression. *, ** and *** denote significance at the 10%, 5% and 1% level, respectively. Robust standard errors clustered at the school by year level are reported in parentheses. All grades presented have been standardized. Control indicates that prior performance in that specific course is being controlled for. Nevertheless, due to presentation purposes it is not showed.

Table 9: Estimates of the effect of a superstar in selected long-term outcome variables by STEM and non-STEM of superstar

| | (1) | | (2) | | (3) | |
|---|---|---|---|---|---|---|
| | #Standard Superstar | | #STEM superstar | | #Non - STEM Superstar | |
| VARIABLES | STEM track choice | Applied to STEM | STEM track choice | Applied to STEM | Non- STEM Track choice | Applied to STEM |
| Exposure to a #superstar | -0.001 | 0.001 | -0.002 | -0.001 | -0.020* | 0.020* |
| | (0.006) | (0.007) | (0.007) | (0.008) | (0.010) | (0.011) |
| Female | -0.361*** | -0.346*** | -0.360*** | -0.345*** | 0.359*** | -0.344*** |
| | (0.004) | (0.005) | (0.004) | (0.005) | (0.004) | (0.004) |
| Class size | -0.001 | -0.001 | -0.001 | -0.001 | 0.001 | -0.001 |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| Proportion of females | -0.003 | 0.012 | -0.004 | 0.013 | 0.003 | 0.012 |
| | (0.019) | (0.019) | (0.018) | (0.019) | (0.018) | (0.019) |
| GPA (baseline exam) | 0.103*** | 0.060*** | 0.103*** | 0.060*** | -0.105*** | 0.062*** |
| | (0.002) | (0.003) | (0.002) | (0.003) | (0.002) | (0.003) |
| Age at grade 10 | 0.023*** | 0.001 | 0.023*** | 0.001 | -0.023*** | 0.000 |
| | (0.007) | (0.009) | (0.007) | (0.009) | (0.007) | (0.009) |
| Morning students | 0.038*** | 0.015 | 0.038*** | 0.015 | -0.037*** | 0.015 |
| | (0.009) | (0.011) | (0.009) | (0.011) | (0.009) | (0.011) |
| Avg. class GPA (baseline exam) | -0.036*** | -0.029*** | -0.036*** | -0.029*** | 0.039*** | -0.031*** |
| | (0.011) | (0.011) | (0.011) | (0.011) | (0.010) | (0.011) |
| Observations | 63,944 | 55,389 | 64,331 | 55,689 | 64,725 | 55,989 |
| R-squared | 0.199 | 0.172 | 0.198 | 0.170 | 0.200 | 0.171 |
| School by year FE | YES | YES | YES | YES | YES | YES |
| Adj. R-squared | 0.186 | 0.156 | 0.184 | 0.154 | 0.187 | 0.155 |

Notes: Presents estimates from school by year fixed effects regression of selected long-term outcomes. Regressions in (1) use a standard superstar (obtaining a grade over 18/20 in all core courses) exposure, while (2) and (3) use a STEM and non-STEM superstar (obtaining a grade over 19/20 in specialized subjects) exposure, accordingly. Regressions control for school by year fixed effects; student-level controls; student predetermined characteristics; and class-by-year characteristics. Each estimate is generated from a different regression. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively. Robust standard errors clustered at the school by year level are reported in parentheses. All grades presented have been standardized.

Table 10: Estimates of exposure to a girl/boy superstar in the final exam of Algebra and Modern Greek in grade 10

| | (1) #Standard superstar | | (2) #Supergirl | | (3) #Superboy | |
|---|---|---|---|---|---|---|
| VARIABLES | Algebra final exam | Modern Greek final exam | Algebra final exam | Modern Greek final exam | Algebra final exam | Modern Greek final exam |
| Exposure to a #superstar | -0.019 | -0.047** | -0.029** | -0.050** | -0.006 | -0.023 |
| | (0.013) | (0.019) | (0.014) | (0.019) | (0.018) | (0.024) |
| Female | -0.106*** | 0.103*** | -0.108*** | 0.102*** | -0.102*** | 0.103*** |
| | (0.006) | (0.006) | (0.006) | (0.006) | (0.006) | (0.006) |
| Class size | 0.005** | 0.007** | 0.005** | 0.007** | 0.005** | 0.007** |
| | (0.002) | (0.003) | (0.002) | (0.003) | (0.002) | (0.003) |
| Proportion of females | 0.029 | 0.044 | 0.032 | 0.054 | 0.025 | 0.036 |
| | (0.039) | (0.052) | (0.039) | (0.052) | (0.039) | (0.052) |
| GPA (baseline exam) | 0.466*** | 0.254*** | 0.468*** | 0.255*** | 0.470*** | 0.255*** |
| | (0.008) | (0.006) | (0.008) | (0.006) | (0.008) | (0.006) |
| Avg. class GPA (baseline exam) | -0.244*** | -0.102*** | -0.242*** | -0.106*** | -0.248*** | -0.117*** |
| | (0.024) | (0.030) | (0.023) | (0.030) | (0.023) | (0.029) |
| Age at grade 10 | -0.076*** | -0.055*** | -0.077*** | -0.055*** | -0.077*** | -0.056*** |
| | (0.009) | (0.010) | (0.009) | (0.010) | (0.009) | (0.010) |
| Morning students | -0.048*** | -0.065*** | -0.049*** | -0.066*** | -0.050*** | -0.065*** |
| | (0.012) | (0.012) | (0.012) | (0.012) | (0.012) | (0.012) |
| Observations | 63,944 | 63,944 | 64,319 | 64,319 | 64,676 | 64,676 |
| R-squared | 0.604 | 0.619 | 0.609 | 0.623 | 0.612 | 0.627 |
| School by year FE | YES | YES | YES | YES | YES | YES |
| Control | YES | YES | YES | YES | YES | YES |
| Adj. R-squared | 0.597 | 0.613 | 0.602 | 0.617 | 0.606 | 0.621 |

Notes: Regressions in (1) use a standard superstar classroom exposure, while (2) and (3) use a supergirl and superboy classroom exposure, accordingly. Supergirl/superboy is defined as a female/male superstar. Regressions control for school by year fixed effects; student-level controls; student predetermined characteristics; and class-by-year characteristics. Each estimate is generated from a different regression. *, ** and *** denote significance at the 10%, 5% and 1% level, respectively. Robust standard errors clustered at the school by year level are reported in parentheses. Control indicates that prior performance in the baseline exam in that specific course is being controlled for. Nevertheless, due to presentation purposes it is not showed.

## 5.5 Superstar effect by type of superstar separately by gender

Section 5.3 presented some interesting dynamics of the non-STEM superstar and female superstar. Tables 11-13 will try to see if the effects are heterogenous across male and female students exposed to a particular superstar.

Firstly, Table 11 presents the estimates of the standard superstar effect separately by gender. I will do this by changing the comparison group consisting of the within-gender for classes with and without superstars. The following two table described later on in this section will use the same logic applied to the pertaining superstar. Contrary to what I initially found, the standard superstar does have an impact on algebra scores, but just on female student's algebra scores. Previously, I had also determined that there was a superstar effect on geometry scores. But this effect is being dragged by the superstar effect on male geometry scores. It is not significant for the case of women. The effect found initially for Modern Greek is for both genders. But it is important to note that the superstar effects are substantially larger in male students.

Secondly, Table 12 presents the STEM/non-STEM superstar effects separately by gender which are all insignificant! The intuition of this phenomena is not so straightforward. One has first to recall that in Table 8 the only effect consisted of non-STEM superstar on Modern Greek test scores. Thus, one can consider that the effect is evenly distributed across genders in such a way that when the analysis is conditioned separately by gender the effect becomes insignificant.

In comparison, Table 13 presents the female and male superstar effects separately by gender and comes with a straightforward intuition. There is no male superstar effect at all! In the literature, men usually disturb class functioning which usually worsens scholastic outcomes. But these results imply that male superstars do not cause any type of externality in classrooms. The female superstar effect in algebra scores only occurs when the student is female. On the other hand, the female superstar effect on modern Greek scores happen across both genders. Hence, exceptionally high achieving girls do create negative externalities in a classroom environment independently of gender.

Table 11: Estimates of the effect of a superstar separately by gender

| | ------- Females -------- | | | ------- Males -------- | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| | Final exam | Final exam | Final exam | Final exam | Final exam | Final exam |
| VARIABLES | Algebra | Modern Greek | Geometry | Algebra | Modern Greek | Geometry |
| Exposure to a superstar | -0.031** | -0.037* | -0.020 | -0.009 | -0.062*** | -0.044** |
| | (0.015) | (0.020) | (0.016) | (0.016) | (0.022) | (0.018) |
| Class size | 0.005** | 0.009*** | 0.007*** | 0.005* | 0.004 | 0.006** |
| | (0.002) | (0.003) | (0.003) | (0.003) | (0.003) | (0.003) |
| Proportion of females | 0.004 | 0.022 | 0.026 | 0.088* | 0.106* | 0.130** |
| | (0.046) | (0.057) | (0.051) | (0.051) | (0.064) | (0.053) |
| GPA (baseline exam) | 0.476*** | 0.256*** | 0.464*** | 0.458*** | 0.257*** | 0.434*** |
| | (0.010) | (0.007) | (0.010) | (0.010) | (0.008) | (0.011) |
| Age at grade 10 | -0.069*** | -0.067*** | -0.015 | -0.086*** | -0.029* | -0.027** |
| | (0.012) | (0.012) | (0.012) | (0.015) | (0.016) | (0.014) |
| Morning students | -0.047*** | -0.071*** | -0.004 | -0.051*** | -0.045** | 0.017 |
| | (0.016) | (0.016) | (0.016) | (0.020) | (0.021) | (0.019) |
| Avg. class GPA (baseline exam) | -0.223*** | -0.117*** | -0.275*** | -0.281*** | -0.092*** | -0.287*** |
| | (0.028) | (0.033) | (0.028) | (0.029) | (0.035) | (0.031) |
| Observations | 35,368 | 35,368 | 35,368 | 28,226 | 28,226 | 28,226 |
| R-squared | 0.622 | 0.625 | 0.614 | 0.599 | 0.597 | 0.575 |
| School by year FE | YES | YES | YES | YES | YES | YES |
| Control | YES | YES | YES | YES | YES | NO |
| Adj. R^2 | 0.611 | 0.614 | 0.603 | 0.584 | 0.582 | 0.559 |

Notes: Each column variable represents the grade achieved by the student in the final exam of grade10 of the relevant course. Regressions control for school by year fixed effects; student-level controls; student predetermined characteristics; and class-by-year characteristics. Each estimate is generated from a different regression. *, ** and *** denote significance at the 10%, 5% and 1% level, respectively. Robust standard errors clustered at the school by year level are reported in parentheses. Control indicates that prior performance in the baseline exam in that specific course is being controlled for. Nevertheless, due to presentation purposes it is not showed.

Table 12: Estimates of exposure to a STEM/non-STEM superstar separately by gender

| | #STEM superstar | | | | #Non-STEM superstar | | | |
|---|---|---|---|---|---|---|---|---|
| | Female | Male | Female | Male | Female | Male | Female | Male |
| VARIABLES | (1) Algebra final exam in grade 10 | (2) Algebra final exam in grade 10 | (3) Modern Greek final exam in grade 10 | (4) Modern Greek final exam in grade 10 | (5) Algebra final exam in grade 10 | (6) Algebra final exam in grade 10 | (7) Modern Greek final exam in grade 10 | (8) Modern Greek final exam in grade 10 |
| Exposure to #superstar | -0.026 | -0.013 | 0.009 | 0.020 | 0.002 | 0.012 | -0.034 | -0.053 |
| | (0.018) | (0.018) | (0.021) | (0.025) | (0.023) | (0.027) | (0.036) | (0.037) |
| Class size | 0.005** | 0.005* | 0.009*** | 0.004 | 0.005* | 0.005* | 0.008*** | 0.004 |
| | (0.002) | (0.003) | (0.003) | (0.003) | (0.002) | (0.003) | (0.003) | (0.003) |
| Proportion of females | -0.002 | 0.085* | 0.019 | 0.106* | -0.004 | 0.082 | 0.019 | 0.114* |
| | (0.046) | (0.051) | (0.057) | (0.064) | (0.046) | (0.050) | (0.057) | (0.063) |
| GPA (baseline exam) | 0.480*** | 0.460*** | 0.256*** | 0.257*** | 0.482*** | 0.462*** | 0.256*** | 0.258*** |
| | (0.010) | (0.010) | (0.007) | (0.008) | (0.010) | (0.010) | (0.008) | (0.008) |
| Avg. class GPA (baseline exam) | -0.071*** | -0.088*** | -0.069*** | -0.029* | -0.071*** | -0.088*** | -0.069*** | -0.028* |
| | (0.012) | (0.015) | (0.012) | (0.016) | (0.012) | (0.015) | (0.012) | (0.016) |
| Age at grade 10 | -0.049*** | -0.055*** | -0.070*** | -0.047** | -0.049*** | -0.054*** | -0.070*** | -0.045** |
| | (0.016) | (0.020) | (0.016) | (0.021) | (0.016) | (0.020) | (0.016) | (0.021) |
| Morning students | -0.227*** | -0.282*** | -0.134*** | -0.127*** | -0.234*** | -0.285*** | -0.128*** | -0.113*** |
| | (0.027) | (0.028) | (0.033) | (0.035) | (0.027) | (0.028) | (0.032) | (0.035) |
| Observations | 35,721 | 28,257 | 35,721 | 28,257 | 35,861 | 28,507 | 35,861 | 28,507 |
| School by year FE | YES | YES | YES | YES | YES | YES | YES | YES |
| Control | YES | YES | YES | YES | YES | YES | YES | YES |
| Adj. R-squared | 0.617 | 0.585 | 0.620 | 0.586 | 0.622 | 0.594 | 0.621 | 0.591 |

Notes: Each column variable represents the grade achieved by the student in the final exam of grade 10 of the relevant course. A STEM and non-STEM superstar is defined by obtaining a grade over 19/20 in their corresponding specialized subjects. Regressions control for school by year fixed effects; student-level controls; student predetermined characteristics; and class-by-year characteristics. Each estimate is generated from a different regression. *, ** and *** denote significance at the 10%, 5% and 1% level, respectively. Robust standard errors clustered at the school by year level are reported in parentheses. Control indicates that prior performance in the baseline exam in that specific course is being controlled for. Nevertheless, due to presentation purposes it is not showed.

Table 13: Estimates of exposure to a girl/boy superstar in the final exam of Algebra and Modern Greek in grade 10 separately by genre

| | #Supergirl | | | | #Superboy | | | |
|---|---|---|---|---|---|---|---|---|
| | ------- Females -------- | | ------- Males -------- | | ------- Females -------- | | ------- Males -------- | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| VARIABLES | Algebra | Modern Greek | Algebra | Modern Greek | Algebra | Modern Greek | Algebra | Modern Greek |
| Exposure to #superstar | -0.035** | -0.049** | -0.023 | -0.054** | -0.013 | -0.002 | 0.002 | -0.045 |
| | (0.017) | (0.021) | (0.017) | (0.023) | (0.020) | (0.026) | (0.022) | (0.029) |
| Class size | 0.005** | 0.009*** | 0.005* | 0.004 | 0.005* | | 0.005* | 0.004 |
| | (0.002) | (0.003) | (0.003) | (0.003) | (0.002) | (0.003) | (0.003) | (0.003) |
| Proportion of females | 0.009 | 0.024 | 0.084* | 0.109* | -0.007 | 0.010 | 0.085* | 0.087 |
| | (0.046) | (0.057) | (0.050) | (0.064) | (0.046) | (0.057) | (0.051) | (0.064) |
| GPA (baseline exam) | 0.476*** | 0.255*** | 0.462*** | 0.256*** | 0.483*** | 0.256*** | 0.457*** | 0.256*** |
| | (0.010) | (0.007) | (0.010) | (0.008) | (0.010) | (0.007) | (0.010) | (0.008) |
| Avg. class GPA (baseline exam) | -0.224*** | -0.113*** | -0.273*** | -0.099*** | -0.228*** | -0.127*** | -0.283*** | -0.108*** |
| | (0.028) | (0.033) | (0.028) | (0.035) | (0.027) | (0.032) | (0.028) | (0.034) |
| Age at grade 10 | -0.070*** | -0.068*** | -0.088*** | -0.028* | -0.071*** | -0.069*** | -0.086*** | -0.029* |
| | (0.012) | (0.012) | (0.015) | (0.016) | (0.012) | (0.012) | (0.015) | (0.016) |
| Morning students | -0.047*** | -0.071*** | -0.054*** | -0.046** | -0.050*** | -0.070*** | -0.051*** | -0.045** |
| | (0.016) | (0.016) | (0.020) | (0.021) | (0.016) | (0.016) | (0.020) | (0.021) |
| Observations | 35,545 | 35,545 | 28,774 | 28,774 | 36,277 | 36,277 | 28,399 | 28,399 |
| School by year FE | YES | YES | YES | YES | YES | YES | YES | YES |
| Control | YES | YES | YES | YES | YES | YES | YES | YES |
| Adj. R-squared | 0.611 | 0.614 | 0.596 | 0.595 | 0.626 | 0.626 | 0.584 | 0.582 |

Notes: Each column variable represents the grade achieved by the student in the final exam of grade 10 of the relevant course. Regressions (1)- (4) use a supergirl classroom exposure, while (5) - (6) use a superboy classroom exposure. Supergirl/superboy is defined as a female/male superstar. Regressions control for school by year fixed effects; student-level controls; student predetermined characteristics; and class-by-year characteristics and are conditioned on gender. Each estimate is generated from a different regression. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively. Robust standard errors clustered at the school by year level are reported in parentheses. All grades presented have been standardized. Control indicates that prior performance in the baseline exam in that specific course is being controlled for. Nevertheless, due to presentation purposes it is not showed.

## 5.6 Heterogeneity exercises

In this section I will try to analyze the superstar effects separately by income of neighborhood, school quality and if the student is in a private or public school. For starters, Table 14 shows that students that live-in high-income neighborhoods are the only ones sensible to the superstar effect – students living in a high-income neighborhood national exam performance decreases by 0.031 standard deviations, in average, when exposed to a standard superstar. Most of the class size effects have been largely reduced and become insignificant, except for the national exam performance.

A long the same lines, Table 15 shows estimates of the superstar effect separately by school quality. One can observe that superstar effects are significant independent of school quality. Nevertheless, the magnitude of the superstar effects is larger in top ranking schools. For example, in average, being exposed to a superstar in a top-ranking school will decrease a student algebra score in grade 11 by 0.089 standard deviations compared to the 0.058 standard deviations in the case he is in a low-ranking school.

Lastly, Table 16 presents the estimates of the superstar effect conditioned by type of school. Regarding private schools, I arrive at the result that the superstar effect is huge for modern Greek i.e. 0.2 standard deviations decrease in Modern Greek test scores in the final exam of semester 2 in grade 10. There are no superstar effects in algebra and modern Greek. This may have to do to with the small sample size in this regression since the estimate is quite unreasonable. In any case, for public schools the effect is more realistic and in line to previous results i.e. a 0.044 (0.03) decrease in modern Greek (geometry) scores.

Table 14: Estimates of the effect of a superstar separately by income of neighborhood

| VARIABLES | (1) High Income: Algebra grade 10 final | (2) High Income: Algebra grade 11 | (3) High Income: National Exam | (4) Low Income: Algebra grade 10 final | (5) Low Income: Algebra grade 11 | (6) Low Income: National Exam |
|---|---|---|---|---|---|---|
| Exposure to a superstar | -0.011 | -0.071*** | -0.031* | -0.023 | -0.001 | -0.008 |
| | (0.019) | (0.024) | (0.017) | (0.019) | (0.021) | (0.015) |
| Female | -0.106*** | -0.088*** | 0.029*** | -0.131*** | -0.097*** | 0.002 |
| | (0.008) | (0.009) | (0.010) | (0.008) | (0.009) | (0.011) |
| Class size | 0.004 | 0.001 | 0.007*** | 0.005 | 0.001 | 0.007** |
| | (0.003) | (0.004) | (0.003) | (0.003) | (0.003) | (0.003) |
| Proportion of females | 0.079 | -0.007 | 0.022 | -0.005 | 0.047 | -0.002 |
| | (0.060) | (0.070) | (0.053) | (0.054) | (0.058) | (0.049) |
| GPA (baseline exam) | 0.474*** | 0.585*** | 0.766*** | 0.774*** | 0.778*** | 0.771*** |
| | (0.011) | (0.011) | (0.006) | (0.003) | (0.004) | (0.006) |
| Avg. class GPA (baseline exam) | -0.261*** | -0.258*** | -0.309*** | -0.227*** | -0.268*** | -0.293*** |
| | (0.034) | (0.039) | (0.031) | (0.032) | (0.042) | (0.033) |
| Age at grade 10 | -0.076*** | -0.042*** | -0.143*** | -0.084*** | -0.068*** | -0.137*** |
| | (0.014) | (0.013) | (0.021) | (0.011) | (0.012) | (0.017) |
| Morning students | -0.048** | -0.013 | -0.104*** | -0.047*** | -0.033** | -0.088*** |
| | (0.019) | (0.018) | (0.024) | (0.016) | (0.016) | (0.021) |
| Observations | 31,029 | 31,029 | 27,592 | 32,565 | 32,565 | 27,797 |
| R-squared | 0.601 | 0.596 | 0.594 | 0.579 | 0.579 | 0.584 |
| School by year FE | YES | YES | YES | YES | YES | YES |
| Control | YES | YES | NO | YES | YES | NO |
| Adj. R-squared | 0.595 | 0.590 | 0.586 | 0.572 | 0.572 | 0.576 |

Notes: Regressions control for school by year fixed effects; student-level controls; student predetermined characteristics; and class-by-year characteristics conditioning on income of the neighborhood of the student. Each estimate is generated from a different regression. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively. Robust standard errors clustered at the school by year level are reported in parentheses. All grades presented have been standardized. Control indicates that prior performance in the baseline exam in that specific course is being controlled for. Nevertheless, due to presentation purposes it is not showed.

Table 15: Estimates of the effect of a superstar separately by school quality

| VARIABLES | (1)<br>Top ranking:<br>Algebra exam 3 | (2)<br>Top ranking:<br>Algebra grade 11 | (3)<br>Top ranking:<br>National Exam | (4)<br>Low ranking:<br>Algebra exam 3 | (5)<br>Low ranking:<br>Algebra grade 11 | (6)<br>Low ranking:<br>National Exam |
|---|---|---|---|---|---|---|
| Exposure to a superstar | -0.081*** | -0.089*** | -0.060*** | -0.031 | -0.058** | -0.056*** |
| | (0.020) | (0.025) | (0.018) | (0.021) | (0.025) | (0.016) |
| Female | -0.120*** | -0.112*** | 0.013 | -0.095*** | -0.086*** | 0.026** |
| | (0.008) | (0.009) | (0.010) | (0.009) | (0.010) | (0.011) |
| Class size | 0.010*** | 0.005 | 0.009*** | 0.005* | -0.000 | 0.008*** |
| | (0.003) | (0.004) | (0.002) | (0.003) | (0.003) | (0.003) |
| Proportion of females | 0.009 | 0.025 | -0.043 | -0.029 | -0.082 | -0.061 |
| | (0.057) | (0.069) | (0.051) | (0.054) | (0.060) | (0.049) |
| GPA (baseline exam) | 0.459*** | 0.535*** | 0.705*** | 0.417*** | 0.532*** | 0.738*** |
| | (0.011) | (0.012) | (0.005) | (0.012) | (0.011) | (0.006) |
| Age at grade 10 | -0.075*** | -0.051*** | -0.154*** | -0.089*** | -0.058*** | -0.135*** |
| | (0.019) | (0.019) | (0.020) | (0.015) | (0.013) | (0.017) |
| Morning students | -0.050** | -0.029 | -0.107*** | -0.058*** | -0.031* | -0.096*** |
| | (0.022) | (0.023) | (0.024) | (0.020) | (0.018) | (0.021) |
| Observations | 27,344 | 27,344 | 27,344 | 28,045 | 28,045 | 28,045 |
| R-squared | 0.584 | 0.565 | 0.567 | 0.573 | 0.562 | 0.568 |
| School by year FE | YES | YES | YES | YES | YES | YES |
| Control | YES | YES | NO | YES | YES | NO |
| Adj. R-squared | 0.578 | 0.558 | 0.561 | 0.563 | 0.553 | 0.558 |

Notes: Regressions control for school by year fixed effects; student-level controls; student predetermined characteristics; and class-by-year characteristics conditioning on the school quality. School quality is constructed by the performance of students in the national exam. Each estimate is generated from a different regression. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively. Robust standard errors clustered at the school by year level are reported in parentheses. All grades presented have been standardized. Control indicates that prior performance in the baseline exam in that specific course is being controlled for. Nevertheless, due to presentation purposes it is not showed.

.

Table 16: Estimates of the effect of a superstar by type of school

| | ------- Private -------- | | | ------- Public -------- | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| | Final exam | Final exam | Final exam | Final exam | Final exam | Final exam |
| VARIABLES | Algebra | Modern Greek | Geometry | Algebra | Modern Greek | Geometry |
| Exposure to a superstar | -0.070 | -0.204*** | -0.075 | -0.020 | -0.044** | -0.030** |
| | (0.052) | (0.048) | (0.076) | (0.014) | (0.019) | (0.014) |
| Female | -0.104*** | 0.100*** | -0.182*** | -0.105*** | 0.104*** | -0.195*** |
| | (0.028) | (0.027) | (0.022) | (0.006) | (0.006) | (0.006) |
| Class size | -0.002 | -0.022** | -0.002 | 0.005** | 0.008*** | 0.007*** |
| | (0.011) | (0.009) | (0.015) | (0.002) | (0.003) | (0.002) |
| Proportion of females | 0.102 | 0.127 | -0.499 | 0.031 | 0.051 | 0.064 |
| | (0.273) | (0.325) | (0.447) | (0.040) | (0.052) | (0.043) |
| GPA (baseline exam) | 0.457*** | 0.253*** | 0.485*** | 0.467*** | 0.255*** | 0.446*** |
| | (0.061) | (0.026) | (0.054) | (0.008) | (0.006) | (0.008) |
| Age at grade 10 | -0.114 | -0.135 | -0.198** | -0.076*** | -0.055*** | -0.020** |
| | (0.179) | (0.101) | (0.089) | (0.009) | (0.010) | (0.009) |
| Morning students | -0.088 | -0.138 | -0.194* | -0.048*** | -0.065*** | 0.008 |
| | (0.190) | (0.105) | (0.101) | (0.012) | (0.013) | (0.012) |
| Avg. class GPA (baseline exam) | -0.143 | 0.042 | -0.221 | -0.246*** | -0.107*** | -0.281*** |
| | (0.127) | (0.150) | (0.151) | (0.024) | (0.031) | (0.025) |
| Observations | 1,804 | 1,804 | 1,804 | 61,790 | 61,790 | 61,790 |
| R-squared | 0.639 | 0.679 | 0.596 | 0.602 | 0.618 | 0.589 |
| School by year FE | YES | YES | YES | YES | YES | YES |
| Control | YES | YES | YES | YES | YES | YES |
| Adj. R^2 | 0.633 | 0.673 | 0.589 | 0.596 | 0.611 | 0.582 |

Notes: Regressions control for school by year fixed effects; student-level controls; student predetermined characteristics; and class-by-year characteristics conditioning on the student being in a private school or not. Each estimate is generated from a different regression. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively. Robust standard errors clustered at the school by year level are reported in parentheses. All grades presented have been standardized. Control indicates that prior performance in the baseline exam in that specific course is being controlled for. Nevertheless, due to presentation purposes it is not showed.

## 5.7 Extensions: interactions between gender and STEM and non-STEM supergirl/superboy

Table 17 presents the estimates for the female STEM superstar effect on algebra and modern Greek scores separately by gender. Obviously, since I am conditioning on gender I have taken the female variable out of the specification in equation (2). The female STEM superstar effect is also significant for algebra female student scores – if exposed and the student is female the grade will decrease, in average, by 0.044 standard deviations. Consistently to previous results, the proportion of females doesn't have any type of impact on females but this time around there exists a positive effect on men test scores in algebra and modern Greek. Specifically, in average, if the proportion of females increases by 0.01 it induces an increase of 0.00084 (0.00108) standard deviations in the algebra (modern Greek) test score in grade 10.

Furthermore, Table 18 presents the structure as the previous table but for long term outcomes. One can check that being exposed to a female STEM superstar does not have any type of impact at all across both genders. This would seem to indicate that female students don't form gender roles, preferences or at least a belief that women can also achieve success in STEM when observing a female STEM superstar. At the very least, if it does, it's in a way that is not statistically significant.

On the contrary, being exposed to a female non-STEM superstar impacts the three outcome variables positively when the student is female (Table 18). Specifically, if a girl is exposed to a female non-STEM superstar, then the student's probability of applying to a STEM degree increases by 0.048 percentage points. Sadly, it's important to recall that this effect in aggregate would not be considerable since there exists many more STEM than non-STEM superstars.

Finally, Table 20 presents the regression estimates to for the male STEM superstar effect on selected long-term outcomes. There is no superstar effect on selecting the STEM track choice or applying to a STEM degree. Unexpectedly, if a female student is exposed to a male STEM superstar, her probability of being admitted to a STEM degree increases by 0.016 percentage points.

In a way, the evidence gathered in this last table negates the initial hypothesis or the idea that when female students are exposed to a male STEM superstar, they will be discouraged from pursuing further studies in STEM. Since they could end up comparing themselves with the superstar and create the belief that women are not apt for STEM or at least reinforce the idea that STEM is male dominated and hence, that achieving success requires extra effort.

Table 17: Estimates of exposure to a STEM supergirl on Algebra final exam of grade 10 separately by gender.

| VARIABLES | Female Algebra | Male Algebra | Female Modern Greek | Male Modern Greek |
|---|---|---|---|---|
| Exposure to a STEM supergirl | -0.044* | -0.005 | -0.016 | -0.006 |
| | (0.023) | (0.025) | (0.028) | (0.034) |
| Class size | 0.005** | 0.005* | 0.009*** | 0.004 |
| | (0.002) | (0.003) | (0.003) | (0.003) |
| Proportion of females | 0.005 | 0.084* | 0.022 | 0.108* |
| | (0.047) | (0.051) | (0.057) | (0.063) |
| GPA (baseline exam) | 0.480*** | 0.462*** | 0.256*** | 0.259*** |
| | (0.010) | (0.010) | (0.007) | (0.008) |
| Avg. class GPA (baseline exam) | -0.071*** | -0.088*** | -0.069*** | -0.028* |
| | (0.012) | (0.015) | (0.012) | (0.016) |
| Age at grade 10 | -0.050*** | -0.054*** | -0.071*** | -0.046** |
| | (0.016) | (0.020) | (0.016) | (0.021) |
| Morning students | -0.227*** | -0.282*** | -0.129*** | -0.117*** |
| | (0.027) | (0.028) | (0.033) | (0.035) |
| Observations | 35,721 | 28,598 | 35,721 | 28,598 |
| R-squared | 0.628 | 0.610 | 0.631 | 0.609 |
| School by year FE | YES | YES | YES | YES |
| Control | YES | YES | YES | YES |
| Adj. R-squared | 0.617 | 0.596 | 0.620 | 0.595 |

Notes: Regressions control for school by year fixed effects; student-level controls; student predetermined characteristics; and class-by-year characteristics. Each estimate is generated from a different regression. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively. Robust standard errors clustered at the school by year level are reported in parentheses. All grades presented have been standardized. Control indicates that prior performance in the baseline exam in that specific course is being controlled for. Nevertheless, due to presentation purposes it is not showed.

Table 18: Estimates of the effect of a STEM supergirl in selected long-term outcome variables separately by gender

| | ------- Females -------- | | | ------- Males -------- | | |
|---|---|---|---|---|---|---|
| VARIABLES | (1)<br>STEM track choice | (2)<br>Admitted to STEM | (3)<br>Applied to STEM | (4)<br>STEM track choice | (5)<br>Admitted to STEM | (6)<br>Applied to STEM |
| Exposure to a STEM supergirl | 0.006 | -0.012 | -0.006 | 0.002 | -0.004 | 0.008 |
| | (0.012) | (0.007) | (0.013) | (0.011) | (0.012) | (0.011) |
| Class size | -0.002 | 0.001 | -0.002 | -0.001 | 0.001 | -0.000 |
| | (0.001) | (0.001) | (0.002) | (0.001) | (0.001) | (0.001) |
| Proportion of females | -0.018 | 0.016 | 0.025 | 0.000 | 0.011 | -0.005 |
| | (0.027) | (0.015) | (0.028) | (0.024) | (0.024) | (0.025) |
| GPA (baseline exam) | 0.131*** | 0.044*** | 0.079*** | 0.067*** | 0.079*** | 0.035*** |
| | (0.003) | (0.002) | (0.004) | (0.003) | (0.003) | (0.003) |
| Age at grade 10 | 0.033*** | -0.005 | -0.002 | 0.005 | -0.022 | 0.002 |
| | (0.009) | (0.006) | (0.011) | (0.011) | (0.013) | (0.013) |
| Morning students | 0.037*** | 0.002 | 0.011 | 0.033** | -0.015 | 0.015 |
| | (0.012) | (0.008) | (0.015) | (0.013) | (0.016) | (0.016) |
| Avg. class GPA (baseline exam) | -0.055*** | -0.021** | -0.051*** | -0.021 | -0.040*** | -0.009 |
| | (0.014) | (0.009) | (0.016) | (0.013) | (0.014) | (0.014) |
| Observations | 35,721 | 25,629 | 31,235 | 28,598 | 20,673 | 24,741 |
| R-squared | 0.116 | 0.066 | 0.085 | 0.094 | 0.100 | 0.079 |
| School by year FE | YES | YES | YES | YES | YES | YES |
| Adj. R-squared | 0.0896 | 0.0274 | 0.0546 | 0.0602 | 0.0537 | 0.0395 |

Notes: Each column variable represents the grade achieved by the student in the final exam of Year 10 of the relevant course. Regressions control for school by year fixed effects; student-level controls; student predetermined characteristics; and class-by-year characteristics. Each estimate is generated from a different regression. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively. Robust standard errors clustered at the school by year level are reported in parentheses. All grades presented have been standardized. Control indicates that prior performance in the baseline exam in that specific course is being controlled for. Nevertheless, due to presentation purposes it is not showed.

Table 19: Estimates of the effect of a non-STEM female superstar in selected long-term outcome variables separately by gender

| | ------- Females -------- | | | ------- Males -------- | | |
|---|---|---|---|---|---|---|
| VARIABLES | (1)<br>STEM track choice | (2)<br>Admitted to STEM | (3)<br>Applied to STEM | (4)<br>STEM track choice | (5)<br>Admitted to STEM | (6)<br>Applied to STEM |
| Exposure to a non-STEM supergirl | 0.040** | 0.018* | 0.048*** | 0.014 | 0.002 | 0.011 |
| | (0.016) | (0.010) | (0.018) | (0.015) | (0.018) | (0.016) |
| Class size | -0.002 | 0.001 | -0.002 | -0.001 | 0.001 | -0.000 |
| | (0.001) | (0.001) | (0.002) | (0.001) | (0.001) | (0.001) |
| Proportion of females | -0.016 | 0.013 | 0.021 | -0.002 | 0.010 | -0.005 |
| | (0.027) | (0.015) | (0.028) | (0.025) | (0.024) | (0.025) |
| GPA (baseline exam) | 0.134*** | 0.046*** | 0.082*** | 0.067*** | 0.079*** | 0.035*** |
| | (0.003) | (0.002) | (0.004) | (0.003) | (0.003) | (0.003) |
| Age at grade 10 | 0.033*** | -0.005 | -0.003 | 0.005 | -0.022 | 0.002 |
| | (0.009) | (0.006) | (0.011) | (0.011) | (0.013) | (0.013) |
| Morning students | 0.036*** | 0.002 | 0.010 | 0.033** | -0.015 | 0.015 |
| | (0.012) | (0.007) | (0.015) | (0.013) | (0.016) | (0.016) |
| Avg. class GPA (baseline exam) | -0.058*** | -0.026*** | -0.054*** | -0.022 | -0.041*** | -0.008 |
| | (0.014) | (0.009) | (0.015) | (0.013) | (0.014) | (0.014) |
| Observations | 35,861 | 25,713 | 31,320 | 28,598 | 20,673 | 24,741 |
| R-squared | 0.120 | 0.068 | 0.087 | 0.094 | 0.100 | 0.079 |
| School by year FE | YES | YES | YES | YES | YES | YES |
| Adj. R-squared | 0.0939 | 0.0293 | 0.0566 | 0.0602 | 0.0537 | 0.0395 |

Notes: Regressions control for school by year fixed effects; student-level controls; student predetermined characteristics; and class-by-year characteristics. Each estimate is generated from a different regression. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively. Robust standard errors clustered at the school by year level are reported in parentheses. All grades presented have been standardized. Control indicates that prior performance in the baseline exam in that specific course is being controlled for. Nevertheless, due to presentation purposes it is not showed.

Table 20: Estimates of the effect of a STEM superboy in selected long-term outcome variables separately by gender

| | ------- Females -------- | | | ------- Males -------- | | |
|---|---|---|---|---|---|---|
| VARIABLES | (1)<br>STEM track choice | (2)<br>Admitted to STEM | (3)<br>Applied to STEM | (4)<br>STEM track choice | (5)<br>Admitted to STEM | (6)<br>Applied to STEM |
| Exposure to a STEM superboy | -0.012 | 0.016** | 0.008 | -0.007 | -0.009 | -0.015 |
| | (0.012) | (0.007) | (0.013) | (0.011) | (0.012) | (0.012) |
| Class size | -0.002 | 0.001 | -0.002 | -0.001 | 0.001 | -0.000 |
| | (0.001) | (0.001) | (0.002) | (0.001) | (0.001) | (0.001) |
| Proportion of females | -0.017 | 0.017 | 0.025 | -0.001 | 0.004 | -0.006 |
| | (0.027) | (0.015) | (0.028) | (0.025) | (0.024) | (0.025) |
| GPA (baseline exam) | 0.133*** | 0.045*** | 0.082*** | 0.066*** | 0.076*** | 0.034*** |
| | (0.003) | (0.002) | (0.004) | (0.003) | (0.003) | (0.003) |
| Age at grade 10 | 0.033*** | -0.005 | -0.003 | 0.005 | -0.023* | 0.001 |
| | (0.009) | (0.006) | (0.011) | (0.011) | (0.013) | (0.013) |
| Morning students | 0.036*** | 0.002 | 0.010 | 0.033** | -0.015 | 0.015 |
| | (0.012) | (0.007) | (0.015) | (0.013) | (0.016) | (0.016) |
| Avg. class GPA (baseline exam) | -0.053*** | -0.027*** | -0.051*** | -0.019 | -0.035** | -0.004 |
| | (0.014) | (0.009) | (0.015) | (0.014) | (0.014) | (0.015) |
| Observations | 36,093 | 25,898 | 31,505 | 28,257 | 20,386 | 24,454 |
| R-squared | 0.119 | 0.068 | 0.087 | 0.093 | 0.097 | 0.080 |
| School by year FE | YES | YES | YES | YES | YES | YES |
| Adj. R^2 | 0.0933 | 0.0298 | 0.0561 | 0.0590 | 0.0503 | 0.0396 |

Notes: Regressions control for school by year fixed effects; student-level controls; student predetermined characteristics; and class-by-year characteristics. Each estimate is generated from a different regression. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively. Robust standard errors clustered at the school by year level are reported in parentheses. All grades presented have been standardized. Control indicates that prior performance in the baseline exam in that specific course is being controlled for. Nevertheless, due to presentation purposes it is not showed.

## 5.8 Superstar effect on high-achieving students

Table 21 presents the superstar effect on second-best students in the classroom. Intuitively, there is a change in the comparison group which will now consist of the top students in the classrooms without a superstar and the best students after the superstar in classrooms with a superstar. Clearly, the superstar effect on the second best is only significant in algebra, history and modern Greek test scores.

The effects are a bit smaller than our previous estimates but still negative – modern Greek test scores decrease by 0.11 standard deviations. As an example, the initial superstar effect on modern Greek scores in the final exam of grade was -0.047 but it has now been reduced to -0.11. This initial evidence paints the case for the existence of perverse effect of heterogenous abilities in effort does hold in Greek high schools since the second-best student is the closest in ability to the superstar and the effect has been decreased. It is also important that the superstar effect in geometry has actually disappeared! Recall that in Table 5 the estimate of the superstar effect on geometry scores in the final exam of grade 10 was of -0.031 and significant.

Table 21: Estimates of the effect of a superstar on the second-best student of the class by core course

| VARIABLES | (1)<br>Final exam<br>Algebra | (2)<br>Final exam<br>Physics | (3)<br>Final exam<br>History | (4)<br>Final exam<br>Modern Greek | (5)<br>Final exam<br>Geometry |
|---|---|---|---|---|---|
| Exposure to a superstar | -0.08** | -0.01 | -0.05* | -0.11*** | -0.05 |
| | (0.032) | (0.028) | (0.029) | (0.031) | (0.034) |
| Female | -0.07*** | -0.09*** | 0.01 | 0.11*** | -0.08*** |
| | (0.024) | (0.025) | (0.023) | (0.026) | (0.027) |
| Class size | 0.00 | 0.00 | 0.00 | 0.01 | 0.01 |
| | (0.005) | (0.005) | (0.004) | (0.005) | (0.005) |
| Proportion of females | -0.05 | 0.01 | -0.10 | 0.04 | -0.02 |
| | (0.099) | (0.097) | (0.086) | (0.086) | (0.103) |
| GPA (baseline exam) | 0.59*** | 0.64*** | 0.43*** | 0.35*** | 0.59*** |
| | (0.075) | (0.062) | (0.061) | (0.055) | (0.073) |
| Age at grade 10 | -0.33*** | -0.14 | 0.07 | -0.05 | -0.09 |
| | (0.110) | (0.099) | (0.075) | (0.084) | (0.095) |
| Morning students | -0.29** | -0.14 | 0.02 | -0.07 | -0.05 |
| | (0.118) | (0.105) | (0.083) | (0.092) | (0.102) |
| Avg. class GPA (baseline exam) | -0.09 | -0.09* | -0.02 | -0.01 | -0.09 |
| | (0.057) | (0.053) | (0.050) | (0.055) | (0.057) |
| Observations | 3,338 | 3,338 | 3,338 | 3,338 | 3,338 |
| R-squared | 0.467 | 0.472 | 0.509 | 0.529 | 0.459 |
| School by year FE | YES | YES | YES | YES | YES |
| Control | YES | YES | YES | YES | YES |
| Adj. R-squared | 0.230 | 0.237 | 0.291 | 0.320 | 0.219 |

Notes: Each column variable represents the grade achieved by the student in the final exam of grade 10 of the relevant course. Regressions control for school by year fixed effects; student-level controls; student predetermined characteristics; and class-by-year characteristics. Each estimate is generated from a different regression. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively. Robust standard errors clustered at the school by year level are reported in parentheses. All grades presented have been standardized. Control indicates that prior performance in the baseline exam in that specific course is being controlled for. Nevertheless, due to presentation purposes it is not showed.

Similarly, Table 22 presents the estimates for the superstar effect across the top 5 students in the classroom. It uses the specification described in equation (3) that adds a top 5 identifier dummy variable and its interaction with the identifier for when classrooms have been exposed to a superstar (classsuper).

The superstar effect for the average student is significant for algebra, history, modern Greek and geometry. With modern Greek suffering the most in magnitude. It is important to notice that when I control for a high achieving group of students the superstar effects in algebra and history which were originally inexistent become significant and in a credible range of most superstar effect I have encountered till now!

The superstar effect in a top 5 student is distinguishable from the effect on an average student only on history and geometry. Recall that in our initial results (Table 5), just modern Greek and geometry presented a superstar effect. The superstar effect on a top 5 students geometry test score is a decrease of 0.024 standard deviations. In case for history, it is a small decrease of just 0.003 standard deviations compared to the non-top 5 decrease of 0.033 standard deviations. Clearly, this is evidence for the existence of perverse effect of heterogenous abilities in effort. Nevertheless, there doesn’t seem to be perverse effect of heterogenous abilities in effort in algebra and modern Greek.

Consistent with what is expected the estimate of a top 5 student is positive and have a considerable magnitude. This magnitude is especially large in the STEM core courses. In average, a top 5 students’ score in physics (modern Greek) is 0.177 (0.027) standard deviations above the grade of a non-top 5 student.

Table 22: Estimates of the effect of a superstar on the top 5 students of the class

| VARIABLES | (1) Final exam Algebra | (2) Final exam Physics | (3) Final exam History | (4) Final exam Modern Greek | (5) Final exam Geometry |
|---|---|---|---|---|---|
| Exposure to a superstar | -0.033** | -0.005 | -0.036** | -0.053*** | -0.046*** |
| | (0.014) | (0.014) | (0.016) | (0.020) | (0.015) |
| classsuper x top5 | 0.008 | 0.020 | 0.033*** | 0.015 | 0.022* |
| | (0.013) | (0.013) | (0.013) | (0.012) | (0.013) |
| Top 5 | 0.165*** | 0.177*** | 0.138*** | 0.027*** | 0.146*** |
| | (0.011) | (0.010) | (0.010) | (0.008) | (0.010) |
| Female | -0.105*** | -0.174*** | -0.042*** | 0.104*** | -0.195*** |
| | (0.006) | (0.006) | (0.006) | (0.006) | (0.006) |
| Class size | 0.007*** | 0.008*** | 0.008*** | 0.008*** | 0.009*** |
| | (0.002) | (0.002) | (0.002) | (0.003) | (0.002) |
| Proportion of females | 0.029 | 0.012 | -0.031 | 0.050 | 0.055 |
| | (0.039) | (0.039) | (0.044) | (0.052) | (0.043) |
| GPA (baseline exam) | 0.410*** | 0.479*** | 0.327*** | 0.245*** | 0.398*** |
| | (0.008) | (0.008) | (0.008) | (0.007) | (0.009) |
| Age at grade 10 | -0.077*** | -0.034*** | -0.001 | -0.055*** | -0.022** |
| | (0.009) | (0.009) | (0.010) | (0.010) | (0.009) |
| Morning students | -0.048*** | -0.010 | -0.003 | -0.065*** | 0.006 |
| | (0.012) | (0.012) | (0.013) | (0.012) | (0.012) |
| Avg. class GPA (baseline exam) | -0.178*** | -0.163*** | -0.129*** | -0.092*** | -0.217*** |
| | (0.024) | (0.023) | (0.026) | (0.030) | (0.025) |
| Observations | 63,594 | 63,594 | 63,594 | 63,594 | 63,594 |
| R-squared | 0.606 | 0.625 | 0.564 | 0.619 | 0.591 |
| School by year FE | YES | YES | YES | YES | YES |
| Control | YES | YES | YES | YES | YES |
| Adj. R-squared | 0.599 | 0.619 | 0.557 | 0.613 | 0.584 |

Notes: Each column variable represents the grade achieved by the student in the final exam of grade 10 of the relevant course Regressions control for school by year fixed effects; student-level controls; student predetermined characteristics; and class-by-year characteristics. Each estimate is generated from a different regression. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively. Robust standard errors clustered at the school by year level are reported in parentheses. All grades presented have been standardized. Control indicates that prior performance in the baseline exam in that specific course is being controlled for. Nevertheless, due to presentation purposes it is not showed.

Similarly, Table 23 presents the estimates for the superstar effect across the top 3 students in the classroom. The superstar effect for the average student is significant for the same courses as in the previous exercise. Modern Greek superstar effect still has the biggest magnitude.

This time around, the superstar effect in a top 3 student is distinguishable from the effect on an average student only in geometry. The superstar effect on a top 3 student regarding the geometry test score is associated with a decrease of 0.13 standard deviations. In the case of history, the difference in the superstar effect on a non-top3 and top 3 student is not significantly distinguishable from zero. This is not quite consistent with the existence of perverse effect of heterogenous abilities in effort since I am just considering the top 3 students instead of top 5. More precisely, the ability gap has narrowed and thus require that the coefficient of the interaction term to be positive and of a considerable magnitude as to decrease the superstar effect for this to hold.

Table 23: Estimates of the effect of a superstar on the top 3 students of the class

| VARIABLES | (1) Final exam Algebra | (2) Final exam Physics | (3) Final exam History | (4) Final exam Modern Greek | (5) Final exam Geometry |
|---|---|---|---|---|---|
| Exposure to a superstar | -0.029** | 0.001 | -0.028* | -0.048** | -0.042*** |
| | (0.014) | (0.014) | (0.015) | (0.019) | (0.015) |
| classsuper x top3 | 0.012 | 0.015 | 0.021 | -0.003 | 0.029* |
| | (0.014) | (0.014) | (0.014) | (0.014) | (0.015) |
| Top 3 | 0.164*** | 0.186*** | 0.147*** | 0.041*** | 0.150*** |
| | (0.010) | (0.010) | (0.010) | (0.008) | (0.011) |
| Female | -0.105*** | -0.175*** | -0.043*** | 0.104*** | -0.195*** |
| | (0.006) | (0.006) | (0.006) | (0.006) | (0.006) |
| Class size | 0.006*** | 0.007*** | 0.007*** | 0.007*** | 0.008*** |
| | (0.002) | (0.002) | (0.002) | (0.003) | (0.002) |
| Proportion of females | 0.030 | 0.013 | -0.031 | 0.050 | 0.056 |
| | (0.039) | (0.039) | (0.044) | (0.052) | (0.043) |
| GPA (baseline exam) | 0.426*** | 0.494*** | 0.341*** | 0.246*** | 0.411*** |
| | (0.008) | (0.008) | (0.008) | (0.006) | (0.008) |
| Age at grade 10 | -0.078*** | -0.035*** | -0.002 | -0.056*** | -0.022** |
| | (0.009) | (0.009) | (0.010) | (0.010) | (0.009) |
| Morning students | -0.049*** | -0.011 | -0.004 | -0.066*** | 0.005 |
| | (0.012) | (0.012) | (0.013) | (0.012) | (0.012) |
| Avg. class GPA (baseline exam) | -0.197*** | -0.182*** | -0.144*** | -0.093*** | -0.234*** |
| | (0.024) | (0.023) | (0.026) | (0.030) | (0.025) |
| Observations | 63,594 | 63,594 | 63,594 | 63,594 | 63,594 |
| R-squared | 0.606 | 0.625 | 0.564 | 0.619 | 0.591 |
| School by year FE | YES | YES | YES | YES | YES |
| Control | YES | YES | YES | YES | YES |
| Adj. R-squared | 0.599 | 0.619 | 0.557 | 0.613 | 0.584 |

Notes: Each column variable represents the grade achieved by the student in the final exam of grade 10 of the relevant course Regressions control for school by year fixed effects; student-level controls; student predetermined characteristics; and class-by-year characteristics. Each estimate is generated from a different regression. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively. Robust standard errors clustered at the school by year level are reported in parentheses. All grades presented have been standardized. Control indicates that prior performance in the baseline exam in that specific course is being controlled for. Nevertheless, due to presentation purposes it is not showed.

## 5.9 Further robustness exercises

For illustration, in Table 24 regressions, I defined a superstar as a student in the top 1% percentile at the national level for each cohort based on the GPA performance in the baseline exam (Table 24). I find that the modern Greek (geometry) scores are reduced by 0.036 (0.036) standard deviations when a student is exposed to a superstar, consistent with Table 5. Interestingly, I find that the coefficient for physics is positive and significant. In average, being exposed to a superstar increases physics scores by 0.03 standard deviations. In Table 5 the estimate was positive but insignificant.

In addition, when I define the superstar as being one of the top 100 students nationwide based on their GPA performance in the baseline exam there exists a negative superstar effect in algebra, modern Greek and geometry and a positive effect in the case of physics (Table 25). Specifically, I find that when a student is exposed to one of the top 100 students nationwide, algebra (modern Greek) scores are reduced by 0.018 (0.037) standard deviations in average. Once again, I find that the coefficient for physics is positive and significant.

Finally, it is important to notice that the total number of superstars characterized by being in the top 1 percentile and the top 100 best students nationwide do not vary much from the standard quantity (1 097). With a difference of 211 and 197 less superstars with respect to the baseline definition, accordingly.

Table 24: Estimates of the effect of a superstar (top percentile nationwide) by performance in the final exam of the core course in grade 10

| VARIABLES | (1)<br>Final exam<br>Algebra | (2)<br>Final exam<br>Physics | (3)<br>Final exam<br>History | (4)<br>Final exam<br>Modern Greek | (5)<br>Final exam<br>Geometry |
|---|---|---|---|---|---|
| Exposure to a superstar | -0.018 | 0.030** | 0.004 | -0.036* | -0.036** |
| | (0.014) | (0.013) | (0.014) | (0.020) | (0.015) |
| Female | -0.105*** | -0.174*** | -0.042*** | 0.104*** | -0.194*** |
| | (0.006) | (0.006) | (0.006) | (0.006) | (0.006) |
| Class size | 0.005** | 0.005** | 0.005** | 0.007** | 0.007*** |
| | (0.002) | (0.002) | (0.002) | (0.003) | (0.002) |
| Proportion of females | 0.028 | 0.011 | -0.030 | 0.051 | 0.057 |
| | (0.039) | (0.039) | (0.044) | (0.051) | (0.043) |
| GPA (baseline exam) | 0.468*** | 0.539*** | 0.381*** | 0.256*** | 0.449*** |
| | (0.008) | (0.007) | (0.007) | (0.006) | (0.008) |
| Age at grade 10 | -0.078*** | -0.033*** | -0.001 | -0.056*** | -0.021** |
| | (0.009) | (0.009) | (0.010) | (0.010) | (0.009) |
| Morning students | -0.049*** | -0.009 | -0.003 | -0.066*** | 0.007 |
| | (0.012) | (0.012) | (0.013) | (0.012) | (0.012) |
| Avg. class GPA (baseline exam) | -0.247*** | -0.242*** | -0.197*** | -0.111*** | -0.279*** |
| | (0.023) | (0.023) | (0.025) | (0.030) | (0.024) |
| Observations | 63,805 | 63,805 | 63,805 | 63,805 | 63,805 |
| R-squared | 0.605 | 0.624 | 0.564 | 0.621 | 0.591 |
| School by year FE | YES | YES | YES | YES | YES |
| Control | YES | YES | YES | YES | YES |
| Adj. R-squared | 0.599 | 0.618 | 0.557 | 0.615 | 0.584 |

Notes: Each column variable represents the grade achieved by the student in the final exam of grade 10 of the relevant course. Regressions control for school by year fixed effects; student-level controls; student predetermined characteristics; and class-by-year characteristics. Each estimate is generated from a different regression. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively. Robust standard errors clustered at the school by year level are reported in parentheses. All grades presented have been standardized. Control indicates that prior performance in the baseline exam in that specific course is being controlled for. Nevertheless, due to presentation purposes it is not showed.

Table 25: Estimates of the effect of a superstar (top 100 student nationwide) by performance in the final exam of the core course in grade 10

| VARIABLES | (1) Final exam Algebra | (2) Final exam Physics | (3) Final exam History | (4) Final exam Modern Greek | (5) Final exam Geometry |
|---|---|---|---|---|---|
| Exposure to a superstar | -0.024* | 0.023* | 0.002 | -0.037** | -0.041*** |
| | (0.014) | (0.013) | (0.015) | (0.019) | (0.015) |
| Female | -0.105*** | -0.174*** | -0.042*** | 0.104*** | -0.194*** |
| | (0.006) | (0.006) | (0.006) | (0.006) | (0.006) |
| Class size | 0.005** | 0.005** | 0.005** | 0.007** | 0.007*** |
| | (0.002) | (0.002) | (0.002) | (0.003) | (0.002) |
| Proportion of females | 0.029 | 0.011 | -0.030 | 0.051 | 0.058 |
| | (0.039) | (0.039) | (0.044) | (0.051) | (0.043) |
| GPA (baseline exam) | 0.468*** | 0.539*** | 0.381*** | 0.256*** | 0.449*** |
| | (0.008) | (0.007) | (0.007) | (0.006) | (0.008) |
| Age at grade 10 | -0.078*** | -0.033*** | -0.000 | -0.056*** | -0.021** |
| | (0.009) | (0.009) | (0.010) | (0.010) | (0.009) |
| Morning students | -0.049*** | -0.009 | -0.003 | -0.066*** | 0.008 |
| | (0.012) | (0.012) | (0.013) | (0.012) | (0.012) |
| Avg. class GPA (baseline exam) | -0.244*** | -0.240*** | -0.196*** | -0.110*** | -0.276*** |
| | (0.023) | (0.023) | (0.026) | (0.030) | (0.024) |
| Observations | 63,791 | 63,791 | 63,791 | 63,791 | 63,791 |
| R-squared | 0.605 | 0.624 | 0.564 | 0.621 | 0.591 |
| School by year FE | YES | YES | YES | YES | YES |
| Control | YES | YES | YES | YES | YES |
| Adj. R-squared | 0.598 | 0.617 | 0.557 | 0.615 | 0.584 |

Notes: Each column variable represents the grade achieved by the student in the final exam of grade 10 of the relevant course. Regressions control for school by year fixed effects; student-level controls; student predetermined characteristics; and class-by-year characteristics. Each estimate is generated from a different regression. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively. Robust standard errors clustered at the school by year level are reported in parentheses. All grades presented have been standardized. Control indicates that prior performance in the baseline exam in that specific course is being controlled for. Nevertheless, due to presentation purposes it is not showed.

# Chapter 6

# Conclusion

## 6.1 Primary results

In this thesis, I show evidence of a moderate superstar effect that brings test scores down in the following courses: modern Greek, geometry and with less evidence, algebra and history. There is also some evidence, although inconclusive, that there is a positive superstar effect in the short run (end of grade 10) in physics test scores which later becomes negative in grade 11.

When analyzing the superstar effect against top performing students in the classroom, especially against the second-best, I found that the effect is reduced. Providing some evidence that the hypothesis of perverse effect of heterogenous abilities in effort does hold in the Greek setting.

Even further, the superstar effect seemed to also have an impact on longer term scholastic measures such as the general performance in grade 11 and the national exam. All other variables that where related to students' preferences, gender role formations and university outcomes where not significant. They included: STEM track choice in grade 11, STEM application, STEM admission and university admission rates.

Secondly, I found that these results are very sensible to gender. For illustration purposes, recall Table 13 that presents female and male superstar effects separately by gender in modern Greek and math test scores in the final exam of grade 10. The male superstar effect was nonexistent. Meanwhile, the female superstar effect decreased girls' tests scores in algebra and modern Greek. In the case of boys, it was just significant in one of the subjects, namely- modern Greek. Interestingly, the effect on boys had a considerable magnitude which by the literature they are less sensible to peer effects (Fischer, 2017). It was also further evidence of perverse effect of heterogenous abilities in effort since men usually have a lower score in modern Greek.

Throughout the gender heterogeneity exercises, the superstar effects are predominantly dragged by the female superstar effect in the algebra course. In contrast, in geometry, the effect is being dragged by the male superstar effect. In general, male superstars don't seem to provide a negative externality in high school classrooms. Conversely, female superstars do create a negative externality in the education sector. This is an interesting fact since woman are usually not thought of creating negative externalities in a classroom environment since men are usually thought of being more disrupting affecting day to day teaching activities resulting in worse academic outcomes (Lavy & Schlosser, 2011). I also found that there are no STEM and non-STEM superstar effects in the short run (Table 11).

Thirdly, in section 5.6. I presented evidence that girls exposed to a female STEM superstar don't generate a belief that woman can achieve success in STEM since there was no female STEM superstar effect across both genders on long term outcomes such as track choice and stem applications. I also found that female non-STEM superstar increases a woman's probability of being admitted to a STEM degree as well as choosing a STEM track choice by approximately 0.02 and 0.05 percentage points.

Lastly, I found that students that live-in high-income neighborhoods are the only ones sensible to the superstar effect and that superstar effects are significant independently of school quality. Nevertheless, it is important to notice that the magnitude of the superstar effects is larger in top ranking schools.

## 6.2 Policy directives

Since superstar peer effects seem to have a negative effect on scholastic outcomes and there is some evidence that this is reduced in the top students in the classrooms a natural argument is to group students into separate classes by prior achievement. A policy known as tracking.

Nevertheless, I cannot recommend this initial conclusion without considering other type of peer effects. For example, tracking is usually criticized since there is a large literature that shows that peer effects of high achievers (who don't necessarily have a huge difference in ability i.e. they use weaker definitions than the ones utilized in this thesis) benefit other high achievers as well as low achievers (Busso, M., & Frisancho, V., 2021; Modena et al., 2022; Cools et al., 2019). Thus, if this is the case, this policy would just be making matters worse by creating a feedback loop that magnifies inequality in student scholastic outcomes since the students that need the most help cannot benefit from those positive student interactions. Nevertheless, tracking could allow teachers to focalize their effort and teaching level based on the student's specific needs (Duflo, E., Dupas, P., & Kremer, M., 2011).

As mentioned before, the endogenous quality of the subject of study appears to make the effect of policy changes very difficult to predict, complicating the application of fine-tuned peer effects policies. Because of this, further research is still needed. Nevertheless, separating superstars into schools designed to increase their performance could potentially create a pareto-efficiently allocation for superstars, high and low achievers' scholastic outcomes. Since in theory the ability gap between superstars could potentially be small enough that there are no superstar effects between them. And if there is, my hypothesis is that the competition is not as fierce as to bring effort down and with luck, create a better school environment for everyone.

Finally, our results showed that in general, male and female STEM superstars do not seem to be influencing the beliefs of woman's gender roles and preferences. This is evidence that most gender differences in educational setting are mainly driven by differences in preferences that aren't updated so easily in a classroom environment. Nevertheless, a female student exposed to female non-STEM superstar seems to have better STEM long-term outcomes. Thus, if the policy is to inorganically modify STEM gender preferences imbalances one should insert a higher proportion of females in classes exposed to a female non-STEM superstar. This type of policy has applicable limitations since there are almost three times less STEM superstars than non-STEM superstars (approx. 70% of non-STEM superstars are females).

## 6.3 Further research

In my opinion, the most important point for future research would be the study of teacher and superstar effects since there is some evidence that peer effects work indirectly by influencing teacher effort and choice of target teaching level (Duflo et al., 2011). Will teachers focalize their effort and teaching level based on the superstar or in the worst students of the classroom ("the rotten apples")? Or will a middle point predominate?

Furthermore, teachers' gender role attitudes seem to influence students school attendance, performance in admission exams and student's choice of university field of study (Lavy & Megalokonomou, 2019). Without a doubt, I expect that interacting teachers gender roles attitudes with STEM and non-STEM superstars separately by gender will bring about interesting findings. In addition, controlling for teaching experience and quality will alleviate biases in most methodologies (Burke & Sass, 2013).

Finally, it would also seem of interest to study the effects of superstars when they are joined together with high achieving students in a school that groups students into levels of performance and contrast it with non-tracking school results. In general, this will be a titanic task since it will be quite difficult to compare tracking schools with non-tracking schools since they probably are quite different creating selection bias issues that are difficult to alleviate (Duflo et al., 2011; Pischke & Manning, 2006). In addition, as mentioned in the policy directives it will also be interesting to check if there exist peer effects between superstar. These two areas of research will allow a better insight on how superstars can be pareto-efficiently allocated based on the peer-group externalities and complementarity of students and teachers.

# References


Abadie, A., S. Athey, G. W. Imbens, and J. Wooldridge (2017). When should you adjust standard errors for clustering? Technical report, National Bureau of Economic Research.

Bergmann, B. R. (1974). Occupational segregation, wages and profits when employers discriminate by race or sex. *Eastern economic journal*, *1*(2), 103-110.

Betts, J. R., & Shkolnik, J. L. (2000). The effects of ability grouping on student achievement and resource allocation in secondary schools. *Economics of Education Review*, *19*(1), 1-15.

Breda, T., & Napp, C. (2019). Girls' comparative advantage in reading can largely explain the gender gap in math-related fields. *Proceedings of the National Academy of Sciences*, *116*(31), 15435-15440.

Brown, J. (2011). Quitters never win: The (adverse) incentive effects of competing with superstars. *Journal of Political Economy*, *119*(5), 982-1013.

Burke, M. A., & Sass, T. R. (2013). Classroom peer effects and student achievement. *Journal of Labor Economics*, *31*(1), 51-82.

Busso, M., & Frisancho, V. (2021). Good peers have asymmetric gendered effects on female educational outcomes: Experimental evidence from Mexico. *Journal of Economic Behavior & Organization*, *189*, 727-747.

Card, D., & Payne, A. A. (2021). High school choices and the gender gap in STEM. *Economic Inquiry*, *59*(1), 9-28.

Carrell, S. E., Fullerton, R. L., & West, J. E. (2009). Does your cohort matter? Measuring peer effects in college achievement. *Journal of Labor Economics*, *27*(3), 439-464.

Carrell, S. E., Sacerdote, B. I., & West, J. E. (2013). From natural variation to optimal policy? The importance of endogenous peer group formation. *Econometrica*, *81*(3), 855-882.

Casad, B. J., Franks, J. E., Garasky, C. E., Kittleman, M. M., Roesler, A. C., Hall, D. Y., & Petzel, Z. W. (2021). Gender inequality in academia: Problems and solutions for women faculty in STEM. *Journal of neuroscience research*, *99*(1), 13-23.

Ceci, S. J., Ginther, D. K., Kahn, S., & Williams, W. M. (2014). Women in academic science: A changing landscape. *Psychological science in the public interest*,

Cools, A., Fernández, R., & Patacchini, E. (2019). *Girls, boys, and high achievers* (No. w25763). National Bureau of Economic Research.

Dearden, L., Ferri, J., & Meghir, C. (2002). The effect of school quality on educational attainment and wages. *Review of economics and statistics*, *84*(1), 1-20.

Delaney, J., & Devereux, P. J. (2021). Gender and educational achievement: Stylized facts and causal evidence.

Denning, J. T., Murphy, R., & Weinhardt, F. (2018). Class rank and long-run outcomes. *The Review of Economics and Statistics*, 1-45.

Duflo, E., Dupas, P., & Kremer, M. (2011). Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya. *American economic review*, *101*(5), 1739-74.

Feld, J., & Zölitz, U. (2017). Understanding peer effects: On the nature, estimation, and channels of peer effects. *Journal of Labor Economics*, *35*(2), 387-428.

Fischer, S. (2017). The downside of good peers: How classroom composition differentially affects men's and women's STEM persistence. *Labour Economics*, *46*, 211-226.

Genakos, C., & Pagliero, M. (2012). Interim rank, risk taking, and performance in dynamic tournaments. *Journal of Political Economy*, *120*(4), 782-813.

Goulas, S., Griselda, S., & Megalokonomou, R. (2020). Comparative advantage and gender gap in STEM.

Goulas, S., & Megalokonomou, R. (2015). *Knowing who you are: The effect of feedback information on short and long term outcomes* (No. 2068-2018-1783).

Goulas, S., Megalokonomou, R., & Zhang, Y. (2018). Does the girl next door affect your academic outcomes and career choices?. *Available at SSRN 3286169*.

Gross, D. P. (2020). Creativity under fire: The effects of competition on creative production. *Review of Economics and Statistics*, *102*(3), 583-599.

Lavy, V., & Megalokonomou, R. (2019). *Persistency in teachers' grading bias and effects on longer-term outcomes: University admissions exams and choice of field of study* (No. w26021). National Bureau of Economic Research.

Lavy, V., Paserman, M. D., & Schlosser, A. (2008). Inside the black box of ability peer effects: Evidence from variation in low achievers in the classroom.

Lavy, V., & Schlosser, A. (2011). Mechanisms and impacts of gender peer effects at school. *American Economic Journal: Applied Economics*, *3*(2), 1-33

Levitt, S. D., List, J. A., Neckermann, S., & Sadoff, S. (2016). The behavioralist goes to school: Leveraging behavioral economics to improve educational performance. *American Economic Journal: Economic Policy*, *8*(4), 183-219.

Lin, Z., & Tang, X. (2022). Solving the Reflection Problem in Social Interaction Models with Endogeneity. *Available at SSRN 3862178*.

Lu, F., & Anderson, M. L. (2015). Peer effects in microenvironments: The benefits of homogeneous classroom groups. *Journal of Labor Economics*, *33*(1), 91-122.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The review of economic studies*, *60*(3), 531-542.

Modena, F., Rettore, E., & Tanzi, G. M. (2022). Asymmetries in the gender effect of high-performing peers: Evidence from tertiary education. *Labour Economics*, *78*, 102225.

Mouganie, P., & Wang, Y. (2020). High-performing peers and female stem choices in school. *Journal of Labor Economics*, *38*(3), 805-841.

Murphy, R., & Weinhardt, F. (2020). Top of the class: The importance of ordinal rank. *The Review of Economic Studies*, *87*(6), 2777-2826.

Niederle, M., & Vesterlund, L. (2007). Do women shy away from competition? Do men compete too much?. *The quarterly journal of economics*, *122*(3), 1067-1101.

Pischke, J. S., & Manning, A. (2006). Comprehensive versus selective schooling in England in Wales: What do we know?.

Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American economic review*, *94*(2), 247-252.

Tincani, M. (2017). Heterogeneous peer effects and rank concerns: Theory and evidence.

Zimmerman, D. J. (2003). Peer effects in academic outcomes: Evidence from a natural experiment. *Review of Economics and statistics*, *85*(1), 9-23.