A Pleasure That Hurts: The Ambiguous Effects of Elite Tutoring on Underprivileged High School Students

Son Thierry Ly, *Paris School of Economics*

Eric Maurin, *Paris School of Economics*

Arnaud Riegert, *Paris School of Economics*

This paper reports findings from a randomized evaluation of an intensive tutoring program conducted in underprivileged high schools. Within each school, the intervention targets students identified as having the ability to pursue a college education. The program is designed to strengthen their readiness for higher education. We demonstrate that such an intervention can have negative effects on a large fraction of participants, even though participation is entirely voluntary. This result is consistent with a simple model where time invested in extracurricular programs and time invested in homework represent imperfect substitutes in the education production function.

I. Introduction

In most developed countries, students coming from low-income families are massively underrepresented in the most prestigious programs of higher education. This underrepresentation contributes to the exclusion of entire

We thank Laurent Davezies, Pascaline Dupas, and Catherine Moizan for their comments on previous versions of this work. We thank Marie Lorin, Alexandre Vayer, and Clara Wolf for providing excellent research assistance. Funding was

[*Journal of Labor Economics*, 2020, vol. 38, no. 2]
 0734-306X/2020/3802-0005$10.00
Submitted December 28, 2016; Accepted January 14, 2019; Electronically published February 24, 2020

social groups from political and economic elites. It has attracted considerable attention from both policy makers and social scientists, but the mechanisms driving unequal access to higher education are still debated (Bailey and Dynarski 2011).

In this context, many initiatives have flourished around the world to help good students from underprivileged backgrounds get into higher education. In the United States, where tuition fees are important and where the college application process is decentralized, several recent studies emphasize the importance of providing students with assistance with college applications as well as with financial aid or with information on how to obtain financial aid (see, e.g., Bettinger et al. 2012; Hoxby and Turner 2013; Castleman, Page, and Schooley 2014; Kautz and Zanoni 2014; Castleman and Goodman 2018). In a European country like France, where tuition fees are very modest and where the application process is automated and centralized, the Ministry of Education has encouraged institutions of higher education to develop tutoring programs in underprivileged high schools to help the best students from these high schools improve their educational record and form more ambitious plans. There exist about 350 such programs all over the country. These programs are called *cordées de la réussite* (team for success) and are becoming increasingly popular. To the best of our knowledge, however, very little is known about the actual impact of these programs on eligible students. On the one hand, they potentially contribute to bridging the cultural gap between underprivileged students and higher education. But on the other hand, they are often time-consuming and likely contribute to distract students from basic subjects. Also, there is no consensus on who exactly should be eligible for these programs. In particular, the question is open as to whether they should be offered to all students willing to pursue higher education or restricted to the very best students only.

To shed light on these issues, our paper reports the results from a randomized evaluation of one of the oldest *cordée de la réussite*, TALENS. It is operated by one of the most prestigious institutions of higher education in France, which is also one of the most selective in the world, namely, the École Normale Supérieure (ENS) of Paris.[1] Since 2006, the ENS offers each year a 2-year mentoring and tutoring program to a selection of students coming from 12 underprivileged high schools of the Paris region. Each year, participants are randomly selected at the end of their first year of high school (grade 10) from a group of volunteers identified by school principals as having the ability to succeed in high school and to pursue college education.

provided by the Fonds d'Expérimentation pour la Jeunesse. Contact the corresponding author, Eric Maurin, at eric.maurin@ens.fr. Information concerning access to the data used in this paper is available as supplemental material online.

[1] The ENS is the institution with the highest proportion of Nobel Prizes among former students in the world, before Caltech and Harvard (Wai and Hsu 2016). In mathematics, the ENS is second in the world (just below Princeton) in terms of number of Field Medals won by former students.

Participants have typically a much better educational background than the average high school students in the Paris region, even though they come much more often from an underprivileged immigrant family. Once selected, participants are divided into small groups, and each group is randomly assigned to a specific tutor. During the last 2 years of high school (grades 11 and 12), each group is invited each month to participate in one or two tutoring sessions dedicated to deepening subjects or to exploring new fields of study. Tutors are volunteer graduate students from the ENS. Eligible students are also invited to participate in specific sessions dedicated to help them prepare high school exit examination as well as with college choice. The general objective of the program is to further strengthen participants' academic achievement, to improve their readiness for higher education, and to give them a better idea of the requirements of higher education. The traveling costs associated with participation in the program (as well as the cost of attending a 1-week introductory meeting at the ENS in Paris) are all covered by the program. All in all, the intervention represents about 150 hours of tutoring per student and year. It costs about 1,500 euros per student and year.

This paper focuses on volunteer students identified by principals in 2010 and 2011. It shows the results of comparing the achievement and choices of those randomly selected to be eligible (the treatment group) with the achievement and choices of those not selected (the control group). This evaluation reveals that the intervention has very little effect on students' average outcomes. In particular, there is no significant difference between treatment and control groups in average performance on the national high school exams (*baccalauréat*) taken at the end of grade 11 and grade 12. Similarly, there is no significant difference in the proportion of students who get access to (and are able to persist in) the most selective undergraduate programs (called *classes préparatoires aux grandes écoles* [CPGE]).

In the initial stage of the project, our plan was to target higher-ability students only, because we believed that our high-intensity program could be beneficial to good students only. But most school principals had different beliefs and thought that the program could be beneficial to a majority of students, not simply the higher-ability ones. Eventually, it proved impossible to impose minimum academic requirements, and we ended up with a significant amount of heterogeneity among eligible students. In this context, we tested for heterogeneous effects across ability groups and found that the effect of the intervention is actually very different for the best students and for the other ones. Specifically, among the 50% eligible students with the highest level of achievement pretreatment (referred to as "high-ability" students), the intervention induces a significant increase in high school achievement as well as in the probability of getting access to (and persisting in) the most selective undergraduate programs. By contrast, among the other half of eligible students (referred to as "mid-high-ability" students), the intervention induces a significant decrease in both high school achievement and the probability of entering selective undergraduate programs. The depressing effect of the

intervention on mid-high-ability participants is likely one reason why a majority of these students choose to quit the program at the end of the first year, whereas only a minority of high-ability participants choose to quit (54% of quitters among mid-high-ability students vs. 37% among high-ability ones).

All in all, the intervention has no impact on the average outcomes of eligible students but a significant effect on inequalities across the high-ability and mid-high-ability ones. The gap in high school graduation between these two groups is about three times more important in the treatment group than in the control group. The results are qualitatively similar for the first cohort and the second cohort, meaning we get similar results from two independent experiments. Our main findings are also robust to multiple testing corrections.

From a theoretical viewpoint, we show that these findings are consistent with a simple model where the program is assumed to induce the substitution of extracurricular activities for curriculum-related activities (such as school homework). Under relatively mild assumptions about the concavity of the education production function, we show that such a substitution may have very different effects on students with different initial endowment in curriculum-related knowledge, namely, positive effects on the achievement of students whose initial endowment in curriculum-related knowledge is relatively strong (i.e., higher-ability students) and adverse effects on students whose initial endowment is relatively low (i.e., lower-ability students).

To further explore why the program has different effects on the two ability groups, we conducted in 2016 a posttreatment survey on the sample of students who participated in the program in 2010 and 2011. The survey confirms that participants from both groups perceived the program as too time-consuming and difficult to reconcile with school homework. Both groups also report that they became friends with other participants and spent a lot of time with them outside the sessions. Many students report that one reason why they persisted in the program is actually that they wanted to stay with their new friends. Hence, many participants from both ability groups chose to stay in the program even though it was too time-consuming for them, simply because they had great pleasure in participating and did not want to lose their new friends. Overall, the posttreatment survey appears to be consistent with a model where both high- and mid-high-ability participants are induced to distract a significant amount of time from curriculum-related activities but where a given reduction in the amount of time invested in these activities has drastically different implications depending on participants' initial endowment in curriculum-related knowledge.

Building on the fact that tutors were randomly assigned to tutees, we provide additional evidence suggesting that tutors who obtained the more negative results were, paradoxically, those who were on average the closest to their tutees (and the most positively perceived by them), namely, tutors who themselves came from a lower socioeconomic background.

Overall, our paper contributes to the large and long-standing literature that explores the achievement gap between high school students with different family backgrounds (Bailey and Dynarski 2011; Reardon 2011). Our results for high-ability participants show that an intensive tutoring and mentoring program is able to bridge this gap, at least when it targets the very best students. But our findings for mid-high-ability participants suggest that any such intensive intervention also runs the risk of being counterproductive, even when it is based on voluntary participation. Many students can find it pleasurable to participate in a program which, in the end, happens to have negative effects on their performance and negative effects on their probability of gaining access to (and persisting) in the best undergraduate programs. These findings contribute to the literature emphasizing that adolescents and young adults often focus too much on the present, which can lead them to make decisions that are not necessarily in their long-run best interest (Lavecchia, Liu, and Oreopoulos 2014). Our results also add to the literature on youth extracurricular activities, which has repeatedly documented that high levels of participation in these activities can take time away from homework and be associated with lower academic performance (see, e.g., Fredricks and Eccles 2010; Fredricks 2012; Knifsend and Graham 2012).

The paper is organized as follows. The next section describes the context of the experiment. Section III describes the content of the intervention and the randomization process. Section IV develops a simple conceptual framework that helps clarify why the intervention may differentially affect higher- and lower-ability participants. Section V describes the data used in the econometric analysis, while section VI presents the main results of our experiment, namely, the impact of the intervention on students' performance on high school exit exams as well as on their probability of gaining access to (and persisting in) selective undergraduate programs. Section VII explores the variation in the effect of the program across the different types of tutors. Section VIII builds on a survey on former participants conducted in 2016 to test assumptions about why a program designed to help students may end up having a negative effect on a significant fraction of them.

II. Institutional Context

In France, compulsory education encompasses 5 years of elementary school (grades 1–5) and 4 years of middle school (grades 6–9) until age 15. At the end of grade 9, about 60% students enter high school and pursue general education for three additional years (grades 10–12), whereas 40% go to a vocational school or enter the labor market. At the end of grade 10, students in the general education track can enter either a more academic program (about 70% do so) or a more technical one. Students who enter an academic program have to choose a major field of study; about half of them specialize in science, while the other half specialize in humanities (either in literature/languages or

economics/social sciences).[2] As discussed below, students eligible to the ENS tutoring program are selected among the high-achieving tenth graders of 12 underprivileged high schools. Virtually all of them opt for the academic program at the end of grade 10, and about two-thirds specialize in science.

In grades 11 and 12, students prepare for the exams that are required to graduate high school. There is one exam per subject, and graduation is based on the average mark across the different subjects. Graduation is a necessary condition for admission into higher education. Some specific exams take place at the end of grade 11 (most notably oral and written French exams), but most of them take place at the end of grade 12. The overall number of exams and their relative importance depend on whether students chose science or humanities as their major field of study at the end of grade 10.

After high school, students who want to enter a selective undergraduate program[3] have to apply through a centralized assignment system called Admissions Post Bac. They are allowed to apply to a maximum of 36 undergraduate programs that they must list by descending order of preference. Each selective undergraduate program ranks its applicants based on the marks obtained during eleventh and twelfth grade (as assessed by teachers). The system assigns as many students as possible to one of their listed choices using a deferred acceptance mechanism (Roth 2008).

It should be emphasized that the application process is under the direct supervision of the high school administration. When applying to a selective program, students do not have to ask teachers for transcripts since transcripts are automatically put online by the administration. Also, when a student applies to a selective program, each teacher has to provide a short qualitative evaluation of the relevance of this application. These evaluations are automatically put online too. In this context, students do not have to write essays or ask recommendation letters either, as is usually the case in the United States, for instance. Each class has a reference teacher (called *professeur principal*) who has online access to the application file of each student. These reference teachers monitor online whether students provide their lists of application on time. They also organize information sessions with their classes.

The most selective undergraduate programs correspond to the CPGE. Among high school students in an academic track, only about 13% are admitted to a CPGE program (17% of those who specialize in science), and

[2] It should be emphasized that there is an important gap between grade 10 and grade 11, especially for those who specialize in science, so that only the best tenth graders are allowed to enter the more academic programs. When we use scores obtained at the end-of-middle-school exams as a measure of ability, we find that the average score of those who specialize in science after grade 10 is 60% of a standard deviation larger than the average score of those who specialize in humanities and 140% of a standard deviation larger than the average score of those who pursue a nonacademic program.

[3] About half of the undergraduate programs are selective; i.e., they are oversubscribed and admission is based on academic results in high school.

77% make it to the second year. These CPGE programs involve 2 years of intense preparation (in either science or humanities), at the end of which students take competitive exams for entry into the most prestigious graduate programs. Admission to the ENS itself is based on one such competitive exam. Most ENS students have gone through a 2-year CPGE preparation program before entry into the ENS. A last important feature of French higher education is that tuition fees are very low even in selective prestigious programs like the CPGE.[4] In this context, lack of information on tuition fees and financial aid is an unlikely explanation for the underrepresentation of students of low socioeconomic status (SES) in the CPGE.

III. The Experiment

A. The Program and Its Objectives

In 2008, the French government initiated programs (called *cordées de la réussite*) all over the country to increase the proportion of students from underprivileged high schools entering higher education. At the local level, each specific intervention involves the collaboration of an institution of higher education and a set of high schools located in the same region. Each year, a selection of students from these high schools is given the opportunity to participate in a program designed to improve their readiness for higher education. In most cases, the program is conducted by volunteer tutors coming from the higher education institution itself. The program analyzed in this paper corresponds to the network constituted by the ENS and 12 underprivileged high schools from Paris and its region.

The ENS encompasses both a very selective graduate school and a set of world-class research centers. As mentioned above, the ENS is one of the institutions with the highest proportion of Nobel Prizes and Field Medals among former students in the world. This institution has played a leading role in the selection and training of the French intellectual elite for more than a century.

The high schools were selected on the basis of the socioeconomic background of their students as well as the proportion of students they send to CPGE programs: only 8% of their students enter such selective programs (11% of those who specialize in science), which is about two times less than the average high school in the region of Paris (14%, and 20% of those who specialize in science).

B. Identification of Volunteers

Each year, in each high school of the network, participants to the ENS program are randomly selected from a set of tenth-grade volunteers. The

[4] In public universities, tuition fees are only about 200 euros. For historical reasons, CPGE programs do not depend on universities (but on high schools) and are actually tuition-free. For more historical details on the CPGE, see, e.g., Belhoste (2003).

identification of volunteers takes place early April, about 2 months before the end of the academic year. In each school, the principal starts by identifying tenth-grade students who are likely to enter an academic program in grade 11 and likely to pursue a college education after grade 12. The principal invites them to participate in an informational meeting where the program managers provide detailed information on the objectives and contents of the program. During this meeting, the program managers make it clear that the number of seats in the program is limited and that eligible students have to be randomly selected among volunteers.

The project was funded by the French Ministry for Youth. In the proposal we submitted when we applied for the grant, we proposed to restrict invitations to students whose average marks were both in the top half of their class and above the 11/20 threshold; that is, we proposed to use objective criteria to define the group of volunteers who were to be invited to the informational meeting. As written in the proposal, our working assumption was that "these eligibility requirements are necessary if we want the program to have an impact on treated students, most notably in terms of improved access to the *classes préparatoires aux grandes ecoles*" (authors' translation).[5] But several school principals found this approach too restrictive, and our definition proved impossible to implement. In fact, no strong consensus emerged among principals about how to identify students with the potential to take advantage of the intervention. Eventually, we agreed to let principals choose how exactly to define a "good" student and who exactly to invite in the informational meeting. The experiment was expected to help further explore whether the program has the same effect on all students.

At the end of the informational meeting, students who are interested in actual participation are invited to take a questionnaire (about their family background and school experience), fill it out at home, and bring it back 1 week later. Those who come back 1 week later with their completed questionnaire undergo a short interview with the program manager.

The manager checks whether the questionnaire has been filled out well and also whether the student has understood the implications of volunteering well. In particular, students are reminded that not all volunteers will be eligible to participate in the program, only a random selection. At the end of this interview, students who confirm their willingness to participate in the program are considered as volunteers.

[5] The full text of the grant proposal is available at https://www.parisschoolofeconomics.eu/docs/maurin-eric/grant_proposal_a1_p2_ens-eep.pdf. Had we restricted invitations to students in the top half of their class and above the 11/20 threshold (as we initially intended to do), the number of students included in the experiment would have been reduced by 36% (from $N = 542$ to 344). As discussed below, the average impact of the program would have been different; that is, we would have found a significant positive effect on exit scores and high school graduation.

In a typical year, the capacity of the program is about 140 seats, and there are about 200 volunteers. In academic year 2010–11 (our first cohort), the ENS agreed to temporarily increase the capacity of the program, and school principals were encouraged to boost participation in the informational meetings. In this specific year, we ended up with 395 volunteers for 216 seats. In the following year (2011–12), the capacity was back to normal (140 seats), and the number of volunteers was back to 212 students.

C. The Sample of Volunteer Students

Once the lists of volunteers are completed in the different experimental schools, eligible students are randomly selected from these lists.[6] As discussed below, the randomization is stratified by school and major field of study. In 2011, 51 volunteer students did not participate in the draw (and were automatically selected) because there were too few volunteer students in their school with a similar field of study. Overall, for the two cohorts under consideration, a total of 556 volunteer students are included in the draw (395 for the first cohort and 161 for the second cohort).

Building on schools' registers, table 1 provides some statistics describing these 556 volunteer students as well as their nonvolunteer schoolmates and the average students in the Paris region. The table confirms that volunteer students have a much better academic record not only than their nonvolunteer schoolmates but also than the average high school student in the Paris region, even though they come more often from a low-SES family than the average high school student. Specifically, the score obtained by volunteer students on the end-of-middle-school national examination (externally set and marked) is on average about +16% of a standard deviation higher than the score obtained by the average high school student in the Paris region, even though the proportion of students coming from a low-SES family is about 40% greater among volunteer students than among the average high school student in the Paris region. The majority of volunteer students (55%) are actually in the top quintile of the distribution of scores within their high school, whereas only about 5% are below the median of this distribution. Virtually all volunteer students pursue an academic track after grade 10, whereas the average proportion in the Paris region is only about 58%. Generally speaking, the table confirms that the program was able to target relatively high-ability students with relatively low socioeconomic backgrounds, compared with the average student in the Paris region. According to the baseline questionnaire filled out by volunteer students, about 63% have an immigrant background, whereas the national proportion is only about 20% (Caille 2010).

[6] All volunteer students are informed by email about whether they are selected. Both selected and nonselected volunteers are reminded that decisions are the result of a random draw.

Table 1
Characteristics of Students Who Volunteer to Participate in the Program

Variable	Paris Region (1)	Experimental High Schools (2)	Volunteer Students (3)	Volunteers, Cohort 2010 (4)	Volunteers, Cohort 2011 (5)
Female	.528	.536	.601	.618	.559
High-SES family	.416	.200	.255	.258	.248
Low-SES family	.265	.438	.362	.357	.373
Average ninth-grade score	.000	−.557	.156	.137	.203
Has repeated a grade	.257	.337	.191	.200	.168
Pursue an academic track	.583	.533	.998	1.000	.994
Specialize in science	.304	.236	.635	.587	.752
Observations	209,654	7,032	556	395	161

NOTE.—Column 1 shows the average characteristics of general education tenth-grade students in the Paris region. Column 2 shows the same characteristics for general education tenth-grade students in the 12 high schools of the experiment, and col. 3 shows them for the volunteer students in these 12 high schools. Columns 4 and 5 further show the characteristics of volunteer students in each cohort. The ability score corresponds to the average grade (standardized at the Paris region level) obtained on the national middle school exit exams taken at the end of grade 9. The table can be read as follows: 60.1% of volunteer students are female, and 25.5% come from a family with a high socioeconomic status (SES) background. At the end of grade 10, 99.8% choose to pursue an academic track, and 63.5% choose to specialize in science. Their average standardized score at the end-of-middle-school exams (i.e., grade 9) is 0.156.

D. Random Selection of Eligible Students

For both cohorts, the randomization took place in each school just before the start of the summer holidays and just after the principal preassigned each volunteer student to one future eleventh-grade class (based on her choice of field of study). In French high schools, each eleventh-grade class corresponds either to students who specialize in science at the end of grade 10 or to students who specialize in humanities (i.e., social sciences or languages/literature). The randomization was conducted at the class level and stratified by major field of study (science/humanities). In the French system, students stay in the same class throughout the academic year and in every subject. The class is therefore a very distinct and close entity where most interactions with same-age students take place.

For each major field of study, half of the classes—or half rounded up to the nearest integer when there were an odd number of classes—were put in the treatment group. In the end, we have 305 volunteer students in the treatment group and 251 volunteer students in the control group. Only the 305 volunteer students in the treatment group were eventually invited to participate in the program. Most of the results in this paper are based on the comparison of volunteer students in treatment and control groups. Under the assumption that volunteer students in control groups remain unaffected by the treatment

(SUTVA), this comparison provides an estimate of an intention-to-treat parameter, namely, the impact of being invited to participate in the program on the subsequent outcomes of volunteer students.

To assess the similarity between the control and treatment groups, table A.1 (tables A.1–A.12 are available online) builds on the information provided by administrative registers to compare the sociodemographic characteristics of volunteer students in the treatment and control groups (in terms of gender, grade repetition, parental occupation, and pretreatment grades). We find no significant differences between the two groups. To further test for the similarity of the two groups, table A.2 compares the responses of treatment and control groups to the baseline questionnaire that students had to fill out in order to be identified as volunteers. Again, we find little difference in responses across the two groups.

Finally, we augmented our data set with administrative information on teachers' gender, number of years of experience, weekly number of teaching hours, and highest level of educational qualification.[7] Building on this information, table A.3 shows that there is no significant difference in gender, level of experience, teaching hours, or educational qualification across teachers assigned to treated or control classes.

E. High-Ability versus Mid-High-Ability Volunteers

As discussed above, the vast majority of volunteer students are good students, but not all of them are top students. Assuming that the intervention contributes to distract students from curriculum-related learning, it may have very different effects on students with different ability levels, that is, with different stocks of curriculum-related knowledge pretreatment. To test for such heterogeneous effects, most of our regression analysis will be conducted not only on the full sample of volunteers but also separately on the half of the sample with the strongest academic records pretreatment as well as on the half with the weakest academic records pretreatment.[8] We replicate

[7] With respect to teachers' educational qualification, the main distinction is between *agrégés* and *certifiés*. In France, to become a secondary school teacher, students have to take either the *agrégation* exam (to become *agrégés*) or the *certificat d'aptitude au professorat de l'enseignement du second degré* (CAPES) exam (to become *certifiés*). Both are competitive exams, but the *agrégation* is more difficult and selective than the CAPES. It is typically taken after 4 years of higher education, whereas CAPES is normally taken after 3 years. Teachers who are *agrégés* have access to better career opportunities, with fewer teaching hours and higher wages.

[8] To define the two ability groups, students are ranked first according to whether they already repeated a grade (grade repeaters are in the lower-ability group) and second according to their average marks during tenth grade (i.e., the year before the treatment). In this setup, the higher-ability subsample consists of students who received the best average marks among students who never repeated a grade. The two ability groups are constructed so as to have the same size.

the comparison of the pretreatment characteristics of treated and control students separately for the two ability groups in panels B and C of table A.1. We do not detect any significant pretreatment difference across treatment and control students within both ability groups.

Table A.4 also shows the proportion of higher- and lower-ability volunteers in each decile of the distribution of ninth-grade scores in their high school.[9] It confirms that a large majority of higher-ability volunteers are in the two top deciles of this distribution, whereas a majority of lower-ability volunteers are in deciles three to six. In the remainder, we will refer to the first group as the "high-ability group" and to the second group as the "mid-high-ability group."

Table A.5 further compares the responses of the two ability groups to the baseline questionnaire. The table suggests that students from the higher-ability group can get help more easily for homework. They also tend to provide a better assessment of their own academic ability. They seem to have more areas of interest and declare more often being interested in social issues as well as in national or international issues. Finally, they are more likely to think that they know about CPGE programs.

F. Program Content, Tutors, and Take-Up

Generally speaking, the objective of the ENS program is to improve students' academic achievement and readiness for higher education. Just after the randomization, the program managers sent a letter to the students who participated in the random draw to inform them of the results of the draw and to invite those in the treatment group to participate in an introductory week (called campus week) organized at the ENS in late August, just before the start of the academic year.

During this introductory week, students have activities led by tutors from previous cohorts. The objective is to help students improve their methods of work and to prepare them for the grade 11 academic program, which is much more demanding than grade 10. During the first 2 days of the week, students attend courses that mimic grade 11 courses. During the third day, they take exams. The fourth and fifth days are dedicated to correcting exam errors and providing students with feedback on their work. Students also have the opportunity to take additional method courses on how to write essays. Meals and accommodation are paid for by the program. In 2010 and 2011, the vast majority of students participated in this week (see table 2).

After the introductory week, students are asked to choose a theme for the tutoring sessions in which they are going to be involved throughout the academic year. The tutoring sessions are designed to touch on topics that are absent from the high school curriculum but representative of higher education

[9] We cannot show their proportions in the deciles of the distribution of tenth-grade scores because tenth-grade scores are observed for volunteer students only.

Table 2
Take-Up Rates, by Cohort and Ability Group

	All (1)	Mid-High Ability (2)	High Ability (3)
	A. Cohort 2010		
Introductory week (year 1)	.792	.709	.877
More than 10 sessions attended (year 1)	.645	.618	.675
Reenlistment	.569	.482	.660
Observations	216	110	106
	B. Cohort 2011[a]		
Introductory week (year 1)	.742	.744	.739
Reenlistment	.483	.395	.565
Observations	89	43	46

NOTE.—The sample includes eligible students from cohorts 2010 (panel A) and cohort 2011 (panel B). In each panel, col. 1 shows take-up rates for the full sample, whereas col. 2 (col. 3) shows take-up rates for the mid-high-ability (high-ability) group. The high-ability group corresponds to the top half of the distribution of pretreatment ability scores across volunteers, whereas the mid-high-ability group corresponds to the bottom half. The table can be read as follows: among eligible students from cohort 2010, 79.2% attended the (year 1) introductory week at the École Normale Supérieure, 64.5% attended more than 10 tutoring sessions during the first year of the program, and 56.9% reenlisted at the end of the first year.

[a] For cohort 2011, information on attendance was not collected.

curricula. There are three possible themes: science, social sciences (including history), and literature. Students are also asked to list one or two friends with whom they would like to be grouped for these sessions. These lists are used by program managers to define small groups of three to seven students with similar thematic preferences.[10] Eventually, the program managers randomly assigned each group of students to one of the tutors specialized in their theme.

Each year, about 80% of tutors are new ones, whereas 20% have already been involved in the program. We have information on the gender of tutors (40% are women) as well on their fathers' occupation: only a minority of fathers (17%) do not belong to the top occupational category of the French (one-digit) classification, which reflects the very strong overrepresentation of students with a privileged family background.

[10] Specifically, we defined 1 seven-person group, 24 six-person groups, 10 five-person groups, and 1 four-person group in 2009–10. In 2010–11, the average group size and the participant-tutor ratio were lower. Specifically, we had 1 seven-person group, 6 six-person groups, 6 five-person groups, 7 four-person groups, and 10 three-person groups in 2010–11. In cohort 2010 (our first cohort), the program managers first define 59 triplets (and 13 doublets) of students with similar thematic preferences. The managers then randomly paired these 72 subgroups within each theme. By doing so, they created 36 groups of four to six persons. Two students were not included in the definition of triplets and doublets because they were supposed to leave the program. Eventually, they remained in the program and were randomly assigned to two different groups (one of them was a six-person group, which explains why we end up with a seven-person group). In the last section of the paper, we build on this randomized design to explore the role of peer groups.

New tutors are recruited in early October at the beginning of the academic year. They are all ENS students. They first benefit from a 2-day training session where program managers provide them with information on the objectives of the program and on the type of high school students they are going to tutor. Tutors who have already been involved in the program also participate in this training session so as to share their experience with new tutors. The tutoring program starts at the end of October, with a first meeting (at the ENS) between tutors and their groups of six students. The team of tutors meets three times a year to share experience and get feedback from program managers.

The academic year is divided into three terms (September–December, January–March, and April–June). Each term is dedicated to a specific topic. During the first year of treatment, students from cohort 2010 were given the opportunity to participate in four thematic sessions per term. They benefited from an additional session of personal coaching in the first term as well as a session of improvisation theatre in the second term to develop their oral skills. All in all, students from cohort 2010 were given the opportunity to participate in 12 tutoring sessions and two additional activities during the first year. This number of sessions was deemed excessive by a number of students, and it was reduced for the second cohort to two tutoring sessions per term, with only one additional cultural outing between the two sessions of each term—that is, a total of nine sessions for the second cohort (instead of 14 for the first cohort). We have information (collected by tutors) on attendance at the 12 tutoring sessions organized for the first cohort of students. This information suggests that attendance rates were high: about two-third of students participate in 10 (or more) sessions out of 12 (table 2).

At the end of grade 11, participants are asked whether they want to pursue the program in grade 12. A majority chose to do so (57% for the first cohort, 48% for the second one). It should be emphasized, however, that the proportion of students who choose to stay in the program is about 18 percentage points higher for high-ability than for mid-high-ability students. As shown in the following sections, one potential reason for such a gap may be that the intervention has in fact very different effects on the two groups of students. At the start of the second year of treatment (grade 12), most students change tutors. As far as students from the first cohort (2010) are concerned, they benefited during this second year from seven thematic sessions with their new tutors as well as from one cultural outing and from a 1-day informational meeting on higher education. Also, starting in March, six additional sessions were organized to help them prepare for the high school exit exams taken at the end of grade 12. With respect to students from the second cohort, they were given the opportunity to participate in four thematic sessions (and two cultural outings) during the two first terms of the second year of treatment. During the third and last term, they benefited from two additional sessions dedicated to the preparation of the high school exit exams.

IV. Conceptual Framework

In this section, before moving on to the empirical analysis, we develop a conceptual framework that helps clarify the nature of the treatment and why treatment effects may not be necessarily the same across ability groups. In our framework, students' achievement depends not only on what is learned at school but also on what is learned outside school (typically from the family). Participation in the program is interpreted as inducing the substitution of the second type of input for the first one. Under standard assumptions about the education production function, we show that such a substitution may have very different effects on high- and mid-high-ability students.

A. Setup and Notation

We assume that students' achievement at the end of high school depend on two types of knowledge. The first type is mainly transmitted by teachers either in the classroom or through homework. In our context, it corresponds to the curriculum of secondary education. The second type of knowledge is mainly transmitted outside the classroom, typically by the family. This includes knowledge about the various tracks available after high school and about the admission requirements for these tracks. This type of (extracurricular) knowledge helps students identify tracks that fit their taste and academic aptitudes. It also helps them formulate education plans and stay focused at school.

At the end of tenth grade, we assume that student i is endowed with a stock K_{0ci} of curricular knowledge and with a stock K_{0fi} of extracurricular knowledge. We denote T_{ci} as the amount of time devoted to further accumulate the first type of knowledge during grade 11 and 12. Similarly, we denote T_{fi} as the amount of time devoted to further accumulate the second type of knowledge. For students coming from a low social background, T_{fi} is typically very low, and this is precisely the problem that the TALENS program aims to solve. After normalization, the budget-time constraint can be written as $T_{ci} + T_{fi} = 1$. For the sake of simplicity, we assume that there is a one-to-one relationship between the time devoted to accumulate a given type of human capital and its actual accumulation, so that we have,

$$K_{ci} = K_{0ci} + T_{ci} \quad \text{and} \quad K_{fi} = K_{0fi} + T_{fi}, \tag{1}$$

where K_{ci} (K_{fi}) represents the stock of curricular (extracurricular) knowledge accumulated at the end of high school. Eventually, we denote Y_i as the achievement of student i at the end of high school, and we assume that

$$Y_i = F(K_{ci}, K_{fi}) = F(K_{0ci} + T_{ci}, K_{0fi} + T_{fi}), \tag{2}$$

where F represents a strictly quasi-concave production function. Assuming strict quasi concavity ensures that the marginal rate of substitution of K_c for

K_f is strictly decreasing with K_c/K_f. As made clear below, this is the only assumption needed to explain that the program does not necessarily have the same effect on higher- and lower-ability students.[11]

In this paper, we focus on tenth graders who are induced by their school principal to participate in the TALENS program and who are willing to participate in this program. These volunteer students all come from a socially disadvantaged family: we assume that they all have the same very low initial level K_{0f} of extracurricular knowledge and that their family has no means to further increase this stock ($T_{fi} = 0$). By contrast, these students do not all have the same initial level K_{0ci} of curriculum-related knowledge, even though this level is necessarily above a certain minimum.[12] For the sake of simplicity, we will assume that there are only two types of students, one with a relatively high initial level of curriculum-related knowledge (hereafter, high-ability students; $K_{0ci} = K_{0H}$) and one with a relatively low initial level ($K_{0ci} = K_{0L} < K_{0H}$).

B. Interpretation of Treatment Effects

Once the list of volunteer students is finalized, half of them are randomly drawn and become eligible to the program (hereafter, the treatment group). The other half represents the control group.

We assume that the program is designed so as to improve participants' level of noncurricular knowledge. Specifically, we assume that participation in the program induces an increase T in the stock of this type of knowledge.

In this setup, consider a student i whose initial endowment is (K_{0ci}, K_{0f}). If she is assigned to the control group, there is no specific constraint on the time she can allocate to increasing K_{0ci}. Assuming that she seeks to maximize her achievement, she devotes all of her efforts to further increase this stock of curriculum-related knowledge, and we can assume $T_{ci} = 1$. Consequently, the increase in achievement between grade 10 and grade 12 is written

$$\Delta_{\text{cont},i} = F(K_{0ci} + 1, K_{0f}) - F(K_{0ci}, K_{0f}). \tag{3}$$

By contrast, if the same student is assigned to the treatment group, she has to allocate T to increasing K_{fi}, and, as a consequence, she can allocate no more than $(1 - T)$ to increasing K_{ci}. The increase in achievement between grade 10 and grade 12 is now written

[11] Note that imposing strict quasi concavity on the production function amounts assuming that there is some complementarity between the two inputs. When extracurricular and curriculum-related knowledge are perfect substitutes, the substitution of one type of knowledge for another type has the same impact on all students, regardless of their initial endowment in extracurricular or curriculum-related knowledge.

[12] As discussed below, this minimum level is set by the principal. It reflects the principal's prior about the minimum level of ability that is required to take advantage of the program.

$$\Delta_{\text{treat},i} = F(K_{0ci} + 1 - T, K_{0f} + T) - F(K_{0ci}, K_{0f}). \quad (4)$$

Overall, the impact of being assigned to the treatment group rather than the control group can be written

$$\begin{aligned}\Delta_i &= \Delta_{\text{treat},i} - \Delta_{\text{cont},i} \\ &= F(K_{0ci} + 1 - T, K_{0f} + T) - F(K_{0ci} + 1, K_{0f}).\end{aligned}$$

Eventually, assuming that T is small, we have

$$\Delta_i \approx T(F_f'(K_{0ci} + 1, K_{0f}) - F_c'(K_{0ci} + 1, K_{0f})), \quad (5)$$

where F_c' denotes the marginal product of curriculum-related knowledge input and F_f' denotes the marginal product of the extracurricular one.

Hence, the first-order impact of the intervention on achievements depends on whether the marginal rate of technical substitution F_f'/F_c' is larger or smaller than 1. Specifically, the impact is positive when F_f'/F_c' is greater than 1 and negative when F_f'/F_c' is lower than 1. Under the assumption that F is strictly quasi concave, the marginal rate of technical substitution is strictly decreasing with the K_c/K_f ratio, and we may have a positive impact for relatively high values of K_c/K_f and a negative impact for relatively low values of K_c/K_f. In this scenario, Δ_i may well be negative for low initial values of K_{0ci} and positive for high initial values of K_{0ci}—that is, it may be negative for mid-high-ability students and positive for high-ability ones.[13] The next sections explore these issues empirically.

Eventually, appendix B (apps. A and B are available online) develops two possible extensions of our baseline model to assess the robustness of our theoretical predictions. The first extension explores the case where the mapping between time investment and knowledge accumulation is not necessarily one to one anymore (as in eq. [1]) but may vary across the different types of knowledge and ability groups. In such a case, one additional reason for why the program may differentially affect the performance of higher- and lower-ability students is that higher- and lower-ability students may not be equally equipped to take advantage of the time invested in the program.

The second extension explores the case where school-related efforts are costly and where students take these costs into account when deciding their optimal investment strategies. Assuming that the marginal costs of school-related efforts increase more rapidly for lower-ability students, participation

[13] Assuming, e.g., that F is a constant elasticity of substitution production function (with $F(x, y) = A(\lambda x^{-\alpha} + (1 - \lambda)y^{-\alpha})^{-1/\alpha}$ and $\alpha > -1$), we can check that the impact is positive for higher-ability students and negative for low-ability students if and only if $(K_{0L} + 1)/K_{0f} < [\lambda/(1 - \lambda)]^{\sigma} < (K_{0H} + 1)/K_{0f}$, where $\sigma = 1/(1 + \alpha)$ represents the elasticity of substitution. By contrast, when F is linear (which amounts to assuming $\alpha = -1$), the two inputs are perfect substitutes and the impact of the program is the same (and equal to $(2\lambda - 1)T$) for both groups of students.

in the program may induce an endogenous decline in the amount of time invested in curricular-related activities, which is stronger for lower-ability students. This is another reason why the program may differentially affect lower- and higher-ability students.

V. Data

The high schools participating in the experiment provided us with the ID number of their volunteer students as well as their average academic achievement in grade 10 (which we use to define our two ability groups). Using ID numbers, we first augment this data set with exhaustive administrative data on students' performance on the national exam taken at the end of ninth grade (the end-of-middle-school exam called the *diplôme national du brevet*) as well as on the national exam taken at the end of high school (*baccalauréat*). We were also able to augment our initial data set with administrative school registers (*bases centrales scolarité*), which provide information on the major field of study chosen by students at the end of grade 10 as well as on the schools and classes attended in the following years.

In the remainder of the paper, we mainly focus on the sample of 556 students who were included in the randomization procedure. The main dependent variable will be their results at the high school national exit exam (externally set and marked, in grades 11 and 12). The main independent variables will be their treatment status, their average marks at the end of the middle school national exam (externally set and marked, in grade 9), and their average marks in grade 10 (as assessed by teachers). Generally speaking, missing rates are very small and unrelated to students' treatment status. In particular, information on exit scores is available for 97% of the observations (14 missing, six treated, and eight controls), whereas information on high school graduation is available for 100% of the observations.

To further assess the similarity between the control and treatment groups, we were able to use the information coming from the questionnaire that students had to fill out to be identified as a volunteer. As discussed above, this pretreatment survey provides us with information on volunteer students' family background, their preferred extracurricular activities, their school record and school background, their plans for the future, their level of information about higher education institutions, and so on.

Finally and as mentioned above, we were also able to enrich our data set with administrative data on teachers' gender, level of education, weekly number of teaching hours, and number of years of experience.

VI. Effects on Achievement and Choices

In this section, we analyze the effect of the intervention on students' performance on high school exams taken at the end of grades 11 and 12. We focus

on the national exams (externally set and marked) that students have to take to graduate high school. For each exam, we estimate the following model:

$$Y_i = \alpha T_i + \beta X_i + \nu_i, \tag{6}$$

where, for each student i, variable Y_i represents the mark obtained at the exam (or a dummy variable indicating whether i passed the exam), variable T_i is a dummy indicating whether i is in the treatment group, and X_i is a vector of pretreatment control variables that includes dummies for gender, grade repetition, family background, and pretreatment marks as well as school fixed effects and major choice fixed effects. Variable ν_i represents unobserved error terms. The parameter of interest is α. Identification is a direct consequence of the experimental nature of the treatment assignment variable T_i. Standard errors are clustered at the class level.

A. Effects on High School Achievement

Table 3 shows the effect of the intervention on scores on exit exams taken at the end of grade 11 (col. 1), on scores on exit exams taken at the end of grade 12 (col. 2), and on average scores on exit exams (col. 3). The table also shows the impact of the intervention on the probability of high school graduation on time (col. 4) as well as on the probability of graduating at all (i.e., including after repeating grade 12; col. 5).[14]

The first panel refers to the full sample of volunteer students. We do not find any significant effect on the different outcomes in this sample. The intervention has no impact on the average grades obtained at the high school exams by volunteer students or on their high school graduation rate. The second panel of table 3 refers to the half of the sample of volunteers with the strongest academic records pretreatment (i.e., in grade 10), whereas the third panel refers to the half of the sample of volunteers with the weakest academic records pretreatment. As emphasized above, both groups correspond to students whose scores at the end-of-middle-school exam (in grade 9) are in the top half of their high school. But the majority of students in the first group are in the two top deciles (high-ability students), whereas the majority of students in the second group are in the deciles near the median or just above the median of their high school (mid-high-ability students).

The panels reveal that the intervention has completely different effects on these two groups of students: it contributes to a decrease in exit scores for the lower-ability group and an increase in exit scores for the higher-ability group. As discussed in the previous sections, the program contributes to substituting extracurricular activities for curriculum-related ones, and our regression

[14] Graduation requires a minimum average score of 10/20. The vast majority of students who fail to graduate at the end of grade 12 are allowed to repeat grade 12 and to retake grade 12 exams.

Table 3
Effect of the Treatment on Performance on High School Exit Examinations

	Dependent Variable				
	Grade 11 Average Score (1)	Grade 12 Average Score (2)	Overall Average Score (3)	Graduation on Time (4)	Graduation (5)
	A. All Volunteer Students				
Treatment	−.096	.025	.013	.016	.001
	(.075)	(.080)	(.078)	(.041)	(.026)
Observations	542	542	542	556	556
Mean dependent variable	.665	.343	.413	.725	.916
	B. Mid-High Ability				
Treatment	−.254**	−.165	−.185	−.074	−.034
	(.103)	(.137)	(.131)	(.063)	(.045)
Observations	276	276	276	285	285
Mean dependent variable	.314	.101	.138	.667	.902
	C. High Ability				
Treatment	.025	.230**	.221**	.124**	.074**
	(.114)	(.111)	(.108)	(.048)	(.026)
Observations	266	266	266	271	271
Mean dependent variable	1.055	.613	.719	.790	.933
	D. Differential Impact				
Treatment × high ability	.279*	.396**	.406**	.198**	.108**
	(.150)	(.183)	(.177)	(.075)	(.051)
Observations	542	542	542	556	556
Mean dependent variable	.741	.512	.581	.123	.031

NOTE.—The sample includes volunteer students from cohorts 2010 and 2011. The table shows the results from reduced-form regressions in which variables measuring performance on high school exit examinations (*baccalauréat*) are regressed on a treatment dummy, using students' gender, pretreatment ability score, and socioeconomic family background as control variables. Column 1 shows the estimated effect of the treatment when the dependent variable is the average grade obtained on examinations taken at the end of grade 11, col. 2 shows the estimated effect when the dependent variable is the average grade obtained on examinations taken at the end of grade 12, and col. 3 shows the estimated effect when the dependent variable is the average grade across all examinations. Column 4 shows the estimated effect when the dependent variable is a dummy indicating high school graduation on time, and col. 5 shows the estimated effect when the dependent variable is a dummy indicating high school graduation at any time. Panel A refers to the full sample, panel B refers to the mid-high-ability subsample, and panel C refers to the high-ability subsample. The high-ability group corresponds to the top half of the distribution of pretreatment ability scores across volunteers, whereas the mid-high-ability one corresponds to the bottom half. Finally, panel D shows the estimated difference in the effect of the treatment between mid-high- and high-ability students. To estimate this difference, we use the full sample and regress the dependent variable on the interaction between a dummy indicating treatment and a dummy indicating high ability, controlling for the treatment dummy as well as for the full set of sociodemographic controls and their interactions with the high-ability dummy. Standard errors clustered at the class level are reported in parentheses.

* $p < .10$.

** $p < .05$.

analysis confirms that this may have very different effects on students endowed with different levels of curriculum-related knowledge.[15] The effects on mid-high-ability students is less significant on grade 12 exams than on grade 11 exams, which is consistent with the fact that a majority of these students drop out of the program at the end of grade 11. By contrast, the effect on high-ability students tend to be more significant on grade 12 exams, which is consistent with the fact that a large majority of higher-ability students specialize in science and that science exams are taken at the end of grade 12.[16] In figure A.1 (available online), we provide a graphical representation of changes in treatment effects on standardized exam scores across grades and ability groups.

The last panel of table 3 confirms that the differences between the two sets of impacts are significant at the standard level: the intervention contributes to a significant increase in the academic gap between the two ability groups. Specifically, differences in grades obtained at the end of grade 11 or grade 12 as well as differences in high school graduation probability between higher- and lower-ability volunteers are significantly more important in the treatment group than in the control group. For example, the difference in the probability of high school graduation on time is about 12 percentage points in the control group, but this gap becomes about 20 percentage points larger in the treatment group—that is, almost a tripling of the gap occurs within the group that was given the opportunity to participate in the program.[17]

For each one of the five available measures of performance on high school exams, table 3 tests two null assumptions, one per ability subgroup. A well-known issue is that such subgroup comparisons tend to mechanically increase the likelihood of finding significant effects (see, e.g., List, Shaikh, and Xu 2016). In table A.6, we report for each one of the five variables the corresponding unadjusted p-values as well the p-values adjusted to take into account

[15] We have checked that when we focus on the sample of students who were in the top half of their class and above the 11/20 threshold pretreatment (as we initially intended to do), we obtain impacts that are almost as strong as those obtained with the higher-ability sample, that is, a positive impact of 18% of a standard deviation on exit score and a positive impact of 7.5 percentage points on graduation rate.

[16] Seventy-eight percent of high-ability students specialize in science, vs. 46% of mid-high-ability students. It should be noted, however, that differences in estimated treatment effects across exams taken in grade 11 and grade 12 are not significant at the standard level, so that these differences should be interpreted cautiously.

[17] It should be noted that it is not uncommon to find heterogeneous treatment effects across ability groups (see, e.g., Fryer, Devi, and Holden 2012). Also it is not unusual for extracurricular interventions to have relatively large impacts on high school students, especially when they target students who likely lack family support (Hoxby and Turner 2013; Castleman, Page, and Schooley 2014). For example, Hoxby and Turner (2013) analyze an intervention that provides information on the college application process to high-achieving low-income students. This very simple intervention increases admissions at more selective colleges by about 30%.

(ability) subgroup comparisons, using Holm's (1979) method. Generally speaking, adjusted and unadjusted tests provide qualitatively similar results, even though multiple testing yields larger p-values. For example, when we focus on the most synthetic measure of high school performance—namely, high school graduation—the unadjusted test rejects the null assumption for the high-ability group at $p < .01$, whereas the adjusted test rejects the null assumption for the same ability group at $p < .02$.

It is also possible to jointly consider the four null hypotheses defined by the two ability groups and the two elementary exit scores (i.e., on grade 11 exams and on grade 12 exams), so as to test the robustness of our results about the timing of the impact of the treatment on the two ability groups. When we consider this family of four null hypotheses, the multiple testing approach yields more important corrections. Specifically, the adjusted p-value suggests that the negative effect on the grade 11 scores of lower-ability students is significant at $p < .07$ (unadjusted $p < .02$), whereas the positive effect on the grade 12 scores of higher-ability students is significant at $p < .13$ (unadjusted $p < .05$). Hence, the finding that lower-ability students are mostly affected during grade 11 and higher-ability students are mostly affected during grade 12 appears to be less robust to multiple testing.

To further explore the robustness of our findings, table A.7 shows the regression results separately for the two successive cohorts. Comfortingly, we observe an increase in the gap between lower- and higher-ability volunteers within the treatment group for both cohorts. The difference in the probability of high school graduation on time between the two ability groups increases by about 17 percentage points in the first cohort and by about 29 percentage points in the second cohort. Because of the small size of the second cohort, it is not possible, however, to assess whether the increase in the gap in graduation rates between the two ability groups is more significant in the first or in the second cohort.

High-ability students choose more often to specialize in science, and one explanation for our results may be that tutoring is more efficient for students who specialize in science. To test this assumption, tables A.8 and A.9 show the impact of the intervention on the grades obtained on the high school national examination in each subject (French, math, physics, languages, etc.) for each major choice (science/humanities) and ability group. For students who specialize in science, the intervention contributes to an increase in the gap between the two ability groups in all subjects, the estimated increase being significant at the standard level in French (+43% of a standard deviation), languages (+53%), and biology (+49%). For students who specialize in humanities, the gap increases in a majority of subjects, even though the estimated increases are not significant at the standard level, which reflects the relatively small number of observations in the humanities subsample.

Overall, the increased gap in achievement between the two ability groups does not seem to be driven by a specific field of study or a specific subject.

This finding suggests that the intervention did not affect subject-specific inputs but did affect more general determinants of performance at school. One such determinant is likely the amount of time devoted to school homework, revision exercises, and preparation for tests. Participation in the program is time-consuming, and it likely contributes to a reduction in the amount of time that participants are able to devote to these activities. It may be detrimental in all subjects, especially for students who are not among the very best ones. We will come back to these issues in the last section of the paper.

B. Effects on Access to Selective Undergraduate Programs

One of the objectives of the program was to increase the proportion of students who gain admission into the most selective undergraduate programs in France, namely, the CPGE. It may be that the intervention has no average effect on high school grades but contributes nonetheless to an increase in the overall number of students from underprivileged high schools who are aware of the existence of CPGE programs and aspire to get admitted into one of them. Most tutors got access to the ENS after 2 years spent in a CPGE program: it is certainly the undergraduate program they know the best and about which they are able to provide the most comprehensive information.

To shed light on these issues, table 4 shows the effect of the intervention on the proportion of students who gain admission into a CPGE program after high school (col. 1) as well as on the proportion who are still in a CPGE program 2 years after high school graduation (col. 2).[18] The table shows no significant effect on the proportion of students who gain admission into a CPGE program after high school or on the proportion who are still in a CPGE program 2 years after high school, consistent with the intervention having on average no effect on education aspirations or on the ability to persist in this type of program.

When we replicate this analysis on the lower-ability group, we find a significant negative effect on the proportion of students gaining admission into CPGE programs (-6.7 percentage points on year 1 enrollment, which corresponds to about a -50% decrease in this proportion). This strong negative impact is likely a direct consequence of the negative impact of the program on high school achievement for this subgroup of participants. We find a similar negative effect on year 2 enrollment as on year 1 enrollment. This result suggests that those who have been induced by the intervention not to go into a CPGE program would have in fact been able to succeed in this program had they not been treated.

Finally, when we replicate the same analysis on the higher-ability group, we find positive effects on enrollment in both year 1 and year 2. In terms of magnitude, these positive effects are almost as large as the negative effects on

[18] CPGE students are not allowed to repeat year 1, and about 23% drop out of the program before the end of year 1.

Table 4
Effect of the Treatment on Access to Selective Undergraduate Programs (CPGE)

	Major Field of Study					
	All Majors		Science Major		Humanities Major	
	Year 1 (1)	Year 2 (2)	Year 1 (3)	Year 2 (4)	Year 1 (5)	Year 2 (6)
	A. All Volunteer Students					
Treatment	−.014	−.015	.011	−.004	−.063	−.033
	(.030)	(.023)	(.039)	(.031)	(.041)	(.026)
Observations	556	556	353	353	203	203
Mean dependent variable	.163	.112	.156	.123	.175	.093
	B. Mid-High Ability					
Treatment	−.067**	−.058**	−.070*	−.075**	−.072*	−.044*
	(.030)	(.022)	(.039)	(.034)	(.042)	(.025)
Observations	285	285	138	138	147	147
Mean dependent variable	.136	.083	.115	.098	.155	.070
	C. High Ability					
Treatment	.055	.044	.076	.057	−.043	−.013
	(.054)	(.039)	(.059)	(.043)	(.133)	(.059)
Observations	271	271	215	215	56	56
Mean dependent variable	.193	.143	.183	.140	.231	.154
	D. Differential Impact					
Treatment × high ability	.122**	.102**	.146**	.132**	.029	.031
	(.060)	(.044)	(.068)	(.053)	(.118)	(.057)
Observations	556	556	353	353	203	203
Mean dependent variable	.057	.060	.068	.042	.076	.084

NOTE.—The sample includes volunteer students from cohorts 2010 and 2011. The table shows the results from regressions in which the probability to enter a *classes préparatoires aux grandes écoles* (CPGE) program (year 1) as well as the probability to still be in a CPGE program 2 years after high school graduation (year 2) are regressed on a treatment dummy, using students' gender, pretreatment ability score, and socioeconomic family background as control variables. Panel A refers to the full sample, panel B refers to the mid-high-ability subsample, and panel C refers to the high-ability subsample. The high-ability group corresponds to the top half of the distribution of pretreatment ability scores across volunteers, whereas the mid-high-ability one corresponds to the bottom half. Finally, panel D shows the estimated difference in the effect of the treatment between mid-high- and high-ability students. To estimate this difference, we use the full sample and regress the dependent variable on the interaction between a dummy indicating treatment and a dummy indicating high ability, controlling for the treatment dummy as well as for the full set of sociodemographic controls and their interactions with the high-ability dummy. Within each panel, cols. 3 and 4 refer to students who specialize in science at the end of grade 10, whereas cols. 5 and 6 refer to students who specialize in humanities. Standard errors clustered at the class level are reported in parentheses.

* $p < .10$.

** $p < .05$.

lower-ability participants, but they are not significant at the standard level. This improvement is likely driven by their increased academic performance in high school. Overall, there is no effect on the overall proportion of eligible students in CPGE programs but a significant increase in the difference between the two ability groups (+12.2 percentage points increase).

There are two basic types of CPGE program, one specialized in science and one specialized in humanities. Columns 3–6 of table 4 explore whether the effect of the intervention is different across these two program types. They show that the increased gap in CPGE enrollment across ability groups is mainly driven by CPGE programs specialized in science, which are also the most selective. For humanities, we observe a decline in enrollment for both ability groups (although it is not significant at the standard level for the higher-ability one) but no significant change in the gap.

With respect to multiple testing, we checked that the significant negative effect on the proportion of lower-ability students admitted into CPGE programs is robust to adjustments that take into account subgroup comparisons (see again table A.6). We also jointly considered the four null assumptions defined by the two ability subgroups and the two most synthetic outcomes analyzed in this paper, namely, high school graduation and access to the second year of a CPGE program. Both adjusted and nonadjusted tests reject the same two assumptions at $p < .05$, namely, the assumption that there is no effect on high-ability students' graduation probability and the assumption that there is no effect on lower-ability students' probability to get access to the second year of a CPGE program.

VII. Mechanisms: The Role of Tutors and Peers

The previous sections suggest that the program has positive effects on volunteers with the best academic level pretreatment but negative effects on the other (mid-high-ability) volunteers. One important question, however, is whether the intervention induces the same positive and negative effects regardless of the tutors recruited to implement the program. If all tutors had the same negative effect on lower-ability students, it would suggest that the negative effect is mainly related to some of the deep features of the program and not to the way it is implemented. By contrast, if the negative effect was found only for some specific tutors, the problem would also likely be in the way the program is implemented. A better selection (or training) of tutors would be a way to improve the program.

A. Tutors' Characteristics

To explore this issue, it is possible to build on the fact that students were randomly assigned to their first-year (grade 11) tutor. In this setup, the difference in outcomes observed at the end of grade 11 between eligible students assigned to different types of tutors likely provides an evaluation of the effect of tutors. We have information on the gender and family socioeconomic background of tutors. The design of our intervention makes it possible to test whether it makes a difference to be assigned to one type of tutor rather than to another one.

Table 5 implements this test. It shows the results of regressing the grade 11 score on a treatment dummy and on the interactions between this treatment

Table 5
Effect of Tutors' Gender and Family Background on Grade 11 Scores

	(1)	(2)	(3)	(4)
	A. All Volunteer Students			
Treatment	−.096	−.119	−.317**	−.361**
	(.075)	(.095)	(.079)	(.098)
Treatment × female tutor		.069		.087
		(.145)		(.145)
Treatment × higher-background tutor			.311**	.325**
			(.147)	(.149)
Observations	542	542	523	523
	B. Mid-High-Ability Students			
Treatment	−.254**	−.228*	−.405*	−.361
	(.103)	(.117)	(.112)	(.123)
Treatment × female tutor		−.078		−.086
		(.172)		(.196)
Treatment × higher-background tutor			.257	.244
			(.225)	(.235)
Observations	276	276	262	262
	C. High-Ability Students			
Treatment	.025	.003	−.329	−.394
	(.114)	(.139)	(.117)	(.142)
Treatment × female tutor		.069		.126
		(.246)		(.233)
Treatment × higher-background tutor			.410	.437*
			(.253)	(.244)
Observations	266	266	261	261

NOTE.—The sample includes volunteer students from cohorts 2010 and 2011. The table shows the results of regressing the average score obtained at exams taken at the end of grade 11 on a treatment dummy as well as on interactions between a treatment dummy and dummies indicating either the gender or the family background of the tutor. We use students' gender, pretreatment ability score, and socioeconomic family background as control variables. Panel A refers to the full sample, panel B refers to the mid-high-ability subsample, and panel C refers to the high-ability subsample. The high-ability group corresponds to the top half of the distribution of pretreatment ability scores across volunteers, whereas the mid-high-ability one corresponds to the bottom half. Standard errors clustered at the class level are reported in parentheses.
* $p < .10$.
** $p < .05$.

dummy and dummies indicating the gender of the tutor (male/female) and the family background of the tutor (lower/higher family background), controlling for the same basic set of pretreatment variables as in table 3. Regressions are conducted on the full sample as well as on the two ability subsamples.[19]

The table does not show any significant difference in the effect of the treatment across male and female tutors. But it suggests that the program produces

[19] Another possibility is to focus on eligible students and to regress their scores at exit exams directly on dummies indicating the gender and family background of their tutors. This specification yields similar results.

significantly better outcomes when it is implemented by tutors with a higher socioeconomic background, especially when we focus on higher-ability recipients. It suggests that tutors with a higher socioeconomic background contribute to improving the overall impact of the intervention but also contribute to widening the gap between lower- and higher-ability recipients. One reason for the more positive impact of tutors with a higher socioeconomic background may be that they are less close to tutees and better able to act as real teachers.

B. Peer Group Influence

The results in table 5 provide some evidence that tutors matter. They suggest that a better selection (or training) of tutors can be a way to improve the efficiency of the intervention. We also explored whether the impact of the intervention depends on the group of peers. As discussed above, we introduced some controlled randomness in the design of these groups for the first experimental cohort. In early September, at the end of the introductory week, we first asked students to (freely) form subgroups of two or three persons. In a second step, we randomly matched these subgroups to form the final list of 36 groups. In this setup, it is possible to look at whether students randomly assigned to different subgroups of peers obtain different results at the end of the eleventh grade. To implement this test, table A.10 focuses on the first cohort of students and shows the result of regressing their performance at the end of the eleventh grade on a treatment dummy as well as on the interactions between this treatment dummy and variables indicating the proportion of girls and the proportion of higher-ability students in the subgroups with which their own subgroup were randomly matched, controlling for the same basic set of pretreatment variables as in the previous regression analysis. The regression results suggest that the treatment tends to be more efficient when the proportion of higher-ability students or the proportion of girls are more important. The latter result is consistent with Hoxby (2000) or with Lavy and Schlosser (2011), who provide evidence that an increase in the proportion of girls in a classroom leads to an improvement in students' cognitive outcomes. Further explorations suggest that the influence of peer ability tend to be stronger on high-ability students, but these subgroup analyses rely on small samples and should be taken with caution.

VIII. Discussion

Generally speaking, our experiment suggests that students can volunteer to participate in a program and persist in the program even when it ends up having negative effects on a large fraction of them. To shed light on this paradox, we conducted a survey on former program participants. It took place in 2016, 3–4 years after the end of our experiment. We asked them whether and why they found the program difficult to follow. We also asked them whether and why they decided to quit the program before the end and

about the quality of their interactions with tutors as well as with their group of peers. We collected 200 responses from former participants, 92 of whom are lower ability. These 200 respondents represent about two-thirds of former participants.[20]

One distinctive feature of the program under consideration is that it is conducted by graduate students from the ENS, namely, by some of the very best French graduate students. One potential problem with such elite tutors is that they may have a depressing effect on tutees, especially on those who are not themselves top achievers. They may induce these tutees to think that higher education is not for persons like them. As emphasized earlier, a large proportion of participants (about 63%) have an immigrant background and come from families with little experience of the French system of higher education.

The results of the posttreatment survey are not really consistent with this assumption (table 6). About 86% of respondents agree with the statement that tutors were close to tutees, and 85% agree with the statement that tutors were positive and encouraging. Overall, a very large majority of respondents, including lower-ability ones, had positive relationships with their tutor. In addition, about 68% of respondents actually disagree with the statement that the tutor was difficult to understand (and again, with no difference between lower-ability and higher-ability students).

A second potential explanation for our experimental results is that the program takes too much time and prevents students from allocating enough time to school homework and exam preparation. As it happens, the program is time-consuming. There are three to five 3-hour sessions per term. These sessions took place on Saturday afternoons in Paris (2–5 p.m.), and many students needed 1 hour or more to travel there. In addition, most tutors gave specific homework to students (on top of their school homework), with the effect of increasing students' workload between sessions.

The postintervention survey confirms that tutoring sessions were perceived by many participants as too time-consuming. A majority of respondents agreed with the statement that the travel time was long. About 62% said that they did not do any school homework on the Saturdays when the sessions took place. In 2010, several students actually complained about the amount of time required by the program, and the workload was reduced for the second cohort.[21]

[20] Table A.11 compares the baseline characteristics of students who responded to the survey with those of nonrespondents. It shows that respondents have a slightly better educational background than nonrespondents, so that the responses to the survey may not be representative of those of the entire sample of former participants.

[21] Unfortunately, the sizes of the cohorts are too small for us to be able to detect whether the reduction in the workload for the second cohort was followed by a significant variation in the effect of the program.

Table 6
Students' Perceptions of the Program: A Survey with Former Participants

Variable	All Respondents (1)	Mid-High Ability (2)	High Ability (3)	Difference (SE) (4)
Tutor was close to tutees	.860	.837	.880	.043
				(.049)
Tutor was positive and encouraging	.850	.826	.870	.044
				(.051)
Tutor was difficult to understand	.315	.304	.324	.019
				(.030)
Travel time was long	.525	.467	.574	.107
				(.071)
Was unable to do school homework on tutoring days	.620	.587	.648	.061
				(.069)
Had a lot of work between sessions	.340	.293	.380	.086
				(.067)
Had less time to do school homework because of the program	.250	.228	.269	.040
				(.061)
Reenlistment after the first year	.505	.402	.593	.190**
				(.070)
Pleasant atmosphere in the tutoring group	.900	.880	.917	.036
				(.042)
Became friends with other students of the tutoring group	.710	.674	.741	.067
				(.065)
Spent a lot of time outside sessions with friends from the program	.505	.565	.454	−.112
				(.071)
Kept in touch with former participants from the program	.580	.598	.565	−.033
				(.070)
Observations	200	92	108	

NOTE.—This table shows the responses to the survey conducted in 2016 with former participants. Column 1 refers to the full sample of former participants, col. 2 refers to the mid-high-ability subsample, and col. 3 refers to the high-ability subsample. Column 4 shows the difference between the mid-high and high-ability groups, with the standard error for the difference in parentheses. The high-ability group corresponds to the top half of the distribution of pretreatment ability scores, whereas the mid-high-ability group corresponds to the bottom half. For each item and each sample, we report the proportion of individuals who agree (or strongly agree) with the corresponding statement. The table can be read as follows: 86% of students agree or strongly agree with the statement that their tutor was close to tutees.

** $p < .05$.

The program was also time-consuming by providing participants with new friends and new opportunities to spend time with friends. Three or four years after the program, 90% of respondents reported that the atmosphere of their tutoring was pleasant, 71% said that they became friends with other students from their tutoring group, and 58% said that they had kept in touch with former participants from the program. About 51% of students reported

that they spent a lot of time outside the sessions with friends they met during the program. Among students who persisted in the program in grade 12, a majority mentioned friend relationships as one of the main reasons for their decision.

Overall, our posttreatment survey does not really support the assumption that students were discouraged by the personality of tutors or by their relationships with other tutees. Consistent with our conceptual framework, the main problem seems to be that the program took up too much time, be it because of the length of the sessions themselves, the travel time, the between-session homework, or the induced socialization. This feature of the program is likely one reason why it had such a depressing effect on many mid-high-ability participants and why such a large proportion of these mid-high-ability participants quit the program at the end of the first year (54% of this group quit, vs. 37% for the high-ability group). Many of them should probably have quit earlier: as shown in table 2, most mid-high-ability participants attended virtually all first-year sessions, likely because they took pleasure in meeting with their tutor and their new friends, but at the detriment of their subsequent school performance.

It should be emphasized that when we compare the survey responses of lower- and higher-ability participants, we find no evidence that the program was more time-consuming for lower-ability students, and no evidence either that lower-ability students enjoyed the social nature of the meetings more than higher-ability students (see table 6, cols. 2–4). This is consistent with our conceptual framework, where the amount of time invested in the program is assumed to be the same for both ability groups. As it turns out, our baseline model does not rest on higher- and lower-ability students investing different amount of time in the program but on the fact that, under standard concavity assumptions about the education production function, a given substitution of time invested in extracurricular activities for time invested in curriculum-related activities has more adverse effects on students with the lowest level of initial endowment in curriculum-related knowledge.

Eventually, the posttreatment survey makes it possible to explore whether the way tutors are perceived by participants depends on their gender or on their socioeconomic background. Table A.12 shows the results of regressing variables indicating whether tutors are perceived by participants as close to tutees or encouraging to tutees (as well as whether they are difficult to understand) on dummies indicating the gender of the tutor (panel A) and the family background of the tutor (panel B). The table shows that respondents who had a tutor with a lower socioeconomic background report significantly more often that their tutor was encouraging and close to them. Also, they report significantly less often that their tutor was difficult to understand. The table also shows that female tutors were perceived as more positive and encouraging to tutees than male tutors. Hence, tutors who were perceived as the most encouraging are not really those who obtained the best results. These

findings suggest that being encouraging and close to tutees is not sufficient to be an efficient tutor. They further support the assumption that the problem with the program was not that tutors were not encouraging enough or not close enough to tutees.

IX. Conclusion

In this paper, we report the results of a randomized experiment conducted in 12 underprivileged high schools of the Paris region. The intervention targets tenth-grade volunteer students identified by school principals as having the ability to succeed in high school and to pursue a college education. A random selection of these volunteers attend an intensive 2-year tutoring program designed to improve their academic achievement and readiness for higher education. Tutors are graduates of the ENS, one of the most selective institution of higher education in France.

The experiment reveals that the intervention has positive effects on the performance of higher-ability participants as well as on their probability of gaining access to (and persisting in) the most selective programs of higher education. In that sense, the intervention certainly contributes to reducing inequalities in academic achievement and access to higher education between the best students of underprivileged high schools and their counterparts in more privileged high schools. But the intervention also appears to have significant negative effects on the performance and educational prospects of lower-ability participants.

Hence, a late but intensive intervention conducted by very good graduate students is able to improve the motivation and boost the performance of the best students from underprivileged high schools. But it can also be counterproductive for students who are not strong enough to reconcile homework completion with participation in intense and time-consuming extracurricular activities. This issue is all the more problematic in that, because of the quality of the induced socialization, many students can persist in a program even when it is clear that it is too time-consuming and hurts their educational prospects.

References

Bailey, Martha J., and Susan M. Dynarski. 2011. Gains and gaps: Changing inequality in U.S. college entry and completion. NBER Working Paper no. 17633, National Bureau of Economic Research, Cambridge, MA.

Belhoste, Bruno. 2003. Une histoire des CPGE. In *Actes du colloque Démocratie, classes préparatoires et grandes écoles*.

Bettinger, Eric P., Bridget Terry Long, Philip Oreopoulos, and Lisa Sanbonmatsu. 2012. The role of application assistance and information in college decisions: Results from the H&R Block FAFSA experiment. *Quarterly Journal of Economics* 127, no. 3:1205–42.

Caille, Jean-Paul. 2010. Les cours particuliers en première année de collège : Un entrant en sixième sur dix bénéficie de soutien scolaire payant. *Éducation & formations* 79:59–80.

Castleman, Benjamin, and Joshua Goodman. 2018. Intensive college counseling and the enrollment and persistence of low income students. *Education Finance and Policy* 13, no. 1:19–41.

Castleman, Benjamin L., Lindsay C. Page, and Korynn Schooley. 2014. The forgotten summer: Does the offer of college counseling after high school mitigate summer melt among college-intending, low-income high school graduates? *Journal of Policy Analysis and Management* 33, no. 2:320–44.

Fredricks, Jennifer A. 2012. Extracurricular participation and academic outcomes: Testing the over-scheduling hypothesis. *Journal of Youth and Adolescence* 41, no. 3:295–306.

Fredricks, Jennifer A., and Jacquelynne S. Eccles. 2010. Breadth of extracurricular participation and adolescent adjustment among African-American and European-American youth. *Journal of Research on Adolescence* 20, no. 2:307–33.

Fryer, Roland G., Jr., Tanaya Devi, and Richard T. Holden. 2012. Vertical versus horizontal incentives in education: Evidence from randomized trials. NBER Working Paper no. 17752, National Bureau of Economic Research, Cambridge, MA.

Holm, Sture. 1979. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* 6, no. 2:65–70.

Hoxby, Caroline. 2000. Peer effects in the classroom: Learning from gender and race variation. NBER Working Paper no. 7867, National Bureau of Economic Research, Cambridge, MA.

Hoxby, Caroline, and Sarah Turner. 2013. Informing students about their college options: A proposal for broadening the expanding college opportunities project. Discussion Paper no. 3, Hamilton Project.

Kautz, Tim, and Wladimir Zanoni. 2014. Measuring and fostering noncognitive skills in adolescence: Evidence from Chicago Public Schools and the OneGoal program. University of Chicago.

Knifsend, Casey A., and Sandra Graham. 2012. Too much of a good thing? How breadth of extracurricular participation relates to school-related affect and academic outcomes during adolescence. *Journal of Youth and Adolescence* 41, no. 3:379–89.

Lavecchia, Adam M., Heidi Liu, and Philip Oreopoulos. 2014. Behavioral economics of education: Progress and possibilities. NBER Working Paper no. 20609, National Bureau of Economic Research, Cambridge, MA.

Lavy, Victor, and Analia Schlosser. 2011. Mechanisms and impacts of gender peer effects at school. *American Economic Journal: Applied Economics* 3, no. 2:1–33.

List, John, Azeem Shaikh, and Yang Xu. 2016. Multiple hypothesis testing in experimental economics. NBER Working Paper no. 21875, National Bureau of Economic Research, Cambridge, MA.

Reardon, Sean F. 2011. The widening academic achievement gap between the rich and the poor: New evidence and possible explanations. *Whither opportunity? Rising inequality, schools, and children's life chances*, ed. Greg J. Duncan and Richard J. Murnane, 91–116. New York: Russell Sage Foundation.

Roth, Alvin E. 2008. Deferred acceptance algorithms: History, theory, practice, and open questions. *International Journal of Game Theory* 36, no. 3/4:537–69.

Wai, Jonathan, and Stephen Hsu. 2016. Where Nobel winners get their start. *Nature* 538, no. 7624:152.