A pleasure that hurts: the ambiguous effects of elite tutoring on underprivileged high school students

Son Thierry Ly* Eric Maurin† Arnaud Riegert‡

December 2018

Abstract

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.

JEL codes: I23, I24, D63, D85

Keywords: Underprivileged high-school students, college readiness, randomized experiment

*Paris School of Economics. son.thierry.ly@ens.fr
†Paris School of Economics, EHESS. eric.maurin@ens.fr
‡Paris School of Economics. arnaud.riegert@ens.fr

1 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 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 to get into higher education. In the US, 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; Castleman and Goodman, 2016; Kautz and Zanoni, 2014). 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 in order to help the best students from these high schools to 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 poetentially contribute to bridge 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 to 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" called TALENS, 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 of Paris[1] (hereafter, ENS). Since 2006, the

[1] The École Normale Supérieure 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 the second

ENS offers each year a two-year mentoring and tutoring program to a selection of students coming from twelve 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. 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 in small groups and each group is randomly assigned to a specific tutor. During the two last 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 also to give them a better idea of the requirements of higher education. The travelling costs involved by participation in the program (as well as the cost of attending a one-week introductory meeting at 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 in 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 effects 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*, hereafter 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

institution in the world (just below Princeton) in terms of number of Field medals won by former students.

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 pre-treatment (referred to as "high ability" students), the intervention induces a significant increase in high school achievement as well as in the probability to get access to (and persist 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 probability to enter into 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 a minority only 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 post-treatment 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 of 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 post-treatment 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 came themselves 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 are suggestive 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 to gain access to (and persist) 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 take 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 3 describes the content of the intervention and the randomization process. Section 4 develops a simple conceptual framework that helps clarify why the intervention may differentially affect higher and lower ability participants. Section 5 describes the data used in the econometric analysis while section 6 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 to gain access to (and persist in) selective undergraduate program. Section 7 explores the variation in the effect of the program across the different types of tutors. Section 8 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.

2 Institutional context

In France, compulsory education encompasses 5 years of elementary school (grades 1 to 5) and 4 years of middle school (grades 6 to 9) until age 15. At the end of grade 9, about 60% students enter into high school and pursue general education for three additional years (grade 10 to 12) whereas 40% go to a vocational school or enter into the labor market. At the end of grade 10, students in the general education track can either enter into a more academic program (about 70% do so) or into a more technical one. Students who enter into an academic program have to choose a major field of study and 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 10th graders of 12 underprivileged high schools. Virtually all of them opt for 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 major field of study at the end of grade 10.

[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 10th graders are allowed to enter into 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 SD larger than the average score of those who specialize in humanities and 140% of a SD larger than the average score of those who pursue a non-academic program.

After high school, students who want to enter into a selective undergraduate program[3] have to apply through a Centralized Assignment System (called *Admissions Post-Bac*, hereafter APB). 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 11th and 12th 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 to ask recommendation letters either, as it is usually the case in the US 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 *Classes Préparatoires aux Grandes Écoles* (hereafter, CPGE). Among high school students in an academic track, only about 13% are admitted in a CPGE (17% of those who specialize in science), and 77% make it to the second year. These CPGE programs involve two years of intense preparation (either in 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 two-year CPGE preparation program before entry into 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 aids is unlikely an explanation for the underrepresentation of low SES students in CPGE.

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

[4] In public universities, tuition fees are only about 200 euros. For historical reasons, CPGE 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).

3 The experiment

3.1 The program and its objectives

In 2008, the French government initiated programs (called "*cordées de la réussite*") all over the country in order 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 *École normale supérieure* (ENS) and twelve 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 French intellectual elite for more than a century.

The high schools were selected based on the socioeconomic background of their students as well as on the proportion of students they send to CPGE programs: only 8% of their students enter into 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).

3.2 Identification of volunteers

Each year, in each high school of the network, participants to the ENS program are randomly selected from a set of 10th grade volunteers. The identification of volunteers takes place early April, about two months before the end of the academic year. In each school, the principal starts by identifying 10th grade students who are likely to enter in an academic program in grade 11 and likely to pursue a college education after grade 12. The principal invites them to participate in an information 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 that 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, namely we propose to use objective criteria to define the group of volunteers that were to be invited to the information 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".[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 from the intervention. Eventually, we agreed to let principals choose how exactly to define a "good" student and who exactly to invite in the information meeting. The experiment was expected to help further explore whether the program has the same effect on all students.

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

The manager checks whether the questionnaire is well filled and also whether the student has well understood the implications of volunteering. 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.

On a typical year, the capacity of the program is about 140 seats and there are about 200 volunteers. On academic year 2010-2011 (our first cohort), the ENS agreed to temporarily increase the capacity of the program and school principals were encouraged to boost participation in the information meetings. On this specific year, we ended up with 395 volunteers for 216

[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, namely we would have found a significant positive effect on exit scores and high-school graduation.

seats. On the following year (2011-2012), the capacity was back to normal (140 seats) and the number of volunteers back to 212 students.

3.3 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 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 I provides some statistics describing these 556 volunteer students as well as their non-volunteer schoolmates and the average students in the Parisian region. The table confirms that volunteer students have a much better academic record not only than their non-volunteer schoolmates, but also than the average high school student in the Parisian 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 at the end-of-middle-school national examination (externally set and marked) is on average about +16% of a SD higher than the score obtained by the average high school student in the Parisian region, even though the proportion of students coming from a low-SES family is about 40% more important among volunteer students than among the average high school student in the Parisian 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 Parisian 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 background, compared to the average student in the Parisian region. According to the baseline questionnaire filled 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 or not. Both selected and non-selected volunteers are reminded that decisions are the result of a random draw.

3.4 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 pre-assigned each volunteer student to one future 11th grade class (based on her choice of field of study). In French high schools, each 11th grade class corresponds to either 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.I in the online appendix builds on the information provided by administrative registers to compare the socio-demographic characteristics of volunteer students in the treatment and control groups (in terms of gender, grade repetition, parental occupation, pre-treatment grades). We find no significant differences between the two groups. To further test for the similarity of the two groups, online appendix Table A.II compares the responses of treatment and control groups to the baseline questionnaire that students had to fill in order to be identified as volunteers. Again, we find little difference in responses across the two groups.

Finally, we augmented our dataset 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 these information, Table A.III in the online Appendix shows that

[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 (so as to

there is no significant difference in gender, level of experience, teachning hours or educational qualification across teachers assigned to treated or control classes.

3.5 High ability *vs* 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 pre-treatment. To test for such heterogenous 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 pre-treatment as well as on the half with the weakest academic records pre-treatment.[8] We replicate the comparison of the pre-treatment characteristics of treated and control students separately for the two ability groups in panels B and C of appendix Table A.I. We do not detect any significant pre-treatment difference across treatment and control students within both ability groups.

Table A.IV in the online appendix also shows the proportion of higher and lower ability volunteers in each decile of the distribution of 9th 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 3 to 6. 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".

Online appendix Table A.V further compares the responses of the two ability groups to the baseline questionnaire. The table is suggestive that students from the higher ability group can get help more easily for homework. They also tend to provide better assessment of their

become *agrégés*) or the *Certificat d'Aptitude au Professorat de l'Enseignement du Second Degré* exam (herafter the CAPES, so as to become *certifiés*). Both are competitive exams, but the *Agrégation* is more difficult and selective than the CAPES. It is typically taken after four years of higher education whereas CAPES is normally taken after three 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 10th grade (*i.e.*, the year before the treatment). In this set-up, 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.

[9] We cannot show their proportions in the deciles of the distribution of 10th grade scores because 10th grade scores are observed for volunteer students only.

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 the CPGE programs.

3.6 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 in order to inform them of the results of the draw and to invite those of the treatment group to participate in an introductory week (called campus week) organized at ENS 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 grade 11 academic program, which is much more demanding than grade 10. During the first two 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 have also 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 did participate to this week (see Table II).

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 curricula. There are four possible themes: science, social sciences and history or 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 3 to 7 students with similar thematic preferences.[10] Eventually, the program

[10] Specifically we defined one 7-person group, twenty-four 6-person groups, ten 5-person groups and one 4-person group in 2009-2010. In 2010-2011, the average group size and the participant-tutor ratio were lower. Specifically, we had one 7-person group, six 6-person groups, six 5-person groups, seven 4-person groups and ten 3-person groups in 2010-2011. On cohort 2009-2010 (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 4 to 6 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 6-person group, which explains why we end up with a 7-person group). In the last section of the paper,

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 over representation of students with a privileged family background.

New tutors are recruited early October at the beginning of the academic year. They are all ENS students. They first benefit from a two-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 ENS) between tutors and their groups of six students. The team of tutors meets three times a year in order to share experience and get feedback from program managers.

The academic year is divided in 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 twelve 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 9 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 is suggestive that attendance rates were high: about two third of students participate in 10 (or more) sessions out of 12 (Table II).

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

we build on this randomized design to explore the role of peer groups.

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 one-day information meeting on higher education. Also, starting in March, six additional sessions were organized in order to help them prepare 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.

4 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 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.

4.1 Set up and notations

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 to identify tracks that fit their taste and academic aptitudes. It helps them to formulate education plans and to stay focused at school.

At the end of 10th 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} the amount of time devoted to further accumulate the first type of knowledge during grade 11 and 12. Similarly, we denote T_{fi} 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 $T_{ci} + T_{fi} = 1$. 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 the achievement of student i at the end of high school and we assume,

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

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 10th 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

[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.

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 ($K_{0ci} = K_{0L} < K_{0H}$).

4.2 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 non-curricular 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 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,

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

Overall, the impact of being assign to the treatment group rather than to the control group

[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.

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:

$$\Delta_i \approx T(F'_f(K_{0ci} + 1, K_{0f}) - F'_c(K_{0ci} + 1, K_{0f})) \tag{5}$$

where F'_c denotes the marginal product of curriculum-related knowledge input and F'_f 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}, namely it may be negative for mid-high ability students and positive for high ability ones.[13] The next sections explore these issues empirically.

Eventually, online appendix B develops two possible extensions of our baseline model in order 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 Equation (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 from 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. As-

[13] Assuming for example that F is a CES 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 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.

suming that the marginal costs of school-related efforts increase more rapidly for lower ability students, participation 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.

5 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 dataset with exhaustive administrative data on students' performance on the national exam taken at the end of 9th grade (end of middle school exam called *Diplôme National du Brevet*, hereafter the DNB) as well as on the national exam taken at the end of high school (*Baccalauréat*). We were also able to augment our initial dataset with administrative school registers (*Bases Centrales Scolarité*, hereafter the BCS) 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 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, 6 treated and 8 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 in order to be identified as volunteer. As discussed above, this pre-treatment 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, etc.

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

6 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 grade 11 and grade 12. We focus on the national exams (externally set and marked) that students have to take in order 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 pre-treatment control variables that includes dummies for gender, grade repetition, family background, pre-treatment 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.

6.1 Effects on high school achievement

Table III shows the effect of the intervention on scores on exit exams taken at the end of grade 11 (column I), on scores on exit exams taken at the end of grade 12 (column II) and on average scores on exit exams (column III). The Table also shows the impact of the intervention on the probability of high school graduation on time (column IV) as well as on the probability of graduating at all (*i.e.*, including after repeating grade 12, column V).[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 nor on their high school

[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.

graduation rate. The second panel of Table III refers to the half of the sample of volunteers with the strongest academic records pre-treatment (*i.e.*, in grade 10) whereas the third panel refers to the half of the sample of volunteers with the weakest academic records pre-treatment. 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 of the first group are in the two top deciles (high ability students) whereas the majority of students of 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 one. As discussed in the previous sections, the program contributes to substituting extracurricular activities for curriculum-related ones and our regression 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 from 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.I in the online appendix, we provide a graphical representation of changes in treatment effects on standardized exam scores across grades and ability groups.

The last panel of Table III confirms that the differences between the two sets of impacts are significant at 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

[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 pre-treatment (as we initially intended to do), we obtain impacts that are almost as strong as those obtained with the higher ability sample, namely a positive impact of 18% of a SD on exit score and a positive impact of 7.5 percentage points on graduation rate.

[16] 78% of high ability students specialize in science, versus 46% of mid-high ability students. It should be noted, however, that differences in estimated treatment effects across exams taken on grade 11 and grade 12 are not significant at standard level, so that these differences should be interpreted cautiously.

about 20 percentage points larger in the treatment group, namely almost a tripling of the gap within the group that was given the opportunity to participate in the program.[17]

For each one of the five available measure of performance on high school exams, Table III tests two null assumptions, one per ability subgroup. A well known issue is that such subgroups comparisons tend to increase mechanically the likelihood of finding significant effects (see *e.g.* List, Shaikh, and Xu, 2016). In Table A.VI in the online Appendix, we report for each one of the five variables the corresponding unadjusted p-values as well the p-values adjusted to take into account (ability) subgroup comparisons, using Holm's method (Holm, 1979). Generally speaking adjusted and unadjusted tests provide qualitatively similar results, even though multiple testing yields larger p-value. 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 $p < 0.01$ whereas the adjusted test rejects the null assumption for the same ability group at $p < 0.02$.

It is also possible to consider jointly 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, adjusted p-value suggests that the negative effect on the grade 11 scores of lower ability students is significant at $p < 0.07$ (unadjusted $p < 0.02$) whereas the positive effect on the grade 12 scores of higher ability students is significant at $p < 0.13$ (unadjusted $p < 0.05$). Hence, the finding that lower ability students are mostly affected during grade 11 and higher ability mostly during grade 12 appears to be less robust to multiple testing.

To further explore the robustness of our findings, Table A.VII in the online appendix 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

[17] It should be noted that it is not uncommon to find heterogeneous treatment effects across ability groups (see *e.g.* Roland 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 et al., 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 college by about 30%.

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 for this assumption, Table A.VIII and Table A.IX in the online appendix show the impact of the intervention on the grades obtained at 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 standard level in French (+43% of a SD), 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 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 more general determinants of performance at school. One such determinant is likely the amount of time devoted to school homeworks, revision exercises and preparation of tests. The participation in the program is time consuming and it likely contributes to reduce 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.

6.2 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 *Classes Préparatoires aux Grandes Écoles* (herafter, CPGE program). 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 ENS after two years spent in a CPGE program: it is certainly the undergraduate program that they know the best and about which they are able to provide the most comprehensive information.

To shed light on these issues, Table IV shows the effect of the intervention on the proportion of students who gain admission into a CPGE program after high school (column I) as well as on the proportion who are still in a CPGE program two years after high school graduation[18] (column II). The table shows no significant effect on the proportion of students who gain admission into a CPGE program after high school nor on the proportion who are still in a CPGE program two years after high school, consistent with the intervention having on average no effect on education aspirations nor 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 percent points on year 1 enrollment, which corresponds to an about -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 is suggestive 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 lower ability participants, but they are not significant at 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, but a significant increase in the difference between the two ability groups (+12.2 percentage points increase).

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

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

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 adjustment that take into account subgroup comparisons (see again online Appendix Table A.VI). We also considered jointly 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 non-adjusted tests reject the same two assumptions at $p < 0.05$, 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.

7 Mechanisms: the role of tutors and peers

The previous sections are suggestive that the program has positive effects on volunteers with the best academic level pre-treatment, 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 be suggestive 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 only, 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.

7.1 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 set-up, 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 socio-economic 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 tutors rather than to another one.

Table V implements this test. It shows the results of regressing the grade 11 score on a treatment dummy and on the interactions between this treatment 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 pre-treatment variables as in Table III. 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 significantly better outcomes when it is implemented by tutors with a higher socioeconomic background, especially when we focus on higher ability recipients. It is suggestive that tutors with a higher socioeconomic background contribute to improve the overall impact of the intervention, but also contribute to widen 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.

7.2 Peer group influence

The results in Table V provide some evidence that tutors matter. They are suggestive 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. Early September, at the end of the introductory week, we first asked students to form (freely) sub-groups of two or three persons. In a second step, we matched randomly these sub-groups in order to form the final list of 36 groups. In this set-up, it is possible to look whether students randomly assigned to different sub-groups of peers obtain different results at the end of 11th grade. To implement this test, Table A.X focuses on the first cohort of students and shows the result of regressing their performance at the end of the 11th 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 pre-treatment variables as in the previous regression analysis. The regression results suggest that the treatment tend 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

[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.

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 analysis rely on small samples and should be taken with caution.

8 Discussion

Generally speaking, our experiment suggests that students can volunteer to participate in a program, and persist in this program, even when this program 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, three to four 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 2/3 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 post-treatment survey are not really consistent with this assumption (Table VI). 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).

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

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 exams' 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 (2pm-5pm) and many students need 1 hour or more to go to Paris. In addition, most tutors used to give specific homework to students (on top of their school homework) with the effect of increasing students' workload between sessions.

The post-intervention 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% say 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]

The program was also time consuming by providing participants with new friends and with new opportunities to spend time with friends. Three or four year after the program, 90% of respondents report that the atmosphere of their tutoring was pleasant, 71% say that they became friends with other students from their tutoring group and 58% say they had kept in touch with former participants from the program. About 51% of students report that they spent a lot of time outside of 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 post-treatment 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 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 quitters for this group, against 37% for the high ability group). Many of them should probably have quit earlier: as shown in Table II, most mid-high participants attended virtually all first-year sessions, likely because they took pleasure in meeting with their tutor and their

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

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, no evidence either that lower ability students enjoyed the social nature of the meetings more than higher ability students (see Table VI, columns (II) to (IV)). 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 post-treatment 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.XII in the online appendix 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.

9 Conclusion

In this paper, we report the results from a randomized experiment conducted in twelve underprivileged high schools of the Paris region. The intervention targets 10th 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 two-year tutoring program designed to improve their academic achievement and readiness for higher education. Tutors are graduates from the *École normale supérieure* of Paris, namely 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 to gain access to (and persist 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 education 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 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 education prospects.

References

Avery, Christopher and Kane, Thomas J. Student perceptions of college opportunities. the boston coach program. In Hoxby, Caroline, editor, *College choices: The economics of where to go, when to go, and how to pay for it*, pages 355–394. University of Chicago Press, 2004.

Bailey, Martha J. and Dynarski, Susan M. Gains and gaps: Changing inequality in u.s. college entry and completion. Working Paper 17633, National Bureau of Economic Research, dec 2011.

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

Bettinger, Eric P., Long, Bridget Terry, Oreopoulos, Philip, and Sanbonmatsu, Lisa. The role of application assistance and information in college decisions: Results from the h&r block fafsa experiment*. *The Quarterly Journal of Economics*, 127(3):1205–1242, 2012.

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

Castleman, Benjamin and Goodman, Joshua. Intensive college counseling and the enrollment and persistence of low income students. *Education Finance and Policy*, 2016.

Castleman, Benjamin L., Page, Lindsay C., and Schooley, Korynn. 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(2):320–344, 2014.

Cornwell, Christopher, Mustard, David B., and Sridhar, Deepa J. The enrollment effects of merit-based financial aid: Evidence from georgia's hope program. *Journal of Labor Economics*, 24(4):761–786, 2006.

Dee, Thomas S. A teacher like me: Does race, ethnicity, or gender matter? *The American economic review*, 95(2):158–165, 2005.

Dynarski, Susan. The new merit aid. In *College choices: The economics of where to go, when to go, and how to pay for it*, pages 63–100. University of Chicago Press, 2004.

Fairlie, Robert W., Hoffmann, Florian, and Oreopoulos, Philip. A community college instructor like me: Race and ethnicity interactions in the classroom. *The American Economic Review*, 104(8):2567–2591, 2014.

Fredricks, Jennifer A. Extracurricular participation and academic outcomes: Testing the over-scheduling hypothesis. *Journal of Youth and Adolescence*, 41(3):295–306, 2012.

Fredricks, Jennifer A and Eccles, Jacquelynne S. Breadth of extracurricular participation and adolescent adjustment among african-american and european-american youth. *Journal of Research on Adolescence*, 20(2):307–333, 2010.

Grodner, Andrew and Rupp, Nicholas G. The role of homework in student learning outcomes: Evidence from a field experiment. *The Journal of Economic Education*, 44(2):93–109, 2013.

Holm, Sture. A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, pages 65–70, 1979.

Hoxby, Caroline. Peer Effects in the Classroom: Learning from Gender and Race Variation. NBER Working Papers 7867, National Bureau of Economic Research, Inc, August 2000.

Hoxby, Caroline and Turner, Sarah. Informing students about their college options: A proposal for broadening the expanding college opportunities project. *The Hamilton Project, Discussion Paper*, 3, 2013.

Kane, Thomas J. A quasi-experimental estimate of the impact of financial aid on college-going. NBER Working Papers 9703, National Bureau of Economic Research, Inc, 2003.

Kautz, Tim and Zanoni, Wladimir. Measuring and fostering non-cognitive skills in adolescence: Evidence from chicago public schools and the onegoal program. *University of Chicago*, 2014.

Knifsend, Casey A and Graham, Sandra. 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(3):379–389, 2012.

Lavecchia, Adam M., Liu, Heidi, and Oreopoulos, Philip. Behavioral economics of education: Progress and possibilities. NBER Working Papers 20609, National Bureau of Economic Research, Inc, 2014.

Lavy, Victor and Schlosser, Analia. Mechanisms and impacts of gender peer effects at school. *American Economic Journal: Applied Economics*, 3(2):1–33, 2011.

List, John, Shaikh, Azeem, and Xu, Yang. Multiple Hypothesis Testing in Experimental Economics. Working Paper 21875, National Bureau of Economic Research, 2016.

Oreopoulos, Philip and Salvanes, Kjell G. Priceless: The nonpecuniary benefits of schooling. *The journal of economic perspectives*, 25(1):159–184, 2011.

Oreopoulos, Philip, Brown, Robert S., and Lavecchia, Adam M. Pathways to education: An integrated approach to helping at-risk high school students. Technical report, National Bureau of Economic Research, 2014.

Reardon, Sean F. The widening academic achievement gap between the rich and the poor: New evidence and possible explanations. *Whither opportunity*, pages 91–116, 2011.

Roland G. Fryer, Jr, Devi, Tanaya, and Holden, Richard T. Vertical versus horizontal incentives in education: Evidence from randomized trials. Working Paper 17752, National Bureau of Economic Research, January 2012.

Roth, Alvin E. Deferred acceptance algorithms: History, theory, practice, and open questions. *International Journal of Game Theory*, 36(3-4):537–569, 2008.

Scott-Clayton, Judith. On money and motivation: A quasi-experimental analysis of financial incentives for college achievement. *Journal of Human Resources*, 46(3):614–646, 2011.

Wai, Jonathan and Hsu, Stephen. Where nobel winners get their start. *Nature*, 538(7624), 2016.

Table I: The characteristics of students who volunteer to participate in the program

	Paris region	Experimental high schools	Volunteer students	Volunteers, cohort 2010	Volunteers, cohort 2011
Variable	(I)	(II)	(III)	(IV)	(V)
Female	0.528	0.536	0.601	0.618	0.559
High-SES family	0.416	0.200	0.255	0.258	0.248
Low-SES family	0.265	0.438	0.362	0.357	0.373
Average 9th grade score	0.000	−0.557	0.156	0.137	0.203
Has repeated a grade	0.257	0.337	0.191	0.200	0.168
Pursue an academic track	0.583	0.533	0.998	1.000	0.994
Specialize in science	0.304	0.236	0.635	0.587	0.752
Obs.	209,654	7,032	556	395	161

Note: Column (I) shows the average characteristics of general education 10th grade students in the region of Paris. Column (II) shows the same characteristics for general education 10th grade students in the twelve high schools of the experiment and column (III) for the volunteer students in these twelve high schools. Columns (IV) and (V) further show the characteristics of volunteer students in each cohort. The ability score corresponds to the (standardized at the Paris region level) average grade obtained at the national middle school exit exams taken at the end of grade 9.

Reading: 60.1% of volunteer students are female, 25.5% come from a high SES family background. At the end of grade 10, 99.8% choose to pursue an academic track, 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.

Table II: Take up rates, by cohort and ability group

	All	Mid-high ability	High ability
	(I)	(II)	(III)
Panel A: Cohort 2010			
Introductory week (year 1)	0.792	0.709	0.877
More than 10 sessions attended (year 1)	0.645	0.618	0.675
Re-enlistment	0.569	0.482	0.660
Obs.	216	110	106
Panel B: Cohort 2011			
Introductory week (year 1)	0.742	0.744	0.739
Re-enlistment	0.483	0.395	0.565
Obs.	89	43	46

Note: The sample includes eligible students from cohorts 2010 (panel A) and from cohort 2011 (panel B). In each panel, column (I) shows take up rates for the full sample whereas column II (column III) shows take up rates for the mid-high (high) ability group. The high-ability group corresponds to the top half of the distribution of pre-treatment ability scores across volunteers whereas the mid-high ability group corresponds to the bottom half

Reading: Among eligible students from cohort 2010, 79.2% attended the (year 1) introductory week at ENS, 64.5% attended more than 10 tutoring sessions during the first year of the program, 56.9% re-enlisted at the end of the first year.

For cohort 2011, information on attendance was not collected.

Table III: The effect of the treatment on performance on high school exit examinations

Dependent variable	Grade 11 average score	Grade 12 average score	Overall average score	Graduation on time	Graduation
	(I)	(II)	(III)	(IV)	(V)
			Panel A: All volunteer students		
Treatment	−0.096	0.025	0.013	0.016	0.001
	(0.075)	(0.080)	(0.078)	(0.041)	(0.026)
Obs.	542	542	542	556	556
Mean dep. var.	0.665	0.343	0.413	0.725	0.916
			Panel B: Mid-high ability		
Treatment	−0.254**	−0.165	−0.185	−0.074	−0.034
	(0.103)	(0.137)	(0.131)	(0.063)	(0.045)
Obs.	276	276	276	285	285
Mean dep. var.	0.314	0.101	0.138	0.667	0.902
			Panel C: High ability		
Treatment	0.025	0.230**	0.221**	0.124**	0.074**
	(0.114)	(0.111)	(0.108)	(0.048)	(0.026)
Obs.	266	266	266	271	271
Mean dep. var.	1.055	0.613	0.719	0.790	0.933
			Panel D: Differential impact		
Treatment × high ability	0.279*	0.396**	0.406**	0.198**	0.108**
	(0.150)	(0.183)	(0.177)	(0.075)	(0.051)
Obs.	542	542	542	556	556
Mean dep. var.	0.741	0.512	0.581	0.123	0.031

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, pre-treatment ability score and socioeconomic family background as control variables.
Column (I) shows the estimated effect of the treatment when the dependent variable is the average grade obtained at examinations taken at the end of grade 11, column (II) when the dependent variable is the average grade obtained at examinations taken at the end of grade 12 and column (III) when the dependent variable is the average grade across all examinations. Column (IV) shows the estimated effect when the dependent variable is a dummy indicating high school graduation on time and column (V) when the dependent variable is a dummy indicating high school graduation at any time.
The first panel refers to the full sample, the second panel to the mid-high ability subsample and the third panel to the high-ability subsample. The high-ability group corresponds to the top half of the distribution of pre-treatment ability scores across volunteers whereas the mid-high ability one corresponds to the bottom half. Finally, the fourth panel 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 socio-demographic controls and their interactions with the high ability dummy. Standard errors clustered at the class level are reported in parentheses. ** and * denote significance at the 5% and 10% levels, respectively.

Table IV: The effect of the treatment on access to selective undergraduate programs (CPGE)

Major field of study	All majors		Science major		Humanities major	
	Year 1	Year 2	Year 1	Year 2	Year 1	Year 2
	(I)	(II)	(III)	(IV)	(V)	(VI)
			Panel A: All volunteer students			
Treatment	−0.014	−0.015	0.011	−0.004	−0.063	−0.033
	(0.030)	(0.023)	(0.039)	(0.031)	(0.041)	(0.026)
Obs.	556	556	353	353	203	203
Mean dep. var.	0.163	0.112	0.156	0.123	0.175	0.093
			Panel B: Mid-high ability			
Treatment	−0.067**	−0.058**	−0.070*	−0.075**	−0.072*	−0.044*
	(0.030)	(0.022)	(0.039)	(0.034)	(0.042)	(0.025)
Obs.	285	285	138	138	147	147
Mean dep. var.	0.136	0.083	0.115	0.098	0.155	0.070
			Panel C: High ability			
Treatment	0.055	0.044	0.076	0.057	−0.043	−0.013
	(0.054)	(0.039)	(0.059)	(0.043)	(0.133)	(0.059)
Obs.	271	271	215	215	56	56
Mean dep. var.	0.193	0.143	0.183	0.140	0.231	0.154
			Panel D: Differential impact			
Treatment × high ability	0.122**	0.102**	0.146**	0.132**	0.029	0.031
	(0.060)	(0.044)	(0.068)	(0.053)	(0.118)	(0.057)
Obs.	556	556	353	353	203	203
Mean dep. var.	0.057	0.060	0.068	0.042	0.076	0.084

Note: The sample includes volunteer students from cohorts 2010 and 2011.
The table shows the results from regressions in which the probability to enter into a CPGE program (year 1) as well as the probability to be still in a CPGE program two years after high school graduation (year 2) are regressed on a treatment dummy, using students' gender, pre-treatment ability score and socioeconomic family background as control variables.
The first panel refers to the full sample, the second panel to the mid-high ability subsample and the third panel to the high-ability subsample. The high-ability group corresponds to the top half of the distribution of pre-treatment ability scores across volunteers whereas the mid-high ability one corresponds to the bottom half. Finally, the fourth panel 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 socio-demographic controls and their interactions with the high ability dummy.
Within each panel, columns (III) and (IV) refer to students who specialize in science at the end of grade 10 whereas columns (V) and (VI) to students who specialize in humanities. Standard errors clustered at the class level are reported in parentheses. ** and * denote significance at the 5% and 10% levels, respectively.

Table V: The effect of tutors' gender and family background on grade 11 scores

	(I)	(II)	(III)	(IV)
Panel A: All volunteer students				
Treatment	−0.096	−0.119	−0.317**	−0.361**
	(0.075)	(0.095)	(0.079)	(0.098)
Treatment × Female tutor	–	0.069	–	0.087
		(0.145)		(0.145)
Treatment × Higher background tutor	–	–	0.311**	0.325**
			(0.147)	(0.149)
Obs.	542	542	523	523
Panel B: Mid-high ability students				
Treatment	−0.254**	−0.228*	−0.405*	−0.361
	(0.103)	(0.117)	(0.112)	(0.123)
Treatment × Female tutor	–	−0.078	–	−0.086
		(0.172)		(0.196)
Treatment × Higher background tutor	–	–	0.257	0.244
			(0.225)	(0.235)
Obs.	276	276	262	262
Panel C: High ability students				
Treatment	0.025	0.003	−0.329	−0.394
	(0.114)	(0.139)	(0.117)	(0.142)
Treatment × Female tutor	–	0.069	–	0.126
		(0.246)		(0.233)
Treatment × Higher background tutor	–	–	0.410	0.437*
			(0.253)	(0.244)
Obs.	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, pre-treatment ability score and socioeconomic family background as control variables.
The first panel refers to the full sample, the second panel to the mid-high ability subsample and the third panel to the high-ability subsample. The high-ability group corresponds to the top half of the distribution of pre-treatment 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. ** and * denote significance at the 5% and 10% levels, respectively.

Table VI: Students' perceptions of the program: a survey with former participants

	All respondents	Mid-high ability	High ability	Difference (s.e.)
Variable	(I)	(II)	(III)	(IV)
Tutor was close to tutees	0.860	0.837	0.880	0.043 (0.049)
Tutor was positive and encouraging	0.850	0.826	0.870	0.044 (0.051)
Tutor was difficult to understand	0.315	0.304	0.324	0.019 (0.030)
Travel time was long	0.525	0.467	0.574	0.107 (0.071)
Was unable to do school homeworks on tutoring days	0.620	0.587	0.648	0.061 (0.069)
Had a lot of work between sessions	0.340	0.293	0.380	0.086 (0.067)
Had less time to do school homework because of the program	0.250	0.228	0.269	0.040 (0.061)
Re-enlistment after the first year	0.505	0.402	0.593	0.190** (0.070)
Pleasant atmosphere in the tutoring group	0.900	0.880	0.917	0.036 (0.042)
Became friends with other students of the tutoring group	0.710	0.674	0.741	0.067 (0.065)
Spent a lot of time outside sessions with friends from the program	0.505	0.565	0.454	−0.112 (0.071)
Kept in touch with former participants from the program	0.580	0.598	0.565	−0.033 (0.070)
Obs.	200	92	108	

Note: This table shows the responses to the survey conducted in 2016 with former participants. Column (I) refers to the full sample of former participants, column (II) refers to the mid-high ability subsample and column (III) to the high-ability subsample. Column (IV) shows the difference between the mid-high and high ability groups, and the standard error on this difference in parentheses. ** and * denote significance of this difference at the 5% and 10% levels, respectively.
The high-ability group corresponds to the top half of the distribution of pre-treatment 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.
Reading: 86% of students agree or strongly agree with the statement that their tutor was close to tutees.

A Online appendix

Table A.I: Balancing tests using administrative information

Variable	Control	T − C	s.e.	Obs.
Panel A: All volunteer students				
Has repeated a grade	0.175	+0.040	(0.035)	556
Female	0.622	−0.034	(0.043)	556
High-SES family	0.247	+0.016	(0.038)	556
9th grade score (Math)	0.725	−0.109	(0.077)	517
9th grade score (French)	0.554	−0.061	(0.084)	519
Pursue an academic track	1.000	−0.004	(0.004)	556
Specialize in science	0.614	+0.004	(0.004)	556
Panel B: Mid-high ability students				
Has repeated a grade	0.333	+0.060	(0.064)	285
Female	0.591	−0.022	(0.063)	285
High-SES family	0.265	−0.008	(0.055)	285
9th grade score (Math)	0.420	−0.094	(0.118)	255
9th grade score (French)	0.196	+0.001	(0.123)	257
Pursue an academic track	1.000	0.000	(0.000)	285
Specialize in science	0.462	0.000	(0.000)	285
Panel C: High ability students				
Has repeated a grade	0.000	0.000	(0.000)	271
Female	0.655	−0.050	(0.065)	271
High-SES family	0.227	+0.032	(0.055)	271
9th grade score (Math)	1.040	−0.064	(0.110)	262
9th grade score (French)	0.926	−0.111	(0.111)	262
Pursue an academic track	1.000	0.000	(0.000)	271
Specialize in science	0.782	+0.005	(0.006)	271

Note: This table shows estimated differences in baseline characteristics between the control group and the treatment group.
For each baseline variable, the first column gives the control group average, the second column gives the estimated difference between the control and the treatment group (using a regression where we control for cohort dummies, high school dummies and grade 11 dummies), the third column gives the corresponding standard error whereas the last column gives the number of observations with non-missing values.
The first panel refers to the full sample of volunteers, the second panel to the mid-high ability subsample and the third panel to the high-ability subsample. The high-ability group corresponds to the top half of the distribution of pre-treatment ability scores across volunteers whereas the mid-high ability group corresponds to the bottom half.
** and * denote significance at the 5% and 10% levels, respectively.

Table A.II: Balancing tests using the baseline survey

Variable	Control	T − C	s.e.	Obs.
Father born out of France	0.580	−0.004	(0.052)	445
Mother born out of France	0.589	−0.014	(0.049)	473
Father finished secondary education	0.625	−0.103*	(0.056)	401
Mother finished secondary education	0.588	−0.084	(0.055)	419
Can get help for homework	0.658	+0.005	(0.049)	474
Positive self-assessment	0.620	+0.057	(0.048)	459
Has ideas about future career	0.722	−0.050	(0.048)	483
Has strong educational aspirations	0.599	−0.053	(0.052)	456
Knows about…				
Grandes Écoles	0.582	−0.060	(0.053)	470
CPGE	0.569	−0.037	(0.051)	471
University	0.549	+0.027	(0.053)	471
IEP	0.361	−0.058	(0.049)	469
IUT	0.143	+0.041	(0.040)	472
BTS	0.398	−0.037	(0.052)	472
Can get help for field choice	0.621	+0.031	(0.051)	465
Believes in equal opportunities	0.421	+0.051	(0.054)	449
Interests:				
Politics	0.347	−0.003	(0.048)	471
Economics	0.335	+0.012	(0.048)	467
National news	0.691	−0.004	(0.048)	472
International news	0.763	+0.007	(0.046)	471
Litterature	0.490	−0.007	(0.052)	470
History	0.634	−0.008	(0.053)	470
Science and tech.	0.592	−0.005	(0.045)	468
Sports	0.582	+0.073	(0.050)	472
Social issues	0.640	−0.028	(0.051)	466
Arts	0.529	−0.030	(0.051)	467

Note: This table shows estimated differences in responses to the baseline survey between the control group and the treatment group.
For each item of the baseline survey, the first column gives the control group average, the second column gives the estimated difference between the control and the treatment group (using a regression where we control for cohort dummies, high school dummies and grade 11 dummies), the third column gives the corresponding standard error whereas the last column gives the number of observations with non-missing values.
** and * denote significance at the 5% and 10% levels, respectively.

Table A.III: Balancing tests using administrative information on teachers

Variable	Control	T − C	s.e.	Obs.
Female	0.585	+0.010	(0.042)	822
Experience (log)	2.261	+0.021	(0.069)	761
Highly qualified	0.343	−0.008	(0.041)	822
Teaching hours per week	16.147	−0.253	(0.262)	822

Note: This table shows estimated differences in characteristics of teachers assigned to classes from the treated and the control groups.
For each baseline variable, the first column gives the control group average, the second column gives the estimated difference between the control and the treatment group (using a regression where we control for cohort dummies, high school dummies and grade 11 major dummies), the third column gives the corresponding standard error whereas the last column gives the number of observations with non-missing values.
** and * denote significance at the 5% and 10% levels, respectively.

Table A.IV: Distribution of volunteer students across deciles of the distribution of 9th grade scores

Decile	Mid-high ability	High ability	All volunteer students
1	0.116	0.417	0.262
2	0.130	0.247	0.187
3	0.158	0.126	0.142
4	0.154	0.070	0.113
5	0.070	0.059	0.065
6	0.105	0.026	0.067
7	0.056	0.011	0.034
8	0.049	0.000	0.025
9	0.032	0.007	0.020
10	0.025	0.004	0.014
Grade missing	0.105	0.033	0.070

Note: The table shows the proportion of volunteer students in the different deciles of the distribution of 9th grade scores in their high school. These scores correspond to the average marks received at the end-of-middle school national exams (externally set and marked).
The high-ability group corresponds to the top half of the distribution of pre-treatment ability scores (as measured by average grades given by teachers during 10th grade) across volunteer students whereas the mid-high ability group corresponds to the bottom half.
Reading: 12.1% of students of the lower-ability group are in the first decile of the distribution of 9th grade scores in their high school.

Table A.V: Differences in responses to the baseline survey between the two ability groups.

Variable	Mid-high ability	High − Mid-high	s.e.	Obs.
Father born out of France	0.578	−0.040	(0.051)	445
Mother born out of France	0.590	−0.066	(0.048)	473
Father has higher education	0.573	+0.027	(0.057)	401
Mother has higher education	0.520	+0.082	(0.057)	419
Can get help for homework	0.606	+0.096*	(0.050)	474
Positive self-assessment	0.500	+0.272**	(0.048)	459
Has ideas about future career	0.690	+0.025	(0.048)	483
Has strong educational aspirations	0.526	+0.017	(0.053)	456
Knows about…				
Grandes Écoles	0.549	−0.015	(0.052)	470
CPGE	0.500	+0.135**	(0.053)	471
university	0.564	−0.044	(0.053)	471
IEP	0.359	−0.082	(0.050)	469
IUT	0.124	+0.057	(0.039)	472
BTS	0.355	+0.075	(0.051)	472
Can get help for field choice	0.588	+0.081	(0.051)	465
Believes in equal opportunities	0.509	−0.186**	(0.052)	449
Interests:				
Politics	0.345	−0.003	(0.050)	471
Economics	0.364	−0.045	(0.048)	467
National news	0.628	+0.141**	(0.049)	472
International news	0.712	+0.118**	(0.047)	471
Litterature	0.556	−0.050	(0.051)	470
History	0.621	+0.071	(0.050)	470
Science and tech.	0.571	−0.072*	(0.042)	468
Sports	0.661	−0.075	(0.050)	472
Social issues	0.595	+0.121**	(0.049)	466
Arts	0.550	+0.001	(0.052)	467

Note: This table shows differences in responses to the baseline survey between mid-high ability students (*i.e.*, below the median of the distribution of pre-treatment ability score across volunteers) and high ability students (*i.e.*, above the median).
For each item of the baseline survey, the first column gives the average of the mid high ability group, the second column gives the estimated difference between the two ability groups (using a regression where we control for cohort dummies, high school dummies and grade 11 major dummies), the third column gives the corresponding standard error whereas the last column gives the number of observations with non-missing values.
** and * denote significance at the 5% and 10% levels, respectively.

Table A.VI: Adjusted and unadjusted p-values

Dependent variable	Grade 11 average score	Grade 12 average score	Overall average score	Graduation on time	Graduation	CPGE Year 1	CPGE Year 2
	(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)
Panel A: All volunteer students							
Treatment	−0.096	0.025	0.013	0.016	0.001	−0.014	−0.015
	(0.075)	(0.080)	(0.078)	(0.041)	(0.026)	(0.030)	(0.023)
Obs.	542	542	542	556	556	556	556
Mean dep. var.	0.665	0.343	0.413	0.725	0.916	0.163	0.112
Panel B: Mid-high ability							
Treatment	−0.254**	−0.165	−0.185	−0.074	−0.034	−0.067**	−0.058**
	(0.103)	(0.137)	(0.131)	(0.063)	(0.045)	(0.030)	(0.022)
Obs.	276	276	276	285	285	285	285
Unadjusted p-value	0.014	0.231	0.159	0.243	0.451	0.025	0.011
Adjusted p-value	0.028	0.231	0.159	0.243	0.451	0.050	0.022
Panel C: High ability							
Treatment	0.025	0.230**	0.221**	0.124**	0.074**	0.055	0.044
	(0.114)	(0.111)	(0.108)	(0.048)	(0.026)	(0.054)	(0.039)
Obs.	266	266	266	271	271	271	271
Unadjusted p-value	.828	.040	.043	.011	.006	0.309	0.254
Adjusted p-value	.828	.080	.086	.022	.012	0.309	0.254

Note: For each one of the estimated effects shown in the panels B and C of Table III and Table IV, this Table shows the corresponding unadjusted p-values as well as the p-value adjusted to take multiple-subgroup comparison into account. We used Holm's method (Holm, 1979).
** and * denote significance at the 5% and 10% levels, respectively.

Table A.VII: The effect of the treatment on performance on high school exit examinations, by cohort

Dependent variable	Cohort 2010					Cohort 2011				
	Grade 11 average score	Grade 12 average score	Overall average score	Graduation on time	Graduation	Grade 11 average score	Grade 12 average score	Overall average score	Graduation on time	Graduation
	(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)	(VIII)	(IX)	(X)
Panel A: All volunteer students										
Treatment	−0.107	0.027	0.019	0.027	−0.000	0.002	0.090	0.078	0.022	0.031
	(0.084)	(0.098)	(0.093)	(0.052)	(0.026)	(0.148)	(0.147)	(0.148)	(0.066)	(0.062)
N	388	388	388	395	395	154	154	154	161	161
Panel B: Mid-high ability										
Treatment	−0.243**	−0.186	−0.194	−0.063	−0.034	−0.280	−0.136	−0.179	−0.114	−0.043
	(0.104)	(0.162)	(0.152)	(0.079)	(0.045)	(0.251)	(0.281)	(0.272)	(0.108)	(0.124)
N	200	200	200	205	205	76	76	76	80	80
Panel C: High ability										
Treatment	0.022	0.284*	0.272*	0.107*	0.049*	0.180	0.103	0.125	0.173**	0.172**
	(0.140)	(0.143)	(0.139)	(0.062)	(0.026)	(0.205)	(0.196)	(0.200)	(0.081)	(0.061)
N	188	188	188	190	190	78	78	78	81	81
Panel D: Differential impact										
Treatment × high ability	0.266	0.470**	0.467**	0.170*	0.084*	0.460	0.239	0.304	0.287**	0.215
	(0.177)	(0.231)	(0.225)	(0.094)	(0.049)	(0.311)	(0.341)	(0.332)	(0.139)	(0.139)
N	388	388	388	395	395	154	154	154	161	161

Note: 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, pre-treatment ability score and socioeconomic family background as control variables. Columns (I) to (V) show the results for cohort 2010 and columns (VI) to (X) for cohort 2011.
Columns (I) and (VI) show the estimated effect of the treatment when the dependent variable is the average grade obtained at examinations taken at the end of grade 11, columns (II) and (VII) when the dependent variable is the average grade obtained at examinations taken at the end of grade 12 and columns (III) and (VIII) when the dependent variable is the average grade across all examinations. Columns (IV) and (IX) show the estimated effect when the dependent variable is a dummy indicating high school graduation on time and columns (V) and (X) when the dependent variable is a dummy indicating high school graduation at any time.
For each cohort, the first panel refers to the full sample of volunteers, the second panel to the mid-high ability subsample and the third panel to the high-ability subsample. The fourth panel shows the estimated difference in the effect of the treatment on high and mid-high students. The high-ability group corresponds to the top half of the distribution of pre-treatment ability scores across volunteers whereas the mid-high ability group corresponds to the bottom half. 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 socio-demographic controls and their interactions with the high ability dummy.
Standard errors clustered at the class level are reported in parentheses.
** and * denote significance at the 5% and 10% levels, respectively.

Table A.VIII: The effect of the treatment on scores at the *Baccalauréat* in the science track

	Grade 11				Grade 12, science subjects				Grade 12, humanities subjects				Final result	
Dependent variable	Average	French (written)	French (oral)	Personal work	Average	Math	Phy-Chem	Biology	Average	History	Philosophy	Lang. 1	Average	Grad.
	(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)	(VIII)	(IX)	(X)	(XI)	(XII)	(XIII)	(XIV)
Panel A: All volunteer students														
Treatment	−0.083	−0.027	−0.037	−0.086	0.171*	0.063	0.230**	0.162*	−0.082	0.050	0.037	−0.166	0.096	0.059
	(0.096)	(0.095)	(0.100)	(0.101)	(0.096)	(0.092)	(0.096)	(0.093)	(0.067)	(0.095)	(0.094)	(0.102)	(0.100)	(0.050)
N	340	340	340	340	339	339	339	326	340	340	340	340	340	353
Panel B: Mid-high ability														
Treatment	−0.278	−0.315**	−0.063	−0.160	−0.035	−0.063	0.129	−0.149	−0.284*	−0.107	−0.014	−0.514**	−0.216	−0.029
	(0.169)	(0.154)	(0.178)	(0.175)	(0.196)	(0.178)	(0.208)	(0.200)	(0.165)	(0.198)	(0.213)	(0.212)	(0.214)	(0.088)
N	129	129	129	129	129	129	129	120	129	129	129	129	129	138
Panel C: High-ability														
Treatment	0.023	0.119	−0.026	−0.035	0.281**	0.113	0.323**	0.341**	0.057	0.161	0.139	0.011	0.290**	0.113**
	(0.118)	(0.117)	(0.128)	(0.129)	(0.106)	(0.099)	(0.114)	(0.108)	(0.074)	(0.121)	(0.138)	(0.109)	(0.118)	(0.053)
N	211	211	211	211	210	210	210	206	211	211	211	211	211	215
Panel D: Differential impact														
Treatment × high ability	0.301	0.434**	0.037	0.125	0.316	0.176	0.194	0.490**	0.341*	0.268	0.153	0.525**	0.505**	0.142
	(0.204)	(0.188)	(0.219)	(0.204)	(0.212)	(0.187)	(0.236)	(0.229)	(0.189)	(0.232)	(0.275)	(0.239)	(0.250)	(0.093)
N	340	340	340	340	339	339	339	326	340	340	340	340	340	353

Note: The sample includes volunteer students from cohorts 2010 and 2011 who specialized in science in grades 11 and 12.
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, pre-treatment ability score and socioeconomic family background as control variables.
Column (I) shows the estimated effect of the treatment when the dependent variable is the average grade obtained at examinations taken at the end of grade 11; columns (II), (III) and (IV) show the effect for each subject taken that year. Column (V) shows the effect when the dependent variable is the average grade obtained at examinations taken at the end of grade 12 in science subjects; columns (VI), (VII) and (VIII) show the effect for each science subject taken that year. Column (IX) shows the effect when the dependent variable is the average grade obtained at examinations taken at the end of grade 12 in humanities subjects; columns (X), (XI) and (XII) show the effect for each humanities subject taken that year. Column (XIII) shows the effect when the dependent variable is the average grade across all examinations and column (XIV) when the dependent variable is a dummy indicating high school graduation on time.
The first panel refers to the full sample of volunteers, the second panel to the mid-high ability subsample and the third panel to the high-ability subsample. The fourth panel shows the estimated difference in the effect of the treatment on high and mid-high ablity students. The high-ability group corresponds to the top half of the distribution of pre-treatment ability scores across volunteers whereas the mid-high ability group corresponds to the bottom half. 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 socio-demographic controls and their interactions with the high ability dummy.
Standard errors clustered at the class level are reported in parentheses.
** and * denote significance at the 5% and 10% levels, respectively.

Table A.IX: The effect of the treatment on scores at the *Baccalauréat* in the humanities tracks

	Grade 11				Grade 12, science subjects			Grade 12, humanities subjects				Final result	
Dependent variable	Average	French (written)	French (oral)	Personal work	Math	Social Sc.	Sciences	Average	History	Philosophy	Lang. 1	Average	Grad.
	(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)	(VIII)	(IX)	(X)	(XI)	(XII)	(XIII)
Panel A: All volunteer students													
Treatment	−0.101	−0.097	−0.066	−0.128	−0.064	−0.115	0.019	−0.092	−0.168	0.092	−0.195	−0.118	−0.058
	(0.117)	(0.141)	(0.129)	(0.144)	(0.119)	(0.241)	(0.137)	(0.078)	(0.136)	(0.157)	(0.127)	(0.123)	(0.073)
N	202	202	202	202	184	110	201	199	202	202	199	202	203
Panel B: Mid-high ability													
Treatment	−0.219*	−0.155	−0.122	−0.316*	−0.130	−0.014	−0.012	−0.108	−0.281	0.097	−0.225	−0.185	−0.136
	(0.122)	(0.153)	(0.148)	(0.181)	(0.158)	(0.319)	(0.167)	(0.081)	(0.172)	(0.200)	(0.154)	(0.137)	(0.094)
N	147	147	147	147	132	72	146	144	147	147	144	147	147
Panel C: High ability													
Treatment	0.087	−0.147	0.049	0.193	0.027	−0.457	0.423	−0.326*	−0.115	−0.078	−0.548	−0.116	0.215*
	(0.316)	(0.366)	(0.186)	(0.366)	(0.255)	(0.516)	(0.359)	(0.186)	(0.300)	(0.300)	(0.336)	(0.230)	(0.117)
N	55	55	55	55	52	38	55	55	55	55	55	55	56
Panel D: Differential impact													
Treatment × high ability	0.306	0.008	0.171	0.509	0.157	−0.443	0.435	−0.218	0.166	−0.175	−0.323	0.069	0.350**
	(0.275)	(0.324)	(0.209)	(0.349)	(0.290)	(0.548)	(0.336)	(0.174)	(0.293)	(0.311)	(0.328)	(0.229)	(0.142)
N	202	202	202	202	184	110	201	199	202	202	199	202	203

Note: The sample includes volunteer students from cohorts 2010 and 2011 who specialized in humanities in grades 11 and 12.
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, pre-treatment ability score and socioeconomic family background as control variables.
Column (I) shows the estimated effect of the treatment when the dependent variable is the average grade obtained at examinations taken at the end of grade 11; columns (II), (III) and (IV) show the effect for each subject taken that year. Column (V) shows the effect when the dependent variable is the average grade obtained at examinations taken at the end of grade 12 in science subjects; columns (VI) and (VII) show the effect for each science subject taken that year. Column (VIII) shows the effect when the dependent variable is the average grade obtained at examinations taken at the end of grade 12 in humanities subjects; columns (X), (XI) and (XII) show the effect for each humanities subject taken that year. Column (XII) shows the effect when the dependent variable is the average grade across all examinations and column (XIII) when the dependent variable is a dummy indicating high school graduation on time.
The first panel refers to the full sample of volunteers, the second panel to the mid-high ability subsample and the third panel to the high-ability subsample. The fourth panel shows the estimated difference in the effect of the treatment on high and mid-high ablity students. The high-ability group corresponds to the top half of the distribution of pre-treatment ability scores across volunteers whereas the mid-high ability group corresponds to the bottom half. 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 socio-demographic controls and their interactions with the high ability dummy.
Standard errors clustered at the class level are reported in parentheses.
** and * denote significance at the 5% and 10% levels, respectively.

Table A.X: Peer group effects on students' achievement

	(I)	(II)	(III)	(IV)
		Panel A: All volunteer students		
Treatment	−0.107	−0.338**	−0.258**	−0.430**
	(0.084)	(0.134)	(0.124)	(0.146)
Treatment × Share of female students		0.385**		0.348**
		(0.152)		(0.153)
Treatment × Share of high-ability students			0.297*	0.225
			(0.168)	(0.175)
Obs.	388	388	388	388
		Panel B: Mid-high ability		
Treatment	−0.243**	−0.257	−0.399**	−0.385**
	(0.104)	(0.159)	(0.150)	(0.174)
Treatment × Share of female students		0.025		−0.031
		(0.229)		(0.236)
Treatment × Share of high-ability students			0.369*	0.374*
			(0.207)	(0.217)
Obs.	200	200	200	200
		Panel C: High ability		
Treatment	0.022	−0.221	−0.065	−0.275
	(0.140)	(0.221)	(0.266)	(0.315)
Treatment × Share of female students		0.373		0.364
		(0.258)		(0.252)
Treatment × Share of high-ability students			0.146	0.100
			(0.321)	(0.311)
Obs.	188	188	188	188

Note: The sample includes volunteer students from cohort 2010 only.
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 variables indicating either the proportion of girls or the proportion of high-ability students among the peers who were randomly assigned to the student's tutoring group. We use students' gender, pre-treatment ability score and socioeconomic family background as control variables.
The first panel refers to the full sample, the second panel to the mid-high ability subsample and the third panel to the high-ability subsample. The high-ability group corresponds to the top half of the distribution of pre-treatment 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. ** and * denote significance at the 5% and 10% levels, respectively.

Table A.XI: Balancing test - Post intervention survey respondents vs. non-respondents

Variable	Non-respondents	R − NR	s.e.	Obs.
Has repeated a grade	0.257	−0.041	(0.054)	305
Female	0.600	−0.009	(0.063)	305
High-SES family	0.210	+0.067	(0.059)	305
9th grade score (Maths)	0.433	+0.247**	(0.109)	283
9th grade score (French)	0.344	+0.118	(0.124)	284
Pursue an academic track	1.000	0.000	–	305
Specialize in science	0.667	+0.007	(0.007)	305

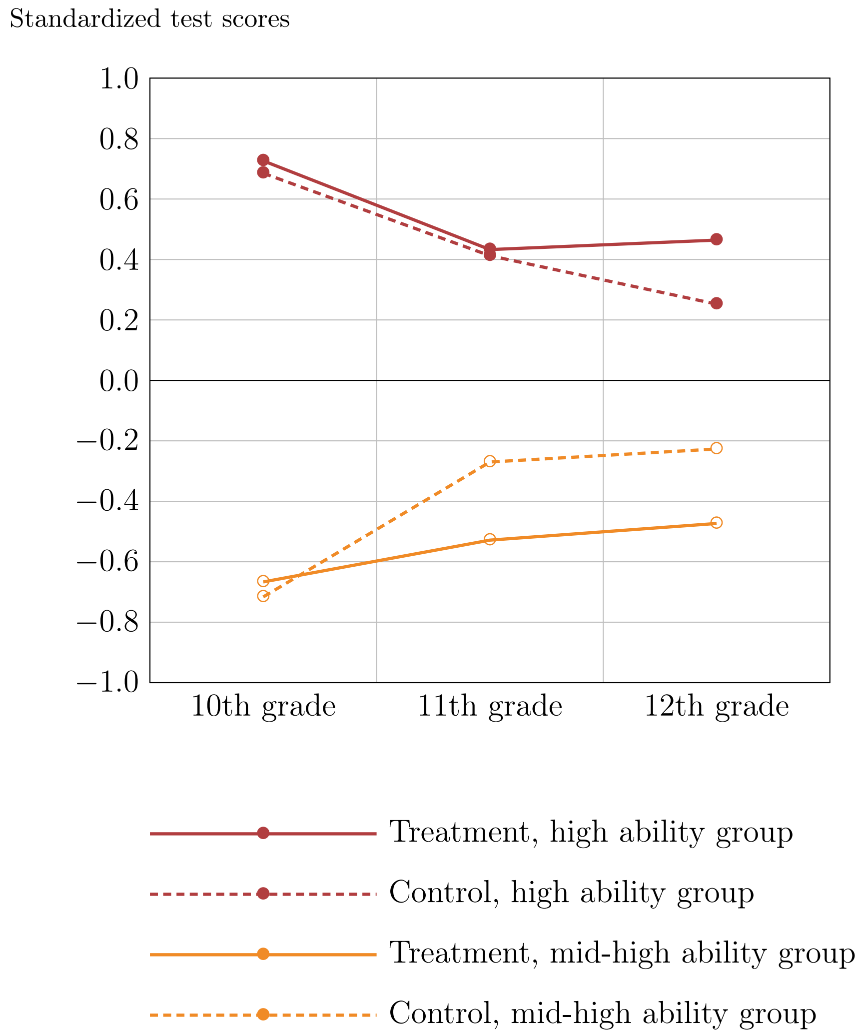
Note: This table shows differences in baseline characteristics between non-respondents (NR) and respondents (R) to the survey conducted in 2016 with former participants.
For each baseline variable, the first column gives the average for the group of non-respondents, the second column give the estimated difference between respondents and non-respondents (using regression where we control for cohort dummies, high school dummies and grade 11 major dummies), the third column gives the corresponding standard error whereas the last column gives the number of observations with non-missing values.
** and * denote significance at the 5% and 10% levels, respectively.

Table A.XII: The impact of tutors' gender and tutors' family background on former participants' perceptions

Variable	Baseline	Difference	s.e.	Obs.
Panel A: Impact of having a female tutor				
Tutor was close to tutees	0.829	+0.080*	(0.047)	200
Tutor was positive and encouraging	0.797	+0.138**	(0.046)	200
Tutor was difficult to understand	0.333	−0.048	(0.067)	200
Re-enlistment after the first year	0.556	−0.033	(0.060)	305
Panel B: Impact of having a tutor from a high family background				
Tutor was close to tutees	0.953	−0.100**	(0.043)	200
Tutor was positive and encouraging	0.884	−0.017	(0.057)	200
Tutor was difficult to understand	0.186	+0.174**	(0.071)	200
Re-enlistment after the first year	0.623	−0.053	(0.071)	305

Note: Building on the survey conducted in 2016 with former participants, this table shows the estimated impact of tutors' gender (panel A) and tutors' family background (panel B) on former participants' perceptions as well as on their probability to re-enlist at the end of the first year of the program. For each item of the 2016 survey, the first column of panel A shows the average for the group of former participants who had a male tutor, the second column shows the estimated impact of having a female tutor (using a regression where we control for cohort, high school and grade 11 major), the third column shows the corresponding standard error, whereas the last column shows the number of observations with non missing values. Similarly, first column of panel B shows the average for the group of former participants who had a tutor with a high SES family background, the second column shows the estimated impact of having a tutor with a low SES family background (using a regression where we control for cohort, high school and grade 11 major), the third column shows the corresponding standard error, whereas the last column shows the number of observations with non missing values.
** and * denote significance at the 5% and 10% levels, respectively.

Figure A.I: Evolution of the score gap between high and mid-high ability students in the treatment and control groups.



The figure shows changes in treatment effects on standardized test scores across grades, for the high (red) and the mid-high (orange) ability groups. 10th grade test scores are derived from continuous assessment by teachers, whereas 11th and 12th grade test scores are derived from high school national exit exams (externally set and marked). All scores are standardized to zero mean and unit variance across the sample of volunteer students

B Extensions of the theoretical framework

In this Appendix, we briefly discuss two possible extensions of our model. First, we consider the case where the mapping between time investment and knowledge accumulation is not one-to-one anymore. Second, we consider the case where time investment in school-related efforts is costly and where students have to take this cost into account when deciding their time investment strategy.

B.1 Time investment and knowledge accumulation

In this section, we assume that the mapping between time investment and knowledge accumulation is not given by Equation (1) anymore. Specifically, we assume,

$$K_{ci} = K_{0ci} + \phi_c(T_{ci}, K_{0ci}, K_{0fi}) \quad \text{and} \quad K_{fi} = K_{0fi} + \phi_f(T_{fi}, K_{0ci}, K_{0fi}) \tag{7}$$

where ϕ_c and ϕ_f represent production functions, namely the functions which relate the acquisition of new curricular and extracurricular knowledge to time investments and to initial endowments. In this framework, the impact of being assigned to the treatment group rather than to the control group can be written,

$$\begin{aligned} \Delta_i \;=\; & F(K_{0ci} + \phi_c(1-T, K_{0ci}, K_{0f}), K_{0f} + \phi_f(T, K_{0ci}, K_{0f})) \\ & -F(K_{0ci} + \phi_c(1, K_{0ci}, K_{0f}), K_{0f} + \phi_f(0, K_{0ci}, K_{0f})) \end{aligned} \tag{8}$$

Assuming that T is small, it yields,

$$\Delta_i \approx T \cdot F'_c \cdot \phi'_{fT} \left(\frac{F'_f}{F'_c} - \frac{\phi'_{cT}}{\phi'_{fT}} \right) \tag{9}$$

with $F'_c = \partial F(K_{0ci} + \phi_c(1, K_{0ci}, K_{0f}), K_{0f})/\partial K_c$, $F'_f = \partial F(K_{0ci} + \phi_c(1, K_{0ci}, K_{0f}), K_{0f})/\partial K_f$, $\phi'_{fT} = \partial\phi_f(0, K_{0ci}, K_{0f})/\partial T$ and $\phi'_{cT} = \partial\phi_c(1, K_{0ci}, K_{0f})/\partial T$.

In this setup, the issue is not anymore whether F'_f/F'_c is above or below 1, but whether it is above or below the ϕ'_{cT}/ϕ'_{fT} ratio, namely the ratio of the marginal product of time invested in curricular knowledge acquisition to the marginal product of time invested in extracurricular knowledge acquisition. If this ratio decreases with K_{c0}, we may observe positive treatment effects on higher ability students and negative treatment effects on lower ability students, even in case where F'_f/F'_c does not really vary across ability groups. For example, let us assume that higher ability students are better equipped to take advantage of the program so that the marginal product of time invested in extracurricular programs (especially elite ones) increases a lot with students' initial endowment in curricular knowledge. Assuming, at the same time, that the marginal product of time invested in curricular activities is not very different for lower and higher ability students, we could observe heterogeneous treatment effects, even in case where the marginal rate of substitution F'_f/F'_c is constant across ability groups.

B.2 Endogenous time investment

Our baseline model assumes that there is no cost associated with time investments in school-related efforts, so that the overall amount of time devoted to curricular or extracurricular activities is the same for all students and exogenously set at its maximum, namely $t_c + t_f = 1$. In this section, we relax this assumption and we assume that time invested in school-related efforts induces positive costs, denoted $C(t_c + t_f)$. In this setup, the amount of time devoted to school-related activities is not exogenous anymore, but the result of an optimization problem solved by the student.

For the sake of clarity, we will keep on assuming that the amount of time invested in extracurricular activities is exogenously set at $t_f = T$ for eligible participants and at $t_f = 0$ for non-eligible participants, so that students have only to choose the optimal amount of time invested in curricular activities. We assume that they make their choice so as to maximize the following objective function,

$$V(t_c) = F(k_{c0} + t_c, k_{f0} + t_f) - C(t_c + t_f) \quad \text{under the constraint} \quad 0 \leq t_c + t_f \leq 1 \tag{10}$$

In this setup, the optimal investment in curricular related activities is a function of K_{c0}, K_{f0} and t_f (denoted $t_c^* = t_c^*(K_{c0}, K_{f0}, t_f)$) which satisfies the familiar first-order condition,

$$F'_1(K_{c0} + t_c^*, K_{f0} + t_f) = C'(t_c^* + t_f) \tag{11}$$

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

$$\Delta_i = F(K_{0ci} + t^*_{cT}, K_{f0} + T) - F(K_{0ci} + t^*_{c0}, K_{f0}) \tag{12}$$

where t^*_{cT} denotes $t^*_c(K_{0ci}, K_{f0}, T)$ while t^*_{c0} denotes $t^*_c(K_{0ci}, K_{f0}, 0)$.

Assuming that T is small, it yields,

$$\Delta_i \approx T \cdot F'_c \left(\frac{F'_f}{F'_c} + t^{*\prime}_{c0} \right) \tag{13}$$

with $F'_c = \partial F(K_{0ci}+t^*_{c0}, K_{0f})/\partial K_c$, $F'_f = \partial F(K_{0ci}+t^*_{c0}, K_{0f})/\partial K_f$ and $t^{*\prime}_{c0} = \partial t^*_c(K_{0ci}, K_{0f}, 0)/\partial T$. Variable $t^{*\prime}_{c0}$ captures how time investment in curricular activities responds to exogenous increases in time invested in non-curricular activities. In the baseline model, any exogenous increase in t_f is followed by a symmetrical decline in t_c, regardless of students' ability (which amounts having $t^{*\prime}_{c0} = -1$ for all students). In the new model, this is not the case anymore, since $t^{*\prime}_{c0}$ likely varies across ability groups. Assuming that additional school-related efforts are more costly for lower ability students, $t^{*\prime}_{c0}$ is likely even more negative for these students, which could be one additional reason for why they do not benefit from the treatment.

For example, let us assume that the cost function is convex (with $C(x) = -cx^2/2$) and that the production function is linear-quadratic (with $F(k_c, k_f) = a_c k_c + a_f k_f - b_c k_c^2/2 - b_f k_f^2/2 + \sigma k_c k_f$). In such a case, it is not difficult to show that the optimal investment in curricular related activities is a linear function of the increase in extracurricular activities (T) induced by the program,

$$t^*_c = t_{c0} - \lambda T \tag{14}$$

where $t_{c0} = (a_c - b_c K_{c0} + \sigma K_{f0})/(b_c + c)$ and $\lambda = 1 - [(\sigma + b_c)/(c + b_c)]$. In this framework, the key parameter is λ, which is an increasing function of the cost parameter c. Assuming that parameter c is higher for lower ability students, the program has an impact on the time invested in curricular-related activities which is even more negative for lower ability students. In this framework, using Equation (13), we can show that the impact of the program is positive if and only if

$$\lambda < \frac{F'_f}{F'_c} = \frac{a_f - b_f K_{f0} + \sigma(K_{c0} + t_{c0})}{a_c - b_c(K_{c0} + t_{c0}) + \sigma K_{f0}} \tag{15}$$