Does the girl next door affect your cognitive outcomes?

Sofoklis Goulas
Department of Economics
University of North Carolina at Chapel Hill
goulas@unc.edu

Rigissa Megalokonomou
Department of Economics
University of Queensland
r.megalokonomou@uq.edu.au

Abstract

Peer effects are potentially important for optimally organising schools and neighbourhoods. In this paper we estimate how the gender of classmates and neighbours affect students' academic achievement, their decision to enrol in college and the choice of university study. We exploit a special institutional setting, in which schools are very close to each other, allowing for students from different schools to interact. Given that students are assigned to schools based on proximity to school, we define as neighbours all same-cohort peer who attend any other school within a mile from ones school. We propose a new methodology in mitigating reflection and endogeneity issues in identifying social interactions. We also exploit within-school and -neighbourhood variation in the proportion of females across consecutive cohorts. We investigate utility spillovers from the educational choices of students in consecutive cohorts. Spatial variation allows us to identify social interactions in groups of various sizes. Using data for the universe of students in Greece between 2004 and 2009 we find that a higher share of females improves females and males academic performance, affects various scholastic outcomes related to the university admission exams and affects the choice of university study for both genders. We also find positive and significant externalities in the decision to enrol in college and the choice of university study among peers who belong to the same social group.

Keywords: college enrollment, social interactions, geography, reflection problem
JEL Classification: I26, J24

1 Introduction

In the recent years the literature on the role of social interactions in economic behavior has expanded rapidly. This doesn't come as surprise when one thinks the importance of those effects in every day decision-making. The basis of decision-making though in almost every context is information. Humans are social beings and we naturally collect information through social interactions in order to inform our goals and choices. This is even more pronounced among adolescents. In developmental science, it has been widely argued that adolescents and young adults regularly mimic the choices and behavior of role models in their environment (Bell (1970)).

Brock and Durlauf (2001) define social interactions as the idea that an individual's marginal utility with respect to other individuals' choices in his reference group is positive. The desire to conform induces prevalent patterns of behavior even among agents with heterogeneous tastes over externalities from other individuals' choices (Bernheim (1994)). Social interactions within a reference group have been shown to affect students' achievement. However, there is little evidence on the effect of social interactions on the decisions of college enrollment and academic mobility. Moreover, social interactions can explain variation in choices across groups with similar characteristics. For example, Schelling (1973) provide early evidence of social interactions in binary choice in a profusion of contexts such as driving style and athletic play. Intuitively, conformity causes social interactions to be interconnected with neighborhood effects. Physical proximity amplifies the interplay of utility spillovers from other agents' choices and the combined effect becomes area specific. In an educational context, Garner and Raudenbush (1991) provide evidence of a positive relation between neighborhood quality and educational attainment.

There is evidence that peers' decision affect scholastic performance in elementary, middle and high school but also during college. Hoxby (2000) examines the effect of social interaction in grade school and finds that students who were randomly assigned to classes with students who have high reading scores relative to the school and grade, received higher reading scores. Hanushek et al. (2003) find that peer achievement has a positive effect on achievement growth. In particular, 0.1 standard deviation increase in peer average achievement leads to a 0.02 increase in student's performance. Zimmerman (2003) examines the effect of social interaction using freshmens' SAT score. He finds strong positive social interaction effects among roommates at almost all parts of the ability distribution. Cipollone and Alfonso (2007) find strong social interactions inter alia the decision to stay longer in school. When men were exempted from the compulsory military services -due to an earthquake- and stayed longer in school, the graduation rates of young women in the affected areas rose by about 2 percentage points. Fletcher (2006) using survey data, finds strong evidence of social interactions college preferences and college enrollment. Giorgi et al. (2007) find that ones' behavior influences the educational decision while in college, indicating the importance of social interaction even at a later stage of someone's academic life. Sacerdote (2011) examines social interaction effects at the room and accommodation level where students are randomly assigned. He does not find any significant influence of peers.

In this paper we examine the effect of social interactions on the decisions of adolescents and young adults regarding college enrollment and academic mobility. We use a new dataset from Greece that contains information on exam scores, college enrollment and educational mobility for every student in six cohorts. We exploit the particular institutional setting in Greece, in which schools are build very close to each other. This setting allows for rich variation of school characteristics within a relatively contained geographical area. We exploit this exogenous variation in group characteristics over time and space to address the endogenous nature of the social interaction groups. The social interaction effects are defined as contextual interactions that induce different mappings from individual characteristics to outcomes (Bryk and Raudenbush (2001)). Reference groups are viewed as ecologies in which the social backgrounds affect individual choices of otherwise similar agents (Raudenbush and Sampson (1999)).

Similar age peers in one's vicinity consist a natural reference group that provide valuable and otherwise costly information, necessary in academic decision making. We widen the reference group and examine social interactions with respect to a series of reference groups: same-cohort school peers, different-cohort school peers, same-cohort peers in the neighborhood and different-cohort peers in the neighborhood.

There are particular advantages in having the universe of high school graduates for a country. First, we can observe the behaviour of all students regarding their education decisions and not only of specific groups of students. Second, we are able to observe different reference groups. A student may be affected by the decisions of same age or older peers in his school and neighborhood. We contribute to the literature by comparing the size of the social interaction effects across distance in space and age.

Empirical analysis of social interactions on students' decisions has been open to question because of the difficulties in disentangling these effects from other confounding influences.[1]. We use an instrumental variable approach and we exploit spatial and time variation to combat potential endogeneity problems and the well known "reflection-problem" (Manski (1993), Manski (2000)). There are two sources of potential endogeneity: Self selection into social groups and common shocks that affect every member of a social group. Reflection may arise from reverse causality between the outcomes of members in the same groups and their decisions are simultaneous. In other words, it is difficult to disentangle if one's actions are the cause or the effect of his peers' influence. These challenges are standard in the social interactions literature. The institutional setting behind our study refrains students from endogenously select their peers in school, facilitating the validity of the identification strategy. Moreover, the geographical density of schools allows us to define social groups wider than a student's schoolmates. Motivating from the idea of role modelship, we battle the simultaneity challenge by investigating social interactions between peers in consecutive cohorts.

By using multiple cohorts and conditioning on school and neighbourhood fixed effects as well as school-and neighborhood- specific time trends we are able to control for unobserved time-varying factors that might confound peer effects in schools and neighbourhoods. We use an instrumental variable approach to combat endogeneity and reflection. We show that within schools and neighbourhoods, there is considerable cohort-to-cohort variation in the proportion of female students that can be attributed to random factors.

We find positive spillover effects between one's decision to enrol in college and that of their peers. More specifically, the results found here indicate that students who attend a high school with a hundred percent more schoolmates who enrol in college are 12.6 percent more

[1] The existing literature that deals with identification of the social comparison effects use either laboratory experiments (Armin and Andrea (2006)), natural experiments (Zimmerman (2003)), quasi-experimental designs (Hoxby (2000)), or fixed effects (Hanushek et al. (2003)

likely to attend college. We also find positive spillovers regarding the decision of educational mobility. Students are 10.7 percent more likely to move to a different city to study if their older peers in school do so, a hundred percent more often. We find that these externalities decrease with the size of reference group.

The policy implications of social interactions can be indirect. The skills and resources that characterise a reference group are usually fixed. As a consequence, an improvement in someone's group characteristics means an equivalent deterioration in someone else's group attributes. Some may argue that the redistribution in favor of disadvantaged students can act as a boost in their scholastic outcomes, when the redistribution comes from more advantaged areas where students might depend less on their peers' quality. For example, Arcidiacono and Nicholson (2005) suggest that the existence of social interaction effects supports claims against school vouchers. This is because, the best students leaving public schools can be detrimental to the students left behind.

The paper is organized as follows. Section 2 describes the unique dataset used and the institutional setting related to college admission. The empirical strategy used to identify social interactions is analysed in Section 3. We present and discuss the results in college enrollment and educational mobility in Section 4. Finally, Section 5 concludes.

2 Data and Institutional Setting

2.1 How are students admitted to college

The transition from high school to post-secondary education in Greece is based on an unusually systematic and transparent allocation of student to university departments[2] In particular, every high school student who completes the twelfth grade receives an admission score, which is the only criterion for university admission and weights: (i) her performance in national twelfth grade exams [3] (ii) her grade twelve within school performance which is a combined

[2]Every tertiary education institute in Greece is public as free education is a constitutional right. Degrees awarded by private colleges are not recognized by the state.

[3]The twelfth grade exams are written exams administered nationally only once every year and last from late May to early June. The exams are proctored and marked externally. Exam markers do not observe the name, school, or even the city of the student whose paper they grade. Students usually take six component

score for homework and midterm exams in each subject.

After receiving their admission scores, students are required to submit a list of ranked choices of specific departments in universities that are relevant to their twelve grade track. For example, students outside the Classics track cannot list Law schools. Each university department generally offers one major of bachelor degree and no minor specializations can be declared. Every university department admits a pre-specified number of students. A computerized system at the Ministry of Education ranks students by their admission score and assigns the highest ranked student to her preferred choice. It then moves to the next student and assigns her to the first department in her list in which there is an available place, and so and so forth. In this context, students have incentives to truthfully reveal their preferences.

University departments must enrol the students assigned to them by the Ministry of Education. The Ministry of Education announces the score of the last admitted student in each university department. The last admitted students in more prestigious departments have generally higher scores in comparison to those in less prestigious ones. Once a student admitted they cannot transfer to a different major. College education is completely publicly funded and every student is exempted for college fees. Private donations to colleges are against the law.

2.2 Data

For the empirical analysis we construct a unique dataset of all students graduating from high school in Greece from 2003 to 2009. We obtain the information from various sources:

1. Administrative data from the Hellenic Ministry of Education containing course taking information and exam grades in the final year, gender, year of birth, graduation year and college admission information. In addition, the total number of places in tertiary

exams, with a combination of common subjects(Language, Mathematics, Physics, Biology or History) and four compulsory track-specific subjects and one elective exam. There are three tracks: Classics, Natural Sciences and Technical Studies. The overall score is the unweighted average of these scores. Students who fail are allowed to retake the exam the next year. In addition, students are not allowed to take the national exams early.

education in each year is provided.

2. School specific information such as name of school, type of school (private, public[4], experimental[5]), geographical coordinates and name of preferacture it belongs to[6]. There are 1319 high schools in Greece[7].

3. The Ministry of Finance provided us with average net income information at the post-code of the school in 2009 Euro.

4. The Ministry of Internal Affairs provided us with urban density information. Urban areas are those with more than 20,000 inhabitants.

5. Geographical coordinates for every tertiary education institute in Greece. There are fifty five college campuses. Not all campuses offer the same majors.

The median distance of a school from each nearest neighbouring school is 0.32 miles.[8] We use cluster analysis to define and construct neighborhoods within a mile from each school. We construct 406 clusters that cover the whole country. Every cluster is a neighborhood that contains all twelve-grade students who attend any other high school within a mile (1.06 miles) radius from one's high school[9]. Figures 1 maps all high schools and tertiary education institutes in our dataset.

Our analysis uses information regarding characteristics and choices of older school peers.

[4]Students are assigned to public schools according to a school district system

[5]Admission to experimental schools is based on a lottery

[6]There are fifty two preferactures in Greece. Preferactures are classified by the Hellenic National Statistical Authority

[7]Of which, 112 are private, and 1207 public. Of those 1207 public schools, 23 are experimental. There are no private experimental schools in Greece. 74 evening high schools for employed people of usually older age are excluded from our analysis

[8]Mean of distance from nearest neighbour: 1.85 miles. Standard deviation: 18.37 miles. 25th percentile:0.07 miles. 75th percentile: 0.77 miles.

[9]We exploit the fact that many schools were built very close to each other in most urban settings in Greece. This is more prevalent in Attica, the region surrounding the city of Athens, the capital of Greece. To give an example, in the cartier of Grava in Athens, there are six high schools next to each other along with several elementary and middle schools that form a humongous school building complex. According to the 2001 census, Attica holds around 36 percent of the total population.

Because of this, we use data on student cohorts from 2004 to 2009[10] Furthermore, our discussion of academic mobility refers to the decisions of students to move to a different preferacture in order to study, given they were admitted to some college. Thus, for this part, we focus only on admitted students[11]. Lastly, we drop 35,808 obs. for which the group of schoolmates overlapped perfectly with the social group of their neighborhood. This exclusion allows us to compare spillover effects from social groups of various sizes. We consolidate our sample by dropping observations with missing values.

Table 5 describes our pooled data across cohorts. Fifty seven percent are females. Ninety percent of the students reside in urban areas. More than 90 % of schools are public. Although, mean postcode income among private schools is significantly higher compared to public schools, mean national exam score doesn't seem to differ much. Experimental schools are in more affluent areas in comparison to other public schools as revealed by their higher mean postcode income. The mean national exam score of students attending experimental schools is much higher than the score achieved by students in private or public schools. Each neighbourhood contains on average 4 schools and 929 student observations.

3 Empirical Strategy

We start off by defining one's reference group as his same-cohort school peers. We investigate the hypothesis that social or collective behaviour patterns drive individual preferences because agents derive utility from conformity or provide access to information.

In particular, we investigate whether a student's decision to enrol in college depends on the decision of his peers in school by using the following regression:

$$IfEnrolled_{i,s,t} = \alpha + \gamma\overline{IfEnrolled_{i-1,s,t}} + \beta X_{i,s,t} + \kappa T_t + \mu S_s + \pi_s year + \epsilon_{ist}(1)$$

where $IfEnrolled_{i,s,t}$ takes the value one if student i in school s and year t enrols into college and $\overline{IfEnrolled_{i-1,s,t}}$ is the fraction of all other students except of student i in school s and year t, who enrol into college. So, we regress a student's i decision to enrol in college

[10] The first cohort in our sample, 2003 (size: 59,102 obs.), is used as a reference group for the 2004 cohort.

[11] In the academic mobility analysis we exclude 60,356 students who did not enrol in college

on the mean enrollment of his peers in school s in year t and other controls. Our covariates include a dummy for being female, a student's admission score, dummies for chosen track in the senior year of high school, dummies for the school each student attended, school specific time trends and year dummies. To control for time-varying unobserved factors that may be correlated with mean college enrolment we include a full set of school-specific linear time trends.

The main coefficient of interest is γ, which captures how the mean enrollment of one's school peers affects his decision to enrol in college. Initially, we employ ordinary least squares to estimate peer effects in education decisions. There are at least two sources of potential bias here: (1) endogeneity and (2) the reflection problem (Manski (1993), Manski (2000)).

Firstly, in many settings individuals self-select themselves into a specific group of peers that generates endogeneity issues if the variables that are responsible for this choice are not fully observable. Students who choose to attend the same high schools might share the same observed and unobserved characteristics. In this case, if we find a relationship between the observed characteristics and the outcome variable it might not be causal. This could be coming from the fact that unobserved characteristics might also affect the outcome variables. This potential unobserved heterogeneity that drives selection into social groups may bias our estimates. Nevertheless, self selection of students into schools is restricted in our setting because students are assigned to public schools[12] based on geographical criteria and they cannot choose their school peers endogenously, by construction. Therefore, social group membership is as good as random, since it does not depend on observables.

Endogeneity may also result from unobserved common group effects, such as teacher and school quality, that affect every student in a social group and render the identification of social interactions challenging. We contribute to the literature by mitigating the endogeneity challenge that stems from common group shocks. We take advantage of a special institutional setting with rich spatial and over time variation in school characteristics. We use cluster analysis to construct geographical units wider than the school district; namely neighborhoods. Those geographical units are big enough to allow for school diversity but also compact enough to capture common behavioral patterns in the area. In particular, we exploit our special

[12]92 % of students in our sample attend public or public experimental schools

institutional setting to identify same-cohort peers who do not attend the same school. We identify same-cohort peers who live in 1 mile radius and attend different schools. We call this group of same-cohort peers who live very close to each other "neighbours". In addition to their same-cohort schoolmates, students are likely to interact with their same-cohort neighbours and they might also be affected by their decisions. A students' neighbours attend different schools and face different school environments. In each neighbourhood, there are students who attend on average four different schools (Table 1). The basic idea here is to compare students' decisions from consecutive cohorts who have similar characteristics and face the same neighbourhood environment but attend different school, except for the fact that one cohort has more female students than the other. Thus, it becomes feasible to isolate the impact of a peer group from the impact of each student's school itself.

Second, reflection may arise because we cannot distinguish whether someone's action is the cause or the effect of his peers' outcomes. In other words, one's decision is simultaneous with that of his peers. We battle the simultaneity challenge by using as an IV the *time lagged* gender composition in the school and neighborhood level. So we compare the decisions of students from consecutive cohorts who had more one-cohort-older peers in their school or neighbourhood who enrolled in college. To build some intuition here, peers one-cohort-older might provide information to younger peers about the costs and the benefits of attending college or migrating to another city or they might function as "role models".

Estimating equation (1) using OLS will lead to biased results. In order to address these concerns we propose the proportion of girls in someone's reference group in the previous period as a source of variation for mean enrollment in college. The intuition is that an individual's academic decision may be related with their gender, but not the gender composition of their environment. This satisfies the exclusion restriction for the validity of our instrument.

We control for unobserved characteristics of schools, students and neighbours that are correlated with the percentage of females that could also be correlated with students' performance by exploiting variation in the gender composition across consecutive years within the same school and neighbourhood. By using multiple cohorts and controlling for school fixed effects, we take into account unobserved factors that might invalidate the school and neighbourhood peer effects analysis.

The first stage regressions are:

$$\overline{IfEnrolled_{g,t}} = \phi_1 + \kappa_1 \overline{IfFemale}_{g,\mathbf{t}} + +\beta_1 X_{i,g,t} + T_t + S_g + \pi_{1g} year + e_{1,g,t} (2)$$

$$\overline{IfEnrolled_{g,t-1}} = \phi_2 + \kappa_2 \overline{IfFemale}_{g,\mathbf{t\text{-}1}} + +\beta_2 X_{i,g,t} + T_t + S_g + \pi_{2g} year + e_{2,g,t-1} (3)$$

$$g \in \{\{school\}, \{neighborhood\}\}$$

where $\overline{IfFemale}_{j,t}$ and $\overline{IfFemale}_{j,t-1}$ is the proportion of females in geographical unit g (school and neighborhood) and year t and year t-1 respectively. The basic idea here is to compare the collective decisions of students (to enrol in college and migrate to another city to pursue tertiary education) from consecutive years who have similar characteristics but the percentage of female peers varies from one year to another. Using the proportion of girls in someone's last year's reference group as an IV relies on the assumption that this proportion has no other effect on someone's decision to enrol in college than through its effect on last year's mean college enrollment and thus this year's someone decision to enrol in college.

The second stage regressions are as follows:

$$IfEnrolled_{igt} = \delta 1 + \kappa_1 \overline{IfEnrolled}_{-i,g,\mathbf{t}} + \psi_1 X_{i,g,t} + T_t + S_g + \lambda_{1g} year + \epsilon_{1ist} (4)$$

$$IfEnrolled_{igt} = \delta 2 + \kappa_2 \overline{IfEnrolled}_{-i,g,\mathbf{t\text{-}1}} + \psi_2 X_{i,g,t} + T_t + S_g + \lambda_{2g} year + \epsilon_{2ist} (5)$$

$$g \in \{\{school\}, \{neighborhood\}\}$$

Our key identifying assumption requires that changes in the proportion of female peers within a school and within a neighborhood are not correlated with changes in unobserved factors that could affect students' decisions. In particular, it is required that changes in the proportion of females within schools and neighbourhoods are not associated with changes in student characteristics ie. age, ethnicity income, parental education. We also provide evidence that these changes in the proportion of girls within a school and within a neighbourhood are not correlated with changes in school enrolment.

Notice that we exploit within school and within neighbourhood variation from one cohort to another. Our analysis does not look at differences in the percentage of females across schools or neighbourhoods. Additionally, we look at the effect of one's peers on their decision to enrol in college and migrate to another city. To do this, we control for one's performance in the senior year national exams.

The fact that students are assigned to schools based on distance alleviates the concern that students respond to these random shocks in gender composition by switching to another school. Students need to provide adequate evidence of residence in a given region in order to have access to the closest in terms of distance school. But even if students could switch schools, then it would be very difficult to choose the destination school based on the percentage of girls in this school for the following reason: the average percentage of female peers by school or neighbourhoods is not publicly known. But even if it was known it would be difficult to know the percentage of females for a cohort that enters the school in a specific year. Additionally, leaving a school should not be correlated to the percentage of female students in this school. It is important to note that any factor affecting the proportion of girls in all geographic units in the same way, such as a female fertility decline 17 years before, will be captured by year fixed effects and would thus not invalidate the identification strategy. We include school or neighbourhood fixed effects to control for school or neighbourhood-invariant unobserved factors respectively. One could be worried that time-varying factors ie. better teachers in some years or a new college in the neighbourhood could affect mean enrolment. To address this concern, we include school- or neighbourhood-specific time trends to control for time-varying factors that could be correlated with changes in the fraction of enrolled students in one's reference group.

Next, we turn to academic mobility. We believe that there might exist social interaction effects in the decision to migrate. We model a person's decision to move to a different city in order to pursue tertiary education, given that they were admitted to some college. This decision is a function of the average decision in one's environment as specified in our regression model:

$$IfMigrate_{igt} = \alpha_1 + \gamma_1 \overline{IfMigrate_{j,g,t}} + \beta_1 X_{i,g,t} + \kappa_1 T_t + \mu_1 S_s + \phi_{1g} year + \epsilon_{1ist} (6)$$

$$IfMigrate_{igt} = \alpha_2 + \gamma_2 \overline{IfMigrate_{j,g,t-1}} + \beta_2 X_{i,g,t} + \kappa_2 T_t + \mu_2 S_s + \phi_{2g} year + \epsilon_{2ist} (7)$$

where $IfMigrate_{igt}$ is the decision of student i in geographical unit g and year t and t-1 respectively to migrate in a different city in order to study, conditional on being accepted to college. $\overline{IfMigrate_{-i,g,t}}$ and $\overline{IfMigrate_{j,g,t-1}}$ are the fractions of students except of student i who migrated to a different city in order to study in geographical unit g and year t and t-1 respectively. We include year fixed effects in order to control for time-invariant unobserved characteristics that could affect the migration decision. When we exploit the within school cohort-to-cohort change in the percentage of female students, we include school fixed effects. When we do the analysis at the neighborhood level, we include neighborhood fixed effects and we exploit the differences in school characteristics in a given year within each neighborhood.

We use an instrumental variable approach in order to estimate the effect of social interaction on the decision of students to move to another city to attend college. Again gender composition seems a likely candidate for an instrumental variable. The proportion of females in a geographical unit g may create an environment more conducive to collective migration as exhibited by average patterns of behavior but it has no direct effect on an individual's decision to migrate.

The first stage regressions is as follows:

$$\overline{IfMigrate_{g,t}} = \phi + \kappa \overline{IfFemale_{g,t}} + \beta X_{i,g,t} + \kappa T_t + \mu S_g + e_{g,t} (8)$$

$$\overline{IfMigrate_{g,t}} = \phi + \kappa \overline{IfFemale_{g,t-1}} + \beta X_{i,g,t} + \kappa T_t + \mu S_g + e_{g,t} (9)$$

$$g \in \{\{school\}, \{neighborhood\}\}$$

Our main specifications are estimated at the neighborhood level. When estimated at the geographical units of neighborhood, these specifications address both the endogeneity and simultaneity issues.

Potential threats to our analysis may include the following: Actual networks may be very different from ecologies in one's vicinity. In addition, social media may allow for peer effects that are independent of proximity and render our analysis of spatial social interactions irrelevant. This is less of a fear though as internet penetration is relatively low in Greece [13]. Parents, relatives and much older individuals in a student's environment may influence his/her academic decisions more than his/her same-cohort or one year older peers within his school and/or within his neighborhood .

4 Validity of Identification Strategy

Our identification strategy requires that fluctuations in the proportion of female students within a school and within a neighborhood should not be correlated with other cohort-to-cohort changes that could affect students' education decisions. In particular, we check if changes in the proportion of female students within a school and within a neighborhood are correlated with changes in students' observable characteristics. For the universe of students (N=355,808 students) the only characteristics we know are: the age of students and if a student enrolled early in school. This is the case if a student is born in the first quarter of his birth year.

However for a smaller sample of 45 schools (observations=18,670)we also know the ethnicity of students. In Table 4 , we present some evidence that the schools in the smaller sample have no different characteristics compared to the whole population. We cannot implement the whole analysis based on this smaller sample because we need the universe of students and schools in order to construct the neighbourhoods and exploit within neighbourhood variation. We use this smaller sample of schools to check if changes in the proportion of girls are correlated with changes in students' ethnicity.

[13] This is more understandable when one takes into account that Greece has 227 inhabited islands, most of which are quite far from the mainland and have outdated telecommunications infrastructure (Ellinikos Organismos Tourismou (EOT), "Greek islands", April 2012).

Tables 2 and 3 provide evidence on the balancing tests for the whole sample and the sub-sample of the 45 schools. Table 2 reports the estimated coefficients from the OLS regression and a within school regression (school fixed effects) of students' characteristics on the proportion of females in each school. We also report the estimated coefficients from a within school regression with school specific time trends (columns (3) and (6). Table 3 reports the estimated coefficients from the within neighbourhood regression (neighborhood fixed effects) with and without adding neighbourhood linear time trends. Again the OLS estimates are reported as a point of comparison.

As we notice from these two tables, the proportion of females is not related to most of the students' characteristics, both in the OLS and the within school/ neighborhood regressions. There are some exceptions in the OLS and within school regression. In particular, the proportion of females within a school seems to be negatively correlated with the proportion of students with Polish and Bulgarian origin, however these correlations are reduced and become statistically insignificant when we add school linear time trends. Within neighbourhoods we find no association between the proportion of females within a neighborhood and students' observed characteristics. All the regressions include year fixed effects. These results suggest that cohort-to-cohort changes in the proportion of female students within a school and within a neighborhood seem to be uncorrelated with changes in students' observed characteristics.

We also examine whether changes in the proportion of female students within a school and within a neighborhood are related to changes in the logarithm of school enrollment. As reported in the first row of Table 2 there seem to be a negative association between changes in the proportion of females within a school and changes in the logarithm of school enrolment. Both, the OLS and within school regressions produce negative and statistically significant at 10% coefficients. However, this correlation largely reduces and becomes insignificant when school specific time trends are added.

One could still have concerns that students might react to the unpredicted changes in gender compositions. Although students are assigned to schools based on geographical characteristics and it is not easy to switch school, one could still be worried that students might drop out from or switch to another school after being exposed to this information. For example, students who are in schools where the proportion of girls is high/low could drop out. Or transfers of students could be observed that might be correlated to the observed proportion of females in a given school. We address this concern by looking at the correlation between the proportion of female students in a school and the probability that a student drops out from or switch to another school in that year. We use the smaller sample of schools because only for these schools we have data for multiple years and we can identify students who drop out and transfers.

Our dependent variable is a dummy variable that takes the value of one if the student drops out from school or if the student is transferred to this school at the beginning of the school year. Table 5 reports the outcome means and the regression estimates separately for boys and girls. The first row in each panel indicates that students' mobility from and to a school in low. Approximately 8% of boys and girls drop out from school in the twelfth grade and around 6-8% of boys and girls respectively transfer to another school at the beginning of the twelfth grade. The second row in each panel reports the regression estimates when school linear trends as well as school and time fixed effects are added. All estimates are small and statistically insignificant. Overall, changes in the proportion of females within a school seem to be uncorrelated with students' mobility across schools and drop out rates.

5 Results and Discussion

Table 6 shows the linear probability model estimates for the decision to enrol in college. Columns (1) and (2) report the effects of the proportion of enrolled students in year t on a student's decision to enrol in college in year t. Columns (3) and (4) report the effects of the proportion of enrolled students in year t-1 on a student's decision to enrol in college in year t. Each cell in the first and second row in Table 6 shows the estimated coefficient from a separate regression. The estimates presented are based on four different specifications. All specifications include track and year fixed effects. Columns (1) and (3) include school fixed effects and school specific linear trends. Columns (2) and (4) include neighbourhood and neighbourhood specific linear trends. In all specifications we control for a student's gender and admission score. We also include a dummy for students who were born in the first quarter of each year, following Angrist and Krueger (1992), who found significant differences in school outcomes for those students.

The coefficients of interest are positive in year t and statistically significant, revealing strong positive externalities at all levels. An increase of a hundred percent in the proportion of same-age school peers who enrol in college increases one's probability of enrol in college by 8.6 percent, ceteris paribus. This effect decreases at the neighborhood level. In particular, an increase of a hundred percent in the proportion of same-age neighbours who enrol in college increases one's probability of enrol in college by 4.3 percent, ceteris paribus. Coefficients of interest are negative for year t-1 and not very precise.

However, OLS estimates are likely to be bias due to endogeneity issues and the reflection problem. To address these but also further potential unobserved heterogeneity issues, we employ the novel identification strategy of relying on variation in gender composition to explain differences in mean college enrollment in school and neighborhood level. We use an instrumental variable approach to explore social interactions in space and time. Our instrument, gender composition, is likely to affect mean college enrollment since female-heavy school environments are found to be less disruptive and less violent (Lavy and Schlosser (2011)).

Tables 7 and 8 report first and second stage estimates, respectively. Both tables distinguish between social interactions among same-age peers and one-cohort-older peers. Each cell in the first and second row in Table 7 shows the estimated coefficient from a separate regression. In our setup, the proportion of girls is a strong predictor of mean enrollment as all first stage estimates are positive and statistically significant at 1%. As we observe in Table 7, our instrument is a better predictor of mean enrollment at the school rather than the neighborhood level. In particular, an increase of a hundred percent in the proportion of same-age girls within a school increases mean enrollment by 13.3 % whereas an increase of a hundred percent in the proportion of same-age girls within a neighborhood increases mean enrollment by 8.5%. When we consider last year's proportion of girls then the coefficient of interest declines. In particular, a 100% increase in the percentage of girls in the previous cohort within a school increases this year's mean enrollment by 12.3 %. Furthermore, mean college enrollement within a neighborhood increases by 10.2% if the percentage of girls in the previous year increases by 100%.

Moreover, the model is just identified as we have one instrumental variable and one endogenous variable. Stock and Yogo (2002) characterize instruments to be weak not only if they lead to biased IV results but also if hypothesis tests of IV parameters suffer from considerable size distortions. They propose values of the Cragg and Donald (1993) minimum eigenvalue statistic for which a Wald test at the 5 percent level will have an actual rejection rate of less than 10 percent. In our case the critical value is 16.38 which is always below the first stage Cragg-Donald statistic we find for the school and neighborhood regressions regarding college enrollment (32.01 and 34.09 respectively) and academic mobility (582.42, 2,849 for the school and neighborhood respectively). So we do not face a weak instrument problem.

Our second stage estimates suggest positive social interactions in education decisions through space and time, with the size of the effect depending on the size of the reference group. Each cell in the first and second row in Table 8 shows the estimated coefficient from a separate regression. In Table 8, we observe that a hundred percent increase in the proportion of students who enrol in college within one's school in a given year, increases a student's probability to enrol in college by 12.6 % in the same year. Similarly, a student is 7% more likely to enrol in college in a given year if the proportion of students who enrol in college in his neighborhood increases by a hundred percent in that year. We find positive and significant spillover effects among peers in consecutive cohorts. Intuitively, social interactions among students of consecutive cohorts are important, as older peers may function as role models or may provide access to information. We find that a hundred percent increase in the proportion of students attending college within one's school or within one's neighborhood a year before, increases his probability of enrolling in college by 29.1% or 9.6 % percent respectively. Year and track fixed effects are included in all specification. When we exploit within school variation, we control for school fixed effects and school specific time trends. When we use within neighborhood variation, we control for neighborhood and neighborhood specific time trends.

Moreover, we explore social interactions in the decision to study in a different city. Educational mobility is found in the literature to be greatly affected by social norms, labor market structure and income (Tremblay (2005)). We focus on those students who enrol in college between 2004 and 2009 (sample size: 355,808). Our models include controls for school or neighbourhood, year and area unobserved time-invariant characteristics. We begin our analysis by estimating specifications (6) and (7) using standard OLS. Our estimates reveal positive social interactions among same-cohort peers and smaller positive externalities coming from students in the previous cohort. Table 10 reports the effects of the proportions of migrated students on the decision to migrate of same-cohort or one-cohort-older students.

The linear probability model coefficients are biased due to reflection and endogeneity and they show a negative relationship between mean migration and a student's decision to migrate to another city. Thus we use the proportion of female peers in one's reference group as an instrumental variable. Table 11 reports first stage estimates. Each column is coming from a separate regression. All coefficients of interest are positive and statistically significant. Again the percentage of female students is a better predictor for the mean migration within a school rather than within a neighbourhood. Ou first stage estimates suggest that changes in the percentage of female peers have significant effects on mean migration in school and neighbourhood among same-cohort students but also in consecutive cohorts. The estimates in columns (1) and (2) are higher than the estimates in columns (3 and (4) respectively implying that the effects are stronger in year t rather than t-1.

Our second stage estimates are reported in Table 9. Each column in based on a separate regression. The coefficients of interest are all positive. Our findings suggest significant positive externalities among same-cohort students but significant and smaller positive externalities among students in consecutive cohorts.

6 Conclusion

In this paper we have estimated the effects of social interactions on a student's education decisions of college enrollment and academic mobility. Despite the vast literature on the topic, two crucial identification challenges remain: common correlated group effects and simultaneity.

Our contribution to the literature is twofold. First, we propose a new approach in alleviating challenges in identifying spillover effects by using time lagged group characteristics. Second, we provide evidence on social interactions using a special institutional setting that allows for spatial variation of group characteristics. So far, the existing literature on social interactions has focused almost exclusively on scholastic performance. The only exemptions to out knowledge are Sacerdote (2011) who identify the effect of social interactions on drinking, drug use, and criminal behavior and Giorgi et al. (2007) who finds significant effects on the choice of college major.

When social interactions are not taken into account, educational treatments may result in misallocation of resources and may fall short of policy goals. Our results aim to inform public policies that target ability mismatch.

We employ instrumental variable techniques to estimate utility linkages at different space and time levels. We battle the reflection problem and the endogeneity issues by using time lagged school and neighborhood student gender compostion as an instrument. Using repeated cross-sectional data, we exploit within-school and within-neighborhood cohort-to-cohort variation to examine the effect of random changes in gender composition on mean college enrollment. Then we look at the effect on a student's decision to enrol in college.

We find that the choices of a student's peers affect their decision to enrol in college and migrate to another city to pursue tertiary education. We use a novel dataset from Greece that contains the universe of high school graduates from 2004 to 2009. We focus our analysis on four reference groups: same-cohort peers in school, one-cohort-older peers in school, same-cohort peers in neighbourhood and one-cohort-older peers in neighbourhood.

Our evidence supports the hypothesis that individuals derive utility from conformity or have access to information, with the size of the externality decreasing in space distance. Our results show that one is more likely to enrol in college and move to another city to pursue post secondary education when many of his peers make the same choices. A hundred percent increase in the percentage of one-cohort-older peers within a school and within a neighborhood who enrolled in college increases a student's probability of college enrollment by 29.1 and 9.6 percent, respectively. In addition, a hundred percent increase in the percentage of same-cohort students who enrol in college within a school and within a neighborhood increases one's own probability to enrol in college by 12.6 and 7 percent respectively.

While our paper has examined several important determinants of college enrollment and migration decision, several avenues of future research remain. Understanding the mechanism that underlies social interactions is the next big question in the literature. Future research could push forward the front of understanding the mechanism that underlies social interactions.

References

Angrist, J. D. and A. B. Krueger (1992). The Effect of Age at School Entry on Educational Attainment: An Application of Instrumental Variables with Moments from Two Samples. *Journal of the American Statistical Association 87*(418), 328–336.

Arcidiacono, P. and S. Nicholson (2005). Peer effects in medical school. *Journal of Public Economics 89*(23), 327–350.

Armin, F. and I. Andrea (2006, January). Clean evidence on peer effects. *Journal of Labor Economics 24*(1).

Bell, A. P. (1970). Role modelship and interaction in adolescence and young adulthood. *Developmental Psychology 2*.

Bernheim, B. D. (1994). A Theory of Conformity. *Journal of Political Economy 102*(5), 841–877.

Brock, W. A. and S. N. Durlauf (2001). Discrete Choice with Social Interactions. *The Review of Economic Studies 68*(2), 235–260.

Bryk, A. and S. Raudenbush (2001). Hierarchical linear models, second edition. Technical report.

Cipollone, P. and R. Alfonso (2007). Social interactions in high school: Lessons from an earthquake. *American Economic Review 97*(3), 948–965.

Cragg, J. G. and S. G. Donald (1993). Testing Identifiability and Specification in Instrumental Variable Models. *Econometric Theory 9*(02), 222–240.

Fletcher, J. (2006). Social interactions in college choice: The interplay of information, preferences, social norms and individual characteristics in predicting college choice. *Working Paper*.

Garner, C. L. and S. W. Raudenbush (1991). Neighborhood Effects on Educational Attainment: A Multilevel Analysis. *Sociology of Education 64*(4), 251–262.

Giorgi, G. D., M. Pellizzari, and S. Redaelli (2007, June). Be as careful of the books you read as of the company you keep: Evidence on peer effects in educational choices. *NBER Working Papers 14948*, 241–275.

Hanushek, E., J. Kain, J. Markman, and S. Rivkin (2003, September). Does peer ability affect student achievement? *Journal of Applied Economics 18*(5), 527–544.

Hoxby, C. (2000). Peer Effects in the Classroom: Learning from Gender and Race Variation. Working Paper 7867, National Bureau of Economic Research.

Lavy, V. and A. Schlosser (2011). Mechanisms and Impacts of Gender Peer Effects at School. *American Economic Journal: Applied Economics 3*(2), 1–33.

Manski, C. F. (1993). Identification of Endogenous Social Effects: The Reflection Problem. *The Review of Economic Studies 60*(3), 531–542.

Manski, C. F. (2000). Economic Analysis of Social Interactions. Working Paper 7580, National Bureau of Economic Research.

Raudenbush, S. and R. Sampson (1999). Ecometrics: Towards a science of assessing ecological settings, with application to the systematic social observation of neighborhoods. Technical report.

Sacerdote, B. (2011). Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far? Handbook of the Economics of Education, Elsevier.

Schelling, T. C. (1973, September). Hockey Helmets, Concealed Weapons, and Daylight Saving: A Study of Binary Choices with Externalities. *The Journal of Conflict Resolution 17*(3), 381–428.

Stock, J. H. and M. Yogo (2002). Testing for Weak Instruments in Linear IV Regression. Working Paper 284, National Bureau of Economic Research.

Tremblay, K. (2005). Academic Mobility and Immigration. *Journal of Studies in International Education 9*(3), 196–228.

Zimmerman, D. J. (2003). Peer Effects in Academic Outcomes: Evidence from a Natural Experiment. *Review of Economics and Statistics 85*(1), 9–23.

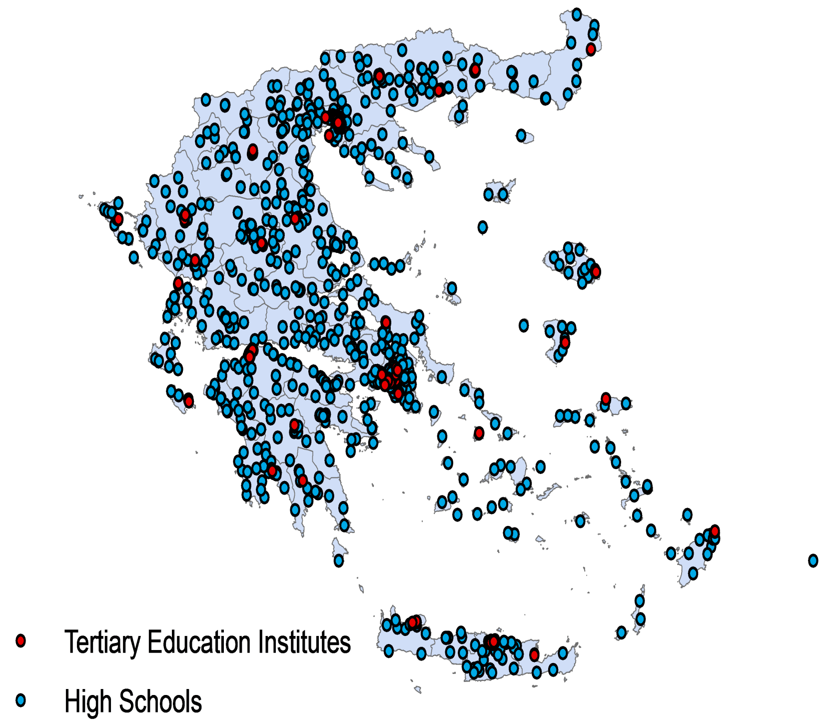

Figure 1: Map of schools

Table 1: Descriptive Statistics

	Mean	Std. Dev.	Min.	Max.	N
Panel A: Individual Level					
First quarter of birth	0.16	0.368	0	1	355,808
Female	0.567	0.495	0	1	355,808
National Exams Score	13.16	4.062	0.52	19.95	355,808
If enrolled	0.812	0.391	0	1	355,808
Mobile students	0.748	0.434	0	1	260,472
Specialty in Classics	0.365	0.481	0	1	355,808
Specialty in Natural Science	0.154	0.361	0	1	355,808
Specialty in Technical Studies	0.484	0.5	0	1	355,808
Postcode Income (Euro, 2009)	29,464	8,441	9,573	122,879	355,808
Aggregate Enrollment	60,206	6,372	52,450	68,136	355,808
Panel B: School Level					
Private	0.081	0.266	0	1	1,319
Income if private (Euro, 2009)	30,575	18,378	16,085	122,879	1,319
National score if private	13.69	2.70	4.7	17.34	1,319
Experimental	0.022	0.149	0	1	1,319
Income if experimental (Euro, 2009)	29,754	14,775	17,583	74,798	1,319
National score if experimental	14.40	1.00	12.23	16.17	1,319
Public	0.89	0.31	0	1	1,319
Income if public (Euro, 2009)	19,327	5,565	9,573	74,798	1,319
National score if public	12.26	1.56	2.97	16.36	1,319
Urban	0.898	0.301	0	1	1,319
Distance to nearest college campus(in miles)	10.871	24.083	0.105	1095.452	1,319
No of students in each school	46	34	0.16	179	1,319
Panel C: Neighborhood Level					
No of schools in each neighborhood	4.449	5.014	2	35	250
No of students in each neighborhood	929.291	1,246.298	8	10,559	250

Note: Data span six cohorts 2004-2009 of 60.119 students on average. Number of schools: 1319. Among those 413 high schools are in Athens or the surrounding suburbs. The national exam score ranges from 0 to 20. Mobile students are those who move to a different city in order to study.

Table 2: BALANCING TESTS FOR PROP. OF FEMALES WITHIN SCHOOL

	WHOLE SAMPLE			**SMALLER SAMPLE**		
	OLS	School FE	School FE +school linear time trends	OLS	School FE	School FE +school linear time trends
	(1)	(2)	(3)	(4)	(5)	(6)
logEnrollment	-0.075	-0.074	0.004	-0.125	-0.125	0.006
	(0.036)*	(0.036)*	(0.024)	(0.053)*	(0.054)*	(0.048)
EarlyEnrollment	-0.002	0.005	0.0009	-0.003	-0.006	0.0010
	(0.019)	(0.019)	(0.001)	(0.016)	(0.016)	(0.001)
Age	-0.002	0.002	-0.002	-0.003	-0.001	-0.004
	(0.002)	(0.002)	(0.003)	(0.010)	(0.010)	(0.006)
Ethnicity						
Greece				0.001	0.001	0.003
				(0.001)	(0.001)	(0.002)
Albany				0.004	0.004	0.016
				(0.013)	(0.013)	(0.011)
Bulgaria				-0.060	-0.060	-0.033
				(0.027)*	(0.027)*	(0.022)
Italy				-0.010	-0.010	-0.001
				(0.011)	(0.011)	(0.006)
Russia				-0.012	-0.012	0.011
				(0.027)	(0.027)	(0.017)
Poland				-0.054	-0.054	-0.022
				(0.024)*	(0.024)*	(0.019)
Ukraine				-0.003	-0.003	0.004
				(0.016)	(0.016)	(0.013)
N	355,808	355,808	355,808	18,670	18,670	18,670

Note: Standard errors are clustered at the school level. A constant is also included. *,**,*** denotes significance at the 10%,5% and 1% level respectively. The table reports OLS and school fixed effects estimates from separate regressions. Columns (3) and (6) report school fixed effects estimates having added school linear time trends. Year dummies are included in all regressions.

Table 3: BALANCING TESTS FOR PROP. OF FEMALES WITHIN NEIGHBOURHOOD

	WHOLE SAMPLE			**SMALLER SAMPLE**		
	OLS	neighb. FE	neighb. FE+ neighb. linear time trends	OLS	neighb.FE	neighb. FE+ neighb. linear time trends
	(1)	(2)	(3)	(4)	(5)	(6)
logEnrollment	-0.081	-0.080	0.002	-0.089	-0.089	0.003
	(0.048)	(0.050)	(0.035)	(0.050)	(0.055)	(0.031)
EarlyEnrollment	0.002	0.005	0.001	0.004	0.004	0.001
	(0.003)	(0.003)	(0.035)	(0.007)	(0.007)	(0.023)
Age	0.001	0.003	-0.002	0.003	0.003	-0.001
	(0.001)	(0.001)	(0.005)	(0.004)	(0.003)	(0.002)
Ethnicity						
Greece				0.001	0.001	0.003
				(0.001)	(0.001)	(0.002)
Albany				-0.007	-0.007	-0.004
				(0.006)	(0.006)	(0.003)
Bulgaria				0.002	0.002	0.008
				(0.009)	(0.009)	(0.006)
Italy				-0.004	-0.004	-0.002
				(0.004)	(0.004)	(0.002)
Russia				0.005	0.005	0.009
				(0.009)	(0.009)	(0.006)
Poland				-0.002	-0.002	0.005
				(0.010)	(0.010)	(0.004)
Ukraine				-0.010	-0.010	-0.009
				(0.006)	(0.006)	(0.005)
N	355,808	355,808	355,808	18,670	18,670	18,670

Note: Standard errors are clustered at the neighborhood level. A constant is also included. *,**,*** denotes significance at the 10%,5% and 1% level respectively. The table reports OLS and neighborhood fixed effects estimates from separate regressions. Columns (3) and (6) report neighborhood fixed effects estimates having added neighborhood linear time trends. Year dummies are included in all regressions.

Table 4: Descriptive statistics for smaller sample and population

	Smaller Sample Mean	**Population Mean**	**Difference**	**Std. Dev.**
log Postcode income	9.962	9.968	0.006	(0.014)
Private school	0.080	0.081	-0.001	(0.001)
Public school	0.897	0.899	-0.002	(0.003)
Experimental school	0.020	0.022	0.002	(0.003)
Urban	0.899	0.898	0.001	(0.001)

Note: 18,670 obs. in smaller sample and 355,808 obs. in population. 45 schools in sample, 1319 schools in population. Evening schools are excluded from the sample and the population

Table 5: Estimation results : Drop out and Transfers

Dependent Variable: Dummy for drop out and Transfers		
	(1)	(2)
Variable	(Males)	(Females)
Drop out		
Outcome mean	0.080	0.078
Regression estimates	0.060	0.020
	(0.046)	(0.038)
Transfers		
Outcome mean	0.068	0.075
Regression estimates	0.020	-0.052
	(0.075)	(0.071)

Note: The table reports means of the dependent variable (first row) and estimates (second row) for the effects of the proportion of females on the probability that a student leaves school the following year. We use the smaller sample here of 45 schools. Clusters at school level. All regressions include controls for student characteristics. Standard errors are clustered at the school level. All regressions include school fixed effects, year fixed effects and school linear time trends.

Table 6: Linear Probability Model Estimates

Dependent Variable: College Enrollment				
	(1)	(2)	(3)	(4)
Sample:	School	Neighborhood	School	Neighborhood
% Enrolled_t	0.086	0.043		
	(0.011)***	(0.012)***		
% Enrolled_{t-1}			-0.039	-0.017
			(0.012)***	(0.014)
Born in 1st quarter	0.006	0.006	0.007	0.007
	(0.001)***	(0.001)***	(0.001)***	(0.001)***
Female	-0.003	-0.003	-0.003	-0.003
	(0.001)**	(0.001)***	(0.001)**	(0.001)***
Admission Score	0.069	0.069	0.069	0.069
	(0.000)***	(0.000)***	(0.000)***	(0.000)***
Speciality FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓
School FE	✓	✓	✓	✓
School specific time trends	✓		✓	
Neighbourhood specific time trends		✓		✓
N	355,808	355,808	355,808	355,808
R^2	0.47	0.47	0.47	0.47

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. Standard errors are clustered at the school level. An intercept is also included.

Table 7: First stage estimates

Dependent variable:	**Mean College Enrollment$_t$**			
	school	neighborhood	school	neighbourhood
Proportion of girls$_t$	0.133	0.085		
	(0.019)***	(0.030)***		
Proportion of girls$_{t-1}$			0.123	0.102
			(0.018)***	(0.003)***
Female	-0.0004	-0.0008	-0.0001	-0.0004
	(0.0003)	(0.0002)***	(0.0003)	(0.0003)
Admission Score	0.001	0.0002	0.00002	0.00003
	(0.0008)***	(0.00004)***	(0.0003)***	(0.00003)
Born in first quarter	0.0001	-0.0006	0.0002	0.0008
	(0.00004)	(0.0002)***	(0.0002)	(0.0007)
Speciality FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓
School FE	✓	✓	✓	✓
School specific time trends	✓		✓	
Neighbourhood specific time trends		✓		✓
N	355,808	355,808	355,808	355,808
R^2	0.54	0.70	0.43	0.62
F-statistic 1st stage	14.22	17.57	12.11	18.13

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. Standard errors are clustered at the school level. An intercept is also included.

Table 8: IV Second Stage Estimates

	Dependent Variable: College Enrollment			
	(1)	(2)	(3)	(4)
Sample:	School	Neighborhood	School	Neighborhood
% Enrolled$_t$	0.126	0.070		
	(0.036)***	(0.033)**		
% Enrolled$_{t-1}$			0.291	0.096
			(0.037)***	(0.035)***
Admission Score	0.069	0.069	0.069	0.069
	(0.001)***	(0.000)***	(0.000)***	(0.000)***
Female	-0.003	-0.003	-0.003	-0.003
	(0.001)***	(0.001)***	(0.001)***	(0.001)***
Born in first quarter	0.006	0.006	0.006	0.007
	(0.001)***	(0.001)***	(0.001)***	(0.001)***
Speciality FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓
School FE	✓	✓	✓	✓
School specific time trends	✓		✓	
Neighbourhood specific time trends		✓		✓
N	355,808	355,808	355,808	355,808

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. Standard errors are clustered at the school level. An intercept is also included.

Table 9: LPM Migration Decision

	Dependent Variable: Migration Decision			
	(1)	(2)	(3)	(4)
Sample:	School	Neighborhood	School	Neighborhood
% Migrated$_t$	-0.099	0.095		
	(0.035)***	(0.037)***		
% Migrated$_{t-1}$			-0.087	-0.067
			(0.026)***	(0.022)***
Born in 1st quarter	-0.001	-0.001	-0.001	-0.001
	(0.002)	(0.002)	(0.002)	(0.002)
Female	-0.013	-0.013	-0.013	-0.013
	(0.002)***	(0.002)***	(0.002)***	(0.002)***
Admission Score	-0.023	-0.023	-0.023	-0.023
	(0.002)***	(0.002)***	(0.002)***	(0.002)***
Speciality FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓
School FE	✓	✓	✓	✓
School specific time trends	✓		✓	
Neighbourhood specific time trends		✓		✓
R^2	0.30	0.30	0.30	0.30
N	355,808	355,808	355,808	355,808

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. Standard errors are clustered at the school level. A intercept is also included.

Table 10: First stage estimates for migration decision

Dependent Variable: Mean Migration Decision				
	(1)	(2)	(3)	(4)
Sample:	School	Neighborhood	School	Neighborhood
Proportion of girls$_t$	0.125	0.098		
	(0.023)***	(0.020)***		
Proportion of girls$_{t-1}$			0.114	0.089
			(0.027)***	(0.022)***
Admission Score	0.0002	0.001	0.0006	0.001
	(0.0001)*	(0.0001)***	(0.0003)***	(0.0001)***
Female	0.002	0.003	0.001	0.002
	(0.0008)**	(0.007)*** (0.0008)	(0.0007)***	
Born in 1t quarter	-0.0006	-0.0007	-0.006	-0.0008
	(0.001)	(0.001)	(0.001)	(0.001)
Speciality FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓
School FE	✓	✓	✓	✓
School specific time trends	✓		✓	
Neighbourhood specific time trends		✓		✓
N	260,472	260,472	260,472	260,472
R^2	0.37	0.39	0.37	0.39
$F - statistic 1st stage$	16.20	14.8	15.9	14.7

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. Standard errors are clustered at the school level. An intercept is also included.

Table 11: IV Estimates for Migration Decision

	Dependent Variable: Migration Decision			
	(1)	(2)	(3)	(4)
Sample:	School	Neighborhood	School	Neighborhood
% Migrated$_t$	0.107	0.091		
	(0.032)***	(0.027)***		
% Migrated$_{t-1}$			0.101	0.065
			(0.036)**	(0.034)*
Admission Score	-0.022	-0.023	-0.023	-0.023
	(0.000)***	(0.000)***	(0.000)***	(0.000)***
Female	-0.014	-0.014	-0.014	-0.014
	(0.002)***	(0.002)***	(0.002)***	(0.002)***
Born in 1st quarter	-0.001	-0.001	-0.001	-0.002
	(0.002)	(0.002)	(0.002)	(0.002)
Speciality FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓
School FE	✓	✓	✓	✓
School specific time trends	✓		✓	
Neighbourhood specific time trends		✓		✓
N	260,472	260,472	260,472	260,472

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. Standard errors are clustered at the school level. An intercept is also included.