The Role of Mindset in Education : A Large-Scale Field Experiment in Disadvantaged Schools

Elise Huillery[1], Adrien Bouguen[2], Axelle Charpentier[3], Yann Algan[4], Coralie Chevallier[5]

December 26, 2023

Abstract

This article shows that a four-year mindset intervention in disadvantaged middle schools led to a 0.05 standard deviation increase in GPA, associated with more optimistic beliefs, more self-criticism, and improved school behavior. Treatment effects were larger for students with relatively better social, academic, and discipline profiles. According to international empirical benchmarks, the mindset intervention is highly cost-effective but the effect size remains small despite repeated exposure over four years. These findings suggests that mindset interventions may not have the potential to transform education outcomes if not directly integrated in everyday-life pedagogy.

JEL classification: J24, I24, I38, J13, C9

Keywords: Non-cognitive skills, Growth mindset, Locus of Control, Diligence, Grit, Education, Gender Inequality, Social Inequality, RCT

We are deeply indebted to the *Energie Jeunes* team who put in endless effort to make the evaluation possible: Philippe Korda, Françoise Yon, Denis Yon, Léa Combette, and many others. We thank the Ministry of Education, in particular Isabelle Robin and Fabienne Rosenwald, for their support during the implementation of the project. We are grateful to Alice Danon, Anaëlle Solnon, Thiphaine Forzy, and Thomas Gesmond for outstanding project management, to Clémence Gleizes, Tereza Varejkova, Cloé Sarrabia, Olivier Frances, Sara Mouhoud, Nathan Viltard, and Clémence Pichon for excellent research assistance, and to many field assistants who collected the data in the schools. Finally, we gratefully acknowledge the staff, teachers, and students for the time and information they shared with us. This research is supported by a grant of the French Fonds d'Expérimentation pour la Jeunesse, the Porticus Foundation, EUR FrontCog grant ANR-17-EURE-0017, and the European Research Council under the European Community's Horizon H2020 Programme (GA n° 647870) for Yann Algan's ERC SOWELL. Adrien Bouguen also acknowledges financial support from the German Research Foundation (DFG) project SFB 884 during his stay at the University of Mannheim. Coralie Chevallier also acknowledges financial support from the Agence nationale de la recherche (project ANR-21-CE28-0009). The project was approved by the Institutional Review Board at the Paris School of Economics (Reference 2014-007). This study is registered in the AEA RCT Registry (AEARCTR-0000376). All errors are our own.

[1]Co-first author, Université Paris-Dauphine, Université PSL, and J-PAL: elise.huillery@dauphine.psl.eu

[2]Co-first author, Santa Clara University, California: abouguen@scu.edu

[3]Ministère de l'Education nationale et de la Jeunesse, Direction de l'évaluation, de la prospective et de la performance: axelle.charpentier@education.gouv.fr

[4]HEC: algan@hec.fr

[5]Ecole Normale Supérieure - PSL, Inserm, Paris: coralie.chevallier@ens.psl.eu

1 Introduction

The way adolescents perceive themselves, how they assess their chances of success and their expected return to effort, might be just as important for academic performance as external factors such as class size or teacher quality. A large number of studies have highlighted a positive correlation between students' mindset and their educational outcomes (Almlund et al., 2011; Castillo et al., 2011; Dohmen et al., 2011; Duckworth and Seligman, 2005; Golsteyn et al., 2014; Moffitt et al., 2011; Sutter et al., 2013), and J. J. Heckman et al. (2006) first established causality from self-esteem and locus of control to schooling decisions. Based on these results, many governments have encouraged a shift of educational priorities to promote the development of socio-behavioral skills at school. The US Department of Education, for instance, identified the promotion of grit, tenacity, and perseverance as "critical factors for success in the 21st century". The UK also launched a multi-million-pound push to improve mindset education.[1] An important question is whether these socio-behavioral skills can indeed be taught, and whether such training has the potential to have a strong downstream impact on educational outcomes. Current scientific evidence suggests that short mindset programs can have positive effects at low cost but these effects are typically small. It is not clear whether repeated mindset interventions may meet policymakers' high hopes in terms of improving academic performance and reducing social inequality at school.

In this paper, we test the impact of *Energie Jeunes*, a program conducted in French disadvantaged middle schools to improve adolescents' academic prospects by changing their mindset. The program builds on three components: first, the growth mindset component teaches students that the brain is highly plastic and grows smarter when it experiences regular schoolwork, and that failures are temporary and signal a learning opportunity. Second, the internal locus of control component emphasizes the role of effort and encourages students to downplay the importance of external constraints such as physical handicaps, family background, teacher quality, or peer influence. Finally, the behavioral component provides tools to operationalize the growth mindset and internal locus of control in practice. These tools promote diligence through routinized effort, concentration, perseverance, and goal setting. They are presented to students as a way to sustain a behavioral change that derives from the change in mindset. We refer to the program as a 'mindset intervention' since it primarily targets beliefs, with operational tools to translate these beliefs into behavioral change. This program may be particularly important in countries, like France, where students demonstrate low feeling of competence and internal locus of control (Algan et al., 2018), and those from disadvantaged families develop biased perceptions of their academic potential (Guyon and Huillery, 2021).

[1] https://www.gov.uk/government/news/england-to-become-a-global-leader-of-teaching-character

The novelty of the program is that it runs from Grade 6 to Grade 9, offering repeated exposure over four years. Each year, students participate in three one-hour in-class sessions presented by external facilitators. This paper thus examines whether repeated exposure to small and low-cost mindset interventions has any chance of generating an economically significant change in student performance. Moreover, we study the effect of an intervention conducted at scale among a large sample of students and use a rich set of subjective and objective measures of behavior and academic outcomes, which provides a unique opportunity to clarify the ongoing debate on the efficacy of mindset interventions.

We conducted a field experiment in 97 disadvantaged middle schools which are representative of French middle schools in Priority Education, i.e., disadvantaged schools receiving extra support from the State. In each school, students who entered Grade 6 in September 2014 and in September 2015 were included in the study. We randomly assigned one of these two cohorts to either the treatment or the control group. Therefore, in half of the schools the 2014 cohort benefited from the program from Grade 6 to Grade 9 while the 2015 cohort did not, and *vice-versa* in the other half of the schools. Our sample consists of nearly a thousand classes and more than 23,000 students, which grants a high degree of internal and external validity to this experiment. We collected data from three sources: first, the administrative data provides information on academic performance through students' grade point average (GPA) and national exam scores, and on students' behavior at school. Second, the teacher survey provides a measure of students' in-class work attitude. Finally, the student survey provides information on students' growth mindset and internal locus of control, self-reported diligence, and educational and professional aspirations. Our hypothesis is that the program first affects students' mindset (beliefs), then their behavior (effort), and finally their academic performance and aspirations.

Intention-to-treat estimates show that the intervention increased GPA by 0.05 standard deviation (hereafter, SD) by the end of middle school. This corresponds to a 0.07 SD increase in GPA for students who participated in at least one *Energie Jeunes* session in Grade 9. Importantly, in earlier grades, treatment effects were somewhat smaller but not significantly different from the one found in Grade 9. Also, students exposed during four years benefited as much as those exposed during shorter time periods. Thus, repeated exposure to the intervention did not amplify treatment effects proportionally. While the program did not increase the national exam scores overall (+0.01 SD), we find a small significant treatment effect on the literacy score (+0.03 SD). Finally, we find a significant +0.04 SD increase in the score of educational and professional aspirations.

Regarding intermediary outcomes, we found a 0.11 SD increase in a mindset index measuring students' growth mindset and internal locus of control, and a 0.06 SD increase in an index of school-reported behavior in Grade 9, which captures the

fact that students are less likely to be absent, late, undisciplined, or violent. We also found improved work attitude in class reported by teachers in Grades 7 and 8. Paradoxically, treatment students themselves did not report more effort: their appreciation of time spent on homework was unchanged and their self-perceived grit, self-discipline, and homework management was lower in Grades 7 and 8 than in the control group. The discrepancy between objective and subjective measures of behavior suggests that students became more self-critical when exposed to the program, may be due to the social comparison effect (Festinger, 1954).

We then examine treatment heterogeneity of academic outcomes, who benefited more, and whether benefits were concentrated among subgroups in a way that helps explain mechanisms. We first use pre-defined heterogeneity groups and then a more synthetic machine learning approach. We found that more disciplined subgroups adjusted their beliefs more intensively and benefited more from the program in academic terms, while less disciplined subgroups benefited less in terms of academic performance but generally gained more in terms of behavior. The machine learning method identifies a subgroup whose GPA improved by 0.09 SD and whose national exam scores improved by 0.05 SD. This subgroup had initially better social, academic, and discipline profiles. We also find that the improvement in mindset is more pronounced in this subgroup, while the improvement in behavior is not. Since strongly affected students already behave better than weakly affected students, it is possible that we could not capture more subtle improvements in behavior other than absences, lateness, misconduct, and work attitude in class. Moreover, the machine learning approach reveals that higher academic gains were associated with losses in self-assessed discipline, which reinforces the idea that students' subjective perceptions should not be taken at face value and raises the question of whether self-criticism was useful to spur greater changes and impacts.

This paper provides several contributions. First, it contributes to the growth mindset literature, which has recently become controversial. A recent meta-analysis raised concerns about the internal and external validity of many studies, concluding that only six studies are high-quality and that the average effect of these studies is non-significant (+0.02 SD) (Macnamara and Burgoyne, 2022). Yet, this review excludes some important, and well-conducted, experiments (Alan et al., 2019; Bettinger et al., 2018; Outes-Leon et al., 2020). On the other hand, David S Yeager and Dweck (2020) review the literature and conclude that mindset interventions have small effects but are cost-effective. Our paper contributes to this literature by providing highly internally and externally valid results showing that mindset interventions do have positive impacts, small in size but highly cost-effective. For instance, the median effect of the randomized trials that evaluate interventions in disadvantaged secondary schools is 0.03 SD, although these interventions cost several thousand US dollars per student (Boulay et al., 2018). Similarly, a review of

747 education studies (randomized or not) found a median effect size of 0.03 SD in studies using large (above 2,000 students) samples, for an average cost of $882 per pupil (Kraft, 2020). By comparison, the cost-effectiveness of the *Energie Jeunes* program is outstanding: the effect size is of 0.05 SD on GPA for a total cost of €65 (i.e., $75) per student.

Besides, this paper uses a rich set of subjective and objective measures of mindset, behavior, and academic outcomes. Most papers in the literature document effects on academic performance and beliefs but not on students' behavior (Bettinger et al., 2018; David S Yeager, Hanselman, et al., 2019). Only one paper, Alan et al. (2019), provides evidence of changes in both mindset and behavior to explain effects on academic performance. Our paper is thus the second to test the full theory of change that goes from beliefs to behavior and finally educational outcomes. Moreover, our study is the first to provide evidence that a mindset intervention can change everyday school-life behavior such as attendance, sanctions, and work attitude in class, and contrast these objective measures with subjective self-assessment.

Third, we improve the understanding of who benefits from mindset interventions by showing that the effects on educational outcomes are larger students with better social, academic and discipline profiles. This result may seem at odds with the existing literature, which finds larger effects for students with initially worst academic predictors (Bettinger et al., 2018; David S Yeager, Hanselman, et al., 2019). The fact that our sample is composed of disadvantaged schools may in part explain the discrepancy: relatively better students in disadvantaged schools may not be high-achievers in the national distribution. This paper sheds new light on the fact that mindset interventions may not help the most fragile students.

Finally, our results shed light on the potential of mindset interventions to improve educational outcomes. The intervention is similar to the ones implemented in all existing papers (Bettinger et al., 2018; Outes-Leon et al., 2020; Paunesku et al., 2015; David S Yeager, Hanselman, et al., 2019), except Alan et al. (2019): it is delivered by external actors who come to the school to deliver the message directly to the students. The homeroom teacher is present at the back of the room but she is not involved and plays no role in the program. Thus far, the literature has shown that such light-touch short interventions led by external actors are cost-effective but generate small impacts of about 0.05 SD on students' performance. Our paper shows that repeating the intervention 12 times over four years does not multiply the impact. In fact, the treatment effect is of similar magnitude as in existing papers (+0.05 SD), and we see in our data that it does not seem to amplify much when exposure increases. This casts doubt on the potential of such light-touch interventions to be truly transformative. In contrast, Alan et al. (2019) train the teachers who are in daily contact with pupils, which produces a much more intense intervention, with a visibly larger impact on a math test two years after the intervention. Taken

together, we think that these studies collectively inform policymakers that relying on external actors may not transform educational outcomes, while involving teachers and parents who are in daily contact with pupils may be more promising.

The rest of the paper is structured as follows. Section 2 describes the intervention and theory of change. Section 3 presents the data and Section 4 the empirical strategy. Section 5 presents the main results, Section 6 discusses treatment effect heterogeneity and what makes the mindset intervention effective, and Section 7 concludes.

2 The Intervention

2.1 The French Context

The study takes place in French middle schools in priority education zones, i.e., disadvantaged schools receiving extra funding from the State. In France, middle school lasts four school years, from Grade 6 to Grade 9, with most students aged 11 to 14. Core courses (math, French, history, life science, etc.) constitute the vast majority of the curriculum. Students may also choose a few optional courses such as Latin or Ancient Greek. Each course follows the same national curriculum.

At the beginning of the school year, students are assigned to one class and take all their courses with this class, i.e. with the same group of peers, except for optional courses. Classes change from one year to another so one student may interact with any other student of her grade over the course of middle school. Regarding teachers, a group of 8-10 teachers, one per discipline, is assigned to each class at the beginning of the year for one year, one of them being designated as homeroom teacher (*professeur principal*). Each teacher usually teaches several grades in the middle school in the same year.

Every quarter, teachers average the grades and fill out a report card. These grades constitute the GPA which is used to determine whether the student can advance to the next grade at the end of the school year. At the end of Grade 9, teachers make a strong recommendation as to which high school (academic, technical or professional) the student should choose. Families can follow or ignore this recommendation. Besides, Grade 9 students take a national exam at the very end of the year which is low stake since their assignment to high school has already been decided. However, this national exam constitutes the first official exam to grant a diploma in France.

2.2 The Intervention

The *Energie Jeunes* program aims at improving students' performance at school by developing their motivation, effort, and self-discipline in disadvantaged schools in

France. The program was developed by a French non-profit organization (hereafter, the NGO) created in 2009. It consists of twelve 55-minute class interventions, three per year from Grade 6 to Grade 9. The NGO enlists dozens of facilitators who are responsible for conducting the interventions in class in pairs. The sessions are included in the homeroom hours allocated to the homeroom teacher to discuss extra-curriculum subjects such as class climate, societal debates, civil rights, etc. (*heures de vie de classe*). Homeroom hours are mandatory, therefore participation to the *Energie Jeunes* program is also mandatory and parents do not give their consent. During the *Energie Jeunes* sessions, facilitators present slides, play videos, organize activities and debates following a detailed standardized script. Homeroom teachers are present at the back of the classroom but play no role in the intervention. In that sense, the mode of delivery is close to most interventions studied in the existing literature, which also rely on external actors. This policy may be attractive to schools because it is cheaper than training the teachers, which was done in Alan et al. (2019). However, relying on external actors may be less efficient because, unlike teachers, they do not interact with students over the entire school year.

The educational content of the program is based on recent research in psychology (Walton, 2014). The program essentially focuses on three components: the growth mindset of intelligence (Dweck and David S Yeager, 2019), the internal locus of control (Duckworth, Quirk, et al., 2019), and diligence. First, the program conveys the message that the human brain is highly plastic, that intelligence is not fixed, and that working hard and on challenging tasks can develop intelligence. The program also emphasizes that setbacks and challenges are normal and constitute opportunities to learn, and that they should not be interpreted as signs of low innate abilities. The growth mindset content is present in every session of the program and is applied to a variety of domains (sports, music, arts, or academia) using different formats (e.g., videos, class discussions, slides, and case studies).

Second, the locus of control is also central in the program. Students are encouraged to interpret experiences as within their own agency and to embrace the idea that success is possible for everyone through hard work, even when one faces external constraints. For example, one video features the story of a handicapped person who became an Olympic medallist; another video features a man who grew up in a slum and became the CEO of an international firm. These materials de-emphazise the role of external constraints and highlight the role of effort and perseverance.

Finally, the program provides operational tools to increase diligence. This component of the program includes advice to routinize effort, increase concentration, build healthy work habits, and minimise distractions, through mini-cases and reflection on students' own experience. Facilitators discuss potential sources of distraction or discouragement and strategies to overcome them, such as keeping one's cellphone away when doing homework, being attentive in class, or being bold enough to resist

peer pressure when one tries to be attentive in class. To materialize intentions into actions, students are asked to make a commitment during the second session of the year. Examples of such commitments include: stop chatting in class, improve one's math average grade, or do one's homework before playing video games. During the third session each year, students assess whether they were able to honour their commitment and analyze the reasons for their success or failure.

We provide in Appendix A the exact program's content of the twelve sessions. For each grade, Figure A1 reports the messages, videos, and activities of the first session (November-January), Figure A2 the second session (January-March), and Figure A3 the third session (March-May).

2.3 The Theory of Change

Our theory of change was pre-registered on AEA RCT Registry (AEARCTR-0000376). We posit that the program affects students in three steps.

First, we expect the program to change students' mindset in the direction of higher perceived chances of success and return to effort. In fact, growth-mindset increases the perception of cognitive ability to succeed, while internal locus of control decreases the perception of the situational or contextual factors that are deemed necessary to succeed.

Second, we expect this change in mindset to translate into a change in behavior. The third component of the program also helps here: the combination of updated beliefs and practical tools is critical to trigger behavioral change. Students with fixed mindsets do not invest in effort because they tend to believe that trying hard or asking for help signals low ability. Change in mindset is therefore a necessary condition for behavioral change. Still, the practical tools may also be important to help students close the intention-to-action gap. Precisely, we expect behavior to change in the direction of increased effort and diligence in the form of, for instance, more homework, higher concentration when doing homework and in class, higher participation in class, fewer disciplinary sanctions, less lateness, fewer absences, etc.

Third, we expect changes in behavior to translate into better performance both at continuous assessment and national exam in Grade 9. We also expect that increased performance together with more optimistic beliefs increase academic ambition and shift aspirations towards higher levels of education and occupation.

Finally, as noted in our pre-analysis plan, we expect some heterogeneity in the impacts according to the level of understanding, attention, and need of the students. On the one hand, students who are more likely to listen attentively to the facilitators and to participate actively in the activities and discussions may benefit more from the intervention. On the other hand, we may find larger impacts for students who need it the most, i.e., those who have a fixed mindset and external locus of control and invest less effort at school, because their margin of improvement is larger. We

will not be able to directly measure precisely these dimensions of heterogeneity but will use proxies based on baseline data.

3 The Data

3.1 Sampling Strategy

Schools In France, priority education schools represent about 20% of middle schools, non-priority public schools 60%, and private schools the remaining 20%.[2] The *Energie Jeunes* program targets public disadvantaged middle schools, mostly in Priority Education. Our baseline sample includes 97 middle schools that volunteered to be part of the experiment, located in seven out of 25 regional school districts in metropolitan France. Appendix Tables A2 compare the characteristics of the sampled schools to priority education schools, public schools, and public+private schools. The sampled schools look fairly similar to all priority education middle schools in terms of socio-economic composition, performance at national exams, and teacher characteristics, whereas they are significantly more disadvantaged than the typical middle school in France. The proportion of students from a high-SES background is half as large, the proportion of financial aid beneficiaries twice as large, and performance in national tests much lower than the national average. Our results are thus likely to generalize to the population of students in priority education areas.

Classes The study includes two cohorts of students, those who entered Grade 6 in 2015 and in 2016. All classes in these two cohorts took part in the experiment except a few special-need classes which host only 3% of students not targeted by the program. Our baseline sample contains 1,026 classes in Grade 6, 521 treatment and 505 control classes.

Students All students registered in a class of a sampled school in one of the two cohorts of study were included in the study. This means all students registered in the sampled schools in Grade 6 in 2015 and 2016, in Grade 7 in 2016 and 2017, in Grade 8 in 2017 and 2018, and in Grade 9 in 2018 and 2019. Our sample includes between 23,000 and 24,000 students in each grade, equally distributed between the control and the treatment group (Table 1, first panel, first column). For the administrative data we aimed to collect information on all the sample, whereas for survey data we randomly selected seven students per class to take the questionnaire, and on whom teachers had also to fill out a questionnaire. We conducted the random selection every year so the students' sub-samples were different from one year to another, and representative of the full population.

[2] https://www.education.gouv.fr/reperes-et-references-statistiques-1316

3.2 Data Sources

All data were collected every year in the Spring, from 2015 to 2018 for the first cohort and from 2016 to 2019 for the second cohort.

Administrative data Administrative data from the schools and the Ministry of Education first provide students' grade point average (GPA), which is based on teachers' continuous assessment. GPA should not be biased in favor of the treatment group because only one teacher per class attended the *Energie Jeunes* sessions, out of ten teachers grading students, so the majority of teachers were unaware of the program and of the classes that were in the treatment or control group. Moreover, teachers teach more than one cohort so many of them likely teach both control and treatment classes. More importantly, teachers had no vested interest in the program, it seems inconceivable that they tried to favor a program that was not designed nor implemented by them, and imposed on them by the school principal. Second, administrative data provides the national end-of-school exam scores. This exam is externally and anonymously graded. Even though it is not required to enroll in high school, it delivers the first official diploma and virtually all ninth graders take it (97% in our sample). The exam includes math, French, history-geography, sciences, and an oral examination based on a personal project. Third, administrative data also includes students' number of absences, lateness, sanctions for minor misconduct, and disciplinary actions for serious offenses. Finally, it also provides students' socioeconomic status, gender, year of birth, and country of birth. The administrative data are available for all students in the sample.

Student survey The student survey was administered on digital tablets to the sub-sample of seven randomly selected students per class to measure participation in the program, mindset, self-assessed diligence, as well as educational and professional aspirations. We use instruments validated in the psychology literature. Students' mindset captures two components: growth mindset and internal locus of control. Self-assessed diligence is measured by six components: self-reported orderliness, grit, school-work impulsivity, work discipline, homework management, and hours of homework (the last two components were not collected in Grade 6).[3] The exact items and sources are detailed in Appendix Tables A1. Finally, educational and professional aspirations are measured based on which type of high school and job students aspire to.

[3]Initially, our student survey included the Academic Diligence Task developed by Galla et al. (2014) as a behavioral measure of diligence. Our prior was that a behavioral task would do better than questionnaires to measure diligence, but it turned out that it was no more reliable, and less valid, than self-reported and teacher-reported measures of diligence, a result that we show in a companion paper (Boon-Falleur et al., 2022). Given that the task is a costly instrument both financially and logistically, we removed it from the survey in the last two years and do not use it in this paper.

Teacher survey The teacher survey provides a third source of information on students' behavior. We administered a French version of the Character Report Card developed by Park et al. (2017) to assess three dimensions of students' in-class attitude: "achievement character" (grit, diligence, schoolwork self-control, which predict higher grades), "intellectual character" (zest and curiosity, which predict greater participation in class and engagement in learning), "social character" (gratitude, optimism, social intelligence, interpersonal self-control, which predict less peer conflict and greater popularity). Teachers' view on student behavior is crucial as it provides a third-party evaluation in case the intervention affected students' reference point i.e. students judging their behavior harsher after having been exposed to the intervention. For the same reasons as for GPA, we consider that teachers provide an objective measure of students' in-class behavior.

3.3 Variables of Interest

Take-Up We use four measures of participation collected yearly at the end of the student questionnaire: one dummy equal to 1 if the student reports having participated in at least one *Energie Jeunes* session during the school year, the number of sessions attended this year (in principle, three), whether the student reports having made a commitment, and whether she reports having honored her commitment.

Mindset Hereafter, we favor summary indices to avoid inference issues due to multiple hypothesis testing. Each index is constructed as a weighted mean of related standardized items. Signs are switched where necessary so that the positive direction always indicates a "better" outcome, and all items are demeaned and divided by the standard deviation of the control group. We weight each item using the methodology proposed in Anderson (2008), which ensures that correlated items receive less weight, while uncorrelated items, which add more information, receive more weight. We use the same aggregation method for all indices. The summary index of mindset combines all items related to the growth mindset of intelligence and to the internal locus of control listed in Appendix Table A1. We also construct the two separate sub-indices of growth mindset and internal locus of control to use in the Appendix.

Self-assessed Diligence We construct a summary index aggregating the items of self-assessed diligence: orderliness, grit, school-work impulsivity, work discipline, homework management, and hours of homework (see all items in Appendix Table A1). We also construct separate indices for each component to use in Appendix.

Teacher-reported Character We construct a summary index of Teacher-reported character using the same methodology as described above and all 24 items included in the Character Report Card (see Appendix Table A1). We also used separate in-

dices for the three main factors measured by the Character Report Card: "achievement character" (grit, diligence, schoolwork self-control), "intellectual character" (zest, curiosity), and "social character" (gratitude, optimism, social intelligence, interpersonal self-control).

School-reported Behavior We construct a summary index of the respect of school rules aggregating information from school administrative registers: yearly number of absences (counted in half-days, meaning that any hour missed counts as a half-day), number of times the student was late, number of sanctions for minor misconduct, and number of disciplinary actions for serious offenses. This index provides a different and complementary measure of behavior than teacher-reported character: it captures violations of school rules in and out of the classroom, whereas teacher-reported character captures variations in work attitude in class. The separate items are used in Appendix.

Educational and Professional Aspirations We measure aspirations at the end of Grade 9. There are two types of high schools in France: technical high schools (including a two-year track and a three-year track) and academic high schools. We use a dummy equal to 1 if the student aspires to go to an academic high school to indicate a "better" outcome, 0 if the student aspires to go to a vocational high school or expects to repeat Grade 9. We also asked students an open-ended question about their career aspiration and coded the answers using National Institute for Statistics and Economic Studies job classification. We create three categories: low-skilled job (farmer, craftsman, storekeeper, manual labourer, low-skilled office worker), intermediate job (e.g., nurse, primary school teacher, accounting officer, technician), and high-skilled occupation (e.g., lawyer, doctor, journalist, computer programmer). Students who want to be soccer players, actors, or singers were assigned to the low-skilled job category—unless they mention selective tracks like the conservatoire or college of music. Finally, we create a dummy for the students who answered that they do not know. We construct a summary index of aspirations with two items: the student aspires to an academic high school, and the student aspires to a medium- or high-skilled job.

Academic Performance We use both GPA, which averages grades from all courses, and scores at the national end-of-middle school exam, which is anonymous, taken in the very last days in Grade 9, and externally graded. GPA is our main final outcome because it plays a crucial role in France. Teachers use a 0-20 point scale and grade the students based on predefined expected competences. It is therefore considered as an absolute measure of academic performance, which contrasts with grading systems in other countries where students are graded in relative terms ("*on the curve*") using percentiles or a predefined distribution. Teachers in elite schools

may adjust the expected competences upwards relative to teachers in disadvantaged schools and some teachers are more demanding than others across all schools, but these variations do not depend on the presence of the *Energie Jeunes* program, of which most teachers are unaware and which they do not take into account when assessing students' expected competences. Moreover, in the French context, GPA has a normal distribution averaging typically at 12-13/20 (11/20 in our sample), and only a small proportion of students achieve more than 16/20 (5% in our sample). In fact, the philosophy of teachers is to place the bar of expected competences high enough to avoid ceiling effects and discriminate finely among students. Finally, GPA is the only measure of performance used by the French education system to assess students' academic performance, which is key to decide whether students go on to the next grade. At the end of Grade 9, GPA is also the only metrics that will determine whether students join vocational or academic high school tracks and allocate them to more or less selective high schools. In contrast, the national final exam is low-stakes as it plays no role in students' assessment and tracking so its importance is purely symbolic. Therefore, we present both measures but lend more weight to GPA. We standardize both the GPA and exam scores.

4 The Empirical Strategy

4.1 The Experimental Design

With the support of the Ministry of Education, the NGO contacted schools to offer them to participate in the study, most of them being already part of the program. In September 2014, 97 middle schools volunteered, located in seven different regional school districts: Aix-Marseille, Créteil, Amiens, Lille, Lyon, Paris, and Versailles. The two cohorts of students included in the study were randomly assigned to the treatment group or the control group. The treatment cohort received the program during the full duration of middle school, whereas the control cohort received no intervention. Only students who were held back at some point may not comply, but this concerns only 0.5% of the students in Grades 6, 7, and 8, and 2% in Grade 9.[4]. The experimental design is represented in Appendix Figure A1.

This design assumes that the risk of spillover across cohorts is small enough to guarantee internal validity. Such spillovers could come from two sources, both minimal. First, friends or siblings enrolled in two consecutive years, but in fact most friendship take place within cohort and the proportion of students who have a sibling in an adjacent cohort is negligible. Second, teachers who attend the program and convey the messages to control students, but the program does not target teachers and it does not modify teaching content. Although minimal, such spillovers would

[4]MENJ-DEPP, Note d'Information 19.46, November 2019, education.gouv.fr/statistiques.

lead to an attenuation bias and run against finding a treatment effect.

On the upside, this design has three advantages: first, the randomization is conducted within schools so all schools benefit equally from the *Energie Jeunes* program, which facilitated willingness to participate in the experiment. Second, this design allows to use within-school variations in treatment assignment, which significantly improves statistical precision compared to school-level randomization. Finally, cohort-level randomization limits the salience of the experiment compared to a class- or individual-level randomization, which reduces the risk of Hawthorne and John Henry effects. In this experiment students do not apply to the program, the school staff does not advertise it, control students are surrounded by classmates who do not benefit either, so they are likely to ignore the existence of the program completely. Moreover, those who may hear about the program being offered in other cohorts should not react much because the program is not particularly attractive. The SUTVA assumption is thus credibly respected.

4.2 Balance Checks and Attrition

Attrition rates vary according to the source of the data. We have minimal (0-4%) attrition in the administrative data (Table 1, Columns 1-4), moderate attrition (5-22%) in the student survey and national exam data (Table 1, Columns 5-9), and substantial attrition (22-42%) in the teacher survey data (Table 1, Columns 10-13). Attrition was due to the increasing difficulty of calling on schools during five years, and more particularly on the teachers.

However, the internal and external validity of the experiment is not affected. First, Table 1 shows that attrition was always balanced across the treatment and control groups. Second, the balance checks in Appendix Tables A3, show that, despite attrition, the students included in the study remain similar across years for all samples (administrative sample, student survey sample, and teacher survey sample), and across the treatment and control groups. We detect some significant but small differences between treatment and control students in some samples and some grades, which is expected due to random variations and our high detection power, but overall randomization worked well and attrition did not modify the composition of samples and experimental groups.

4.3 Estimation Strategy

We use intention-to-treat estimates, meaning that we analyze data for all students enrolled in a school-cohort randomized to an experimental condition. Our sample varies from one year to another because students enter and exit the sampled schools due to registration changes (e.g., house moves, parental decision). Table 1 describes exits from and entries in sampled schools. The proportion of students who were not

registered in Grade 6 (entries) is 7% in Grade 7, 16.5% in Grade 8, and 21.8% in Grade 9 in the control group and does not statistically differ in the treatment group. Similarly, the proportion of students who left after Grade 6 (exits) is 11.8% in Grade 7, 15.1% in Grade 8, and 19.8% in Grade 9 in the control group, and also does not differ in the treatment group. Entries and exits did not affect the comparability of the treatment and control groups (see Section 4.2) but mechanically decreased take-up: 22% of Grade 9 students were not registered in Grade 6 so received partial treatment. We include them in the analysis in a intention-to-treat approach.

To test the null hypothesis that the program had no impact on students, we estimate the average treatment effect separately for each Grade j:

$$Y_{iscj} = \alpha_j + \beta_j T_{sc} + X_i \lambda_i + \theta_s + \theta_c + \epsilon_{iscj} \tag{1}$$

where Y_{iscj} is the outcome of Grade j's student i in school s and cohort c, T_{sc} is a dummy that equals 1 if cohort c in school s is in the treatment group and 0 otherwise, X_i is a vector of student baseline covariates, θ_s is a vector of school fixed effects, θ_c is a cohort fixed effect, and ϵ_{iscj} is the error term. The estimated β_j is the average intention-to-treat effect in Grade j. The equation is estimated via OLS, and standard errors are robust to heteroscedasticity and are clustered at the school-cohort level, which is the unit of randomization. The number of clusters is 194 in Grade 6 and Grade 5, 190 in Grade 8 and 186 in Grade 9 (due to school-level attrition discussed above).

Student baseline covariates included in X_i are selected by a double-LASSO procedure (Belloni et al., 2013). The following baseline variables are included in the procedure: gender, year of birth, whether the student was ever held back before or in Grade 6 / was on time in Grade 6 / was ahead in Grade 6, whether she is a foreigner, whether she was born in a foreign country, whether family receives income-based financial aid (hereafter, "aid recipients" *versus* "non-recipients"), financial aid level (1, 2, 3 or 4), socioeconomic status (advantaged, intermediary, or disadvantaged), at least one parent works, single- or double-parent family, broken family, parent 1's number of children, and parent 2's number of children. Categorical variables are dichotomized and non categorical variables are squared. In the double-LASSO, we also include 'early' academic performance (above or below median GPA in the first quarter in Grade 6, hereafter, "high-achievers" *versus* "low-achievers"), and 'early' school behavior (above or below median school-reported behavior in the first quarter in Grade 6, hereafter, "well-behaved students" *versus* "poorly-behaved students"). We call these variables 'early' and not 'baseline' because the program started in the second or third month in Grade 6 and these variables are measured during the first three months in Grade 6. However, the overlap between these measures and the intervention is minimal: the average student present in the data in grade 9 had received 3.5% of the intervention during her first semester in Grade 6, meaning that

the treatment could not affect substantially the measures of early GPA and early behavior. This provides 20 covariates. Then, we imputed missing values (to avoid dropping unnecessarily useful observations) and created a dummy variable taking the value 1 when the original baseline variable was imputed (so imputation does not have an impact on the estimation). Finally, we included the square of baseline variables that are not dummies. In total, we included 38 baseline variables in the double-LASSO selection procedure: 20 original covariates, 14 missing variable dummies, and 4 squared covariates.

4.4 Participation in the Treatment

Before examining its effects, we measure students' participation in the program. Table 2 shows large differences in program participation between the treatment and the control groups. 16% of control students declared having participated in the program in Grade 6, 19% in Grade 7, 10% in Grade 8, and 4% in Grade 9. This may be due to errors in program implementation: we found out that the NGO made a few mistakes by looking at the classes whose students all responded that they received the program. There were 17 classes in Grade 6 (3.4% of control classes), 7 in Grade 7 (1.4%), 3 in Grade 8 (0.06%) and only 1 in Grade 9 (0.02%). Since several programs may take place in their school, control students may also be confused and mix up *Energie Jeunes* with other interventions. In the treatment group, between 83-96% of students declared that they had participated in at least one session. They reported having attended 2.5 more sessions than in the control group (3 would be expected under perfect compliance). Thus, compliance to the experimental protocol is satisfactory.

Regarding adherence to the program, the proportion of students in the treatment group who declared that they had made a commitment as part of the program is high, although decreasing over time: 75% in Grade 6, while 54% in Grade 9. The proportion of students who declared that they had honored their commitment is lower and also decreases over time, from 54% in Grade 6 down to 35% in Grade 9. Obviously these statistics should not be taken at face value but they suggest that students largely played along with the *Energie Jeunes* program.

5 Main Treatment Effects

Table 3 shows the impact of the program on final and intermediary outcomes, and Figure 1 provides a graphical presentation.

5.1 Final Outcomes

In Grade 9, we find a 0.05 SD treatment effect on GPA (significant at the 5 percent level), which represents 4% of the Priority Education *versus* national average achievement gap (1.4 SD according to Table A2). We find no treatment effect on the final exam scores except in French (+0.03 SD significant at the 5 percent level) and a positive treatment effect on the educational and professional aspiration scores (+0.04 SD significant at the 10 percent level). The average treatment effects are thus small but still visible thanks to the large sample size. As a robustness check, we verify that our results are not sensitive to attrition. To account for attrition, we use inverse probability weights where the probability to attrit is predicted using a set of LASSO selected baseline variables. The results presented appendix Table A6 are unchanged compared to our main specification without weights.

In previous years, when the program was not yet completed, the impact on GPA was already positive in Grades 6 and 7 (+0.03 SD significant at the 1 and 5 percent level respectively), but not in Grade 8 (0.01 SD). We cannot reject equality between treatment effects in Grade 9 and in Grades 6 and 7. As a complementary analysis, Table A5 compares the treatment effects for students who stayed in the sampled schooled during four years and were thus exposed to the whole program on the one hand (67% of the sample), with students who enrolled in sampled schools in Grades 7, 8 or 9 and were thus exposed to the program for a shorter period of time (33% of the sample) on the other. On average, students who arrive later spent two years in sampled schools, hence half as much as those who registered in Grade 6. However, the results show that the differences in treatment effects between both groups are not statistically significant, so we cannot reject the null that four years of program exposure yields the same impact as two years. The weak dynamic in treatment effects indicates that repeated exposure to the mindset intervention does not amplify the magnitude of the treatment effect proportionally. However, this study does not test whether repeated exposure is necessary to sustain the treatment effect on GPA. Since the effect size remains small despite repeated exposure over four years, the main lesson is that mindset interventions may not have the potential to transform education outcomes if not directly integrated in everyday-life pedagogy through teachers' and parents' practices, as in the case of a teacher training in Turkey studied in Alan et al., 2019.

5.2 Cost-Effectiveness

Is the magnitude of the impact meaningful? As mentioned in the introduction, large-scale education interventions in high-income countries often fail or have fairly small effects. Cheung and Slavin (2016) find average effect sizes on academic achievement of 0.16 SD among 197 Randomized Controlled Trials (RCTs), while Fryer Jr (2017)

finds average effect sizes of 0.05 SD in math and 0.07 SD in reading based on 105 school-based RCTs. However, these average effect sizes mask very different program cost and scope. One of the most consistent findings in the education literature is that impressive effects from small and non-representative samples often fail to replicate when programs are scaled up to larger and more representative populations (Slavin and Smith, 2009). Kraft (2020) provides effect-size benchmarks with a corresponding set of per-pupil cost benchmarks from 747 studies evaluating educational programs offering a variety of sample sizes. Focusing on studies using large samples (above 2,000), this review shows that the effect size of *Energie Jeunes* on GPA (0.05 SD) is at the 60th percentile. But *Energie Jeunes* is much more cost-effective than the typical intervention: while the average cost of programs is $4,752 per pupil, *Energie Jeunes* is only €65 (approximately $75) per pupil, hence more than sixty times cheaper than the average program. The small impact of the program looks thus interesting given its relatively minimal cost. The fact that this experiment was conducted on a large number of students (24,000) and on schools that are fairly representative of the population of priority education schools in France confers more importance and external validity to our results.

5.3 Intermediary Outcomes

Following our pre-registered theory of change presented in Section 2.3, Table 3 reports the treatment effects on students' mindset, self-assessed diligence, and behavior observed by teachers and school registers. These different points of view provide interesting results on the subjective and objective effects of the intervention, which is unique in the literature.

First, we find a 0.11 SD treatment effect on mindset in Grades 7, 8, and 9 (every year when the index is available). The sub-indices analysis in Appendix Table A4 shows that the effect concerns both the growth mindset of intelligence and the internal locus of control. Therefore, the average student is prone to update her perceptions and beliefs regarding her chances of success at school.

Second, students' subjective assessment of their own behavior responds to the intervention in a surprising way: we find no treatment effects on self-assessed diligence in Grade 9, and even a negative effect in Grade 7 (-0.04 SD significant at the 10 percent level). Moreover, the analysis of sub-indices in Table A4 reveals several negative treatment effects on self-assessed grit in Grades 6, 7, and 8, on self-assessed work discipline in Grades 7 and 8, and on self-assessed homework management in Grades 7 and 8. We find no effect on self-declared hours spent on homework. Students' self-assessment thus does not provide evidence of any increase in effort, but rather a slackening of self-discipline. This effect is not consistent neither with the core advice conveyed by the program nor with the positive treatment effects on students' mindset and GPA.

Reassuringly, the objective assessment of students' behavior by teachers and school registers shows a positive reaction to the program: we find a 0.06 SD treatment effects on school-reported behavior at the end of the program (Table 3), mostly driven by a lower probability of being late in class (Appendix Table A4). The effect was not present in previous years, and was even negative in Grade 6, which may suggest some initial reluctance in younger students that dissipated quickly. Besides, we find a positive treatment effect on teacher-reported character—which measures work attitude in class—in Grade 7 (+0.08 SD significant at the 5% level). Sub-indices analysis in Appendix Table A4 shows in particular positive treatment effects on "intellectual character" (zest and curiosity) in Grade 7, on "achievement character" (grit, diligence, schoolwork self-control) in Grades 6 and 7, and on "social character" (gratitude, optimism, social intelligence, interpersonal self-control) in Grade 7 and 8. Taken together, the observations by teachers and in schools registers indicate that the program improved students' behavior in different dimensions at different periods of their life in middle school.

The subjective assessment of students' own behavior thus contradicts the more objective outcomes from teachers' observations, school registers, and GPA. The discrepancy between self-perception and actual outcomes is not uncommon in the literature. It may be explained by well-established social comparison effects, by which people have a lower self-regard when exposed to upward comparisons, involving examples of others who are better off or superior to them—the seminal work in this literature being Festinger, 1954. The videos, included as part of Energie Jeunes' session, show several examples of individuals exhibiting high levels of self-discipline, grit, and perseverance over the course of their life (see the details on videos in Appendix Figures A2). As a consequence, the program may have changed the reference against which students judge their levels of grit and conscientiousness in which case students would compare themselves unfavorably to these examples and become more self-critical on how gritty and self-disciplined they are, leading to a reference bias (Duckworth and David Scott Yeager, 2015). This effect would not be a surprise since a recent meta-analyse of 60+ years of social comparison research shows that people generally choose to compare upward (i.e., to people better than themselves) instead of downward, and that upward comparison most likely results in self-deflating contrast (Gerber et al., 2018).

A remaining question that the next section addresses is whether the program improved the behavior and academic outcomes of certain students while deteriorating self-appreciation of *other* students, or whether these effects worked hand in hand.

6 Heterogeneous Effects and Potential Mechanisms

This section answers three questions: (1) whether we can identify treatment heterogeneity for academic outcomes, (2) if so, who are the students benefiting the most from the program, and (3) whether those who benefited more in terms of academic achievement (GPA, national exam) also benefited more in terms of mindset, objective behavior, and subjective assessment. While patterns of heterogeneity may help sort out the mechanisms driving the results, they remain suggestive as the relationship between intermediary outcomes and final outcomes are only correlational.

6.1 Pre-specified Subgroups

As described in our pre-analysis plan, we expected the treatment effect to vary depending on students' gender, socio-economic status ("aid recipients" *versus* "non-aid recipients"), GPA ('early' academic performance) and behavior ('early' school behavior). We pre-selected 'early' academic performance and early' school behavior because we believed that initially better-behaved/performing students, while needing the program less, would be the ones paying more attention to the program's messages during the *Energie Jeunes* sessions. We also expected girls and students from less deprived socioeconomic status to be more diligent and compliant than boys and more deprived students. We present the treatment effects on Grade 9 outcomes by subgroups in Table 4 and in Figure 2. For each characteristic, we added to the model a dummy indicating the sub-group (F for girls, NR for non-aid recipients, $\mathbb{1}_{g>p50}$ for students whose early GPA is above the median, and $\mathbb{1}_{b>p50}$ for students whose early GPA is above the median) and its interaction with the treatment dummy.[5]

Gender Treatment effects on GPA, final exam scores, and aspirations are statistically similar for boys and girls. However, treatment effects on intermediary outcomes are different: girls were slightly more likely to attend the program and experienced a much stronger change their mindset. In contrast, girls exhibited smaller change in behavior, which could be attributed to the fact that they already demonstrated better behavior than boys as reflected by teacher- and school-reported behavior in the control group. Incidentally, note that girls suffer from a downward bias in self-assessed diligence: in the control group we see that they behave better than boys (+0.13 SD and +0.17 SD in teacher- and school-reported behavior respectively) but have lower self-assessed diligence (-0.12 SD), confirming the presence of reference

[5]We also interact the treatment dummy with the school and cohort fixed effects as well as with every covariates selected via double LASSO. The number of observations decreases when we analyze heterogeneity based on early academic performance and school behavior because the sample is restricted to students enrolled in the school from Grade 6 on, for whom early measures are available.

point biases in subjective assessments.

Social Background Non-aid recipients benefited more from the program than aid recipients. The difference in point estimates on national exam scores is positive and significant at the 1 percent level (+0.07 SD on a basis of -0.02 SD for aid recipients). Likewise, the treatment effect on GPA for non-aid recipients is positive and significant (+0.06 SD, p-value=1.4%) while smaller and non-significant for aid recipients—but the difference itself is not statistically significant. Turning to intermediary outcomes, the program improved students' behavior similarly in both sub-groups but non-aid recipients were 7 pp more likely to participate to the program than aid recipients and their mindset improved much more (+0.17 *versus* +0.04 SD). Overall, less deprived students benefited more than the more deprived, potentially related to more responsive mindsets.

Early Academic Performance We do not find much treatment effect heterogeneity depending on baseline academic performance, with the exception of positive treatment effects on aspirations and school-reported behavior only for low-achievers—although the differential treatment effect on aspiration is not precisely estimated. High-achievers were also 6 pp more likely to participate to the program than low-achievers. The null effect on school-reported behavior for high-achievers may be related to strongly better baseline behavior absent the program, as reflected by teacher- and school-reported behavior in the control group.

Early Behavior Well-behaved students benefited more from the program than poorly-behaved students: the differences in point estimates on GPA and on national exam scores are positive and significant. We also find differential impacts on intermediary outcomes: well-behaved students are slightly more likely to attend the program than poorly-behaved students (+ 3 pp), changes in mindset and in self-assessed diligence are larger (respectively +0.06 SD—imprecisely estimated—and +0.11 SD), while on the contrary the change in school-reported behavior is significantly smaller (-0.06 SD). Well-behaved students thus benefited more in terms of mindset and final outcomes, but less in terms of discipline which, as for girls and high-achievers, may be related to the fact that, absent the program, this subgroup already behaves better than the poorly-behaved.

6.2 Machine Learning Approach

The analysis of pre-specified subgroups suggests that stronger profiles tended to gain more in terms of academic outcomes and mindset, but less in terms of discipline, than weaker profiles. However, given the number of hypotheses tested, this analysis may be subject to spurious findings while at the same time heterogeneity may be

driven by the interaction of more than one characteristic at a time. Therefore, the heterogeneity analysis presented above may lead to false positives while missing substantive variation in treatment effects. To estimate treatment heterogeneity more flexibly and tightly, we thus turn to a data-driven approach where the subgroups are defined using a machine learning algorithm and the high-dimensional baseline dataset.

We implement the Generalized Random Forest method proposed by Athey and Wager (2019) and Wager and Athey (2018) to estimate predicted CATEs for each observation with honest causal forest. We apply the approach to two final outcomes: Grade 9 GPA and national exam scores. To fit the CATE, we use all baseline covariates, their square, cube, and interactions (218 variables in total) to form 2000 trees that maximize treatment effect heterogeneity.[6] To rank observations, we proceeded as follows. First, we randomly assigned each observation to a fold from 1 to 50 and then we fitted a causal forest. When fitting and predicting the forest, the procedure ensures that CATE estimates for observations in fold i are computed using trees that do not use any observations of that same fold during fitting, which protect against over-fitting. Then, within each fold, observations were assigned to the strongly or the weakly affected group if their predicted CATE was above or below the median CATE prediction within that fold. Ranking only observations within a fold ensures that the relative rank between two observations does not use data from either observation. In what follows, the group of students predicted to be *weakly affected* are referred to as *Q1* while the group of students predicted to be *strongly affected* are referred to as *Q2*.

Table 5 compares the average treatment effects in *Q2* compared to *Q1* using the same specification as for pre-specified subgroups. On the left part of the table, we use GPA as the targeted outcome. We find strong treatment effect heterogeneity: +0.09 SD in *Q2* compared to +0.02 SD in *Q1*, the difference being significant at the 1 percent level. We also see that the stronger effect on GPA for *Q2* is concomitant with a differential effect on national exam scores (+0.05 SD significant at the 5 percent level), showing a positive correlation between treatment effects on the two academic outcomes. When conducting the same exercise using national exam scores as the targeted outcome to estimate the CATEs (right part of Table 5), we also find heterogeneity between *Q2* and *Q1* with a differential effect of +0.05 SD significant at the 5 percent level. Finally, Table 6 compares the baseline characteristics of the more *versus* less affected students: for both targeted outcomes, *strongly affected* students were slightly older, more often female, they came for relatively less deprived families, and their academic and discipline profiles in the first trimester in Grade 6 were relatively better. The machine learning results thus confirm the pre-registered subgroups analysis while adding clarity and tightness: a subgroup of

[6]The CATE function is fitted using the causal forest function from the R package *grf*

students effectively gained more in academic outcomes than the others, and these students had better social, academic, and discipline profiles.

6.3 Discussion of Potential Mechanisms

Although patterns of heterogeneity are correlational, analyzing them across intermediary outcomes in *Q2* and *Q1* may help sort out the mechanisms driving the results (Davis and Heller, 2020). As there is no pattern of heterogeneity across intermediary outcomes when the national exam is used as the targeted outcome, in what follows we use only GPA as the targeted outcome to define *Q2* and *Q1*. We thus examine if those who benefited more in terms of GPA—who gained more in national exam scores too—also benefited more in terms of mindset, objective behavior, and subjective assessment. The idea is that if benefits in GPA are concentrated among subgroups who did not benefit in a given intermediary outcome, at least that helps rule out that specific mechanism.[7]

We find that treatment effects are significantly larger in *Q2* compared to *Q1* on attendance (+4 pp) and on mindset (+0.07 SD) but, on the contrary, significantly smaller on self-assessed diligence (-0.18 SD) and on teacher-reported character (-0.10 SD), and similar on school-reported behavior (no difference in treatment effects). A simple comparison of the local average treatment effects shows that the heterogeneity in treatment effects on GPA is not driven by differential take-up.[8] If we consider treatment effects on *Q1* and *Q2* separately, we observe that the positive effects on academic outcomes in *Q2* came with a strong improvement in mindset, a moderate improvement in absences, lateness, and misconduct observed in school registers, no change in work attitude in class observed by teachers, and a *negative* change in self-assessed diligence (ATE=-0.08 SD, p-value=0.04). In contrast, the null effect on academic outcomes in *Q1* came with moderate improvements in mindset, in absences, lateness, and misconduct observed in school registers, and in work attitude in class observed by teachers (+ 0.07 SD although statistically not significant), and a *positive* change in self-assessed diligence (ATE=-0.11 SD, p-value<0.01).

[7] An alternative way of exploring the mechanisms is to decompose the average treatment effect into experimentally induced changes in intermediary outcomes and in other unmeasured factors (J. Heckman et al., 2013). This decomposition method relies on two conditions: 1) observed mediating factors are independent from unobserved mediating factors, and 2) the treatment does not modify the way mediating factors affect the final outcome (structural invariance assumption). The first condition is not testable, but the second condition can be tested: we regress GPA on intermediary outcomes (potential mediating factors), the treatment, interactions of each intermediary outcome with the treatment, and baseline covariates selected with a double-LASSO. We find significant coefficients on interaction terms, which violates the structural invariance assumption (coefficients on intermediary outcomes vary with treatment assignment). In this case, multiplying the coefficient on an intermediary outcomes by the treatment effect on this factor does not provide a valid estimate of the contribution of an intermediary outcome to the average treatment effect.

[8] Dividing the Intention-to-Treat effects by the proportion of students who attended at least one session, the LATE is $(0.02)/(0.76) = 0.023$ SD for *Q1*, while $(0.02 + 0.07)/(0.76 + 0.04) = 0.112$ SD for *Q2*. LATEs are thus even more differential than ATEs.

The machine learning approach thus clarifies that stronger gains in GPA came together with stronger gains in mindset but not in objective measures of behavior, at least not the ones we collected. This is not so surprising since *strongly affected* students already behave better than *weakly affected* students, but it implies that our measures of behavior probably missed improvements in more subtle aspects of behavior for *strongly affected* students other than absences, lateness, misconduct, and work attitude in class. For future research, other measurement tools may be more appropriate to capture behavioral gains for well-behaved students. Moreover, the machine learning approach uncovers an interesting finding: higher academic gains were associated with a reduction in self-assessed discipline. This may due to the fact that *strongly affected* students are more prone to self-deflating contrast effect than *weakly affected* students when the program exposes them to upward social comparisons. This finding shows that the negative treatment effect on self-assessment should not be taken at face value because it was associated with better final outcomes, which casts doubt on the reliability of students' subjective perceptions. In any case, it is interesting that the negative effect on self-appreciation worked hand-in-hand with academic gains since it raises the question of whether self-criticism was useful to spur greater changes and impacts.

7 Conclusion

This paper shows that a large-scale mindset intervention in disadvantaged schools was efficient at developing more optimistic beliefs on one's chances of success and at improving school-related behavior, with a downstream positive effect on GPA. It also shows important heterogeneity in the mindset intervention treatment effect on GPA, with much stronger effects for students who have better social, academic, and discipline profiles. The mindset of these students responded more to the program and they also became more self-critical on their own behavior, which may explain why they benefited more. In contrast, weaker profiles, who gained less academically, gained more in terms of behavior and discipline. A future follow-up of the students would thus be interesting to assess whether these behavioral gains translated into better academic outcomes after middle school.

More generally, this paper shows that mindset interventions are a cost-effective policy for disadvantaged schools. Compared to other educational programs, the *Energie Jeunes* program's return is indeed high. Compensatory education policies offering more hours of teaching or after-class tutoring or reduced class size often fail to substantially increase academic achievement despite large costs (Beffy and Davezies, 2013; Bénabou et al., 2009; Bressoux et al., 2016; Goux et al., 2017). These policies focus on external parameters of the education production function, while it may be crucial to also address internal constraints that impede students'

motivation and ambition. This paper sheds light on this matter by pointing to the important role of mindset in education, i.e., how adolescents think about themselves and their chances of success. However, effect sizes remain small, despite the fact that the message is repeated during four years. Therefore, multiplying low-cost mindset interventions led by external actors may not yield substantive impacts. Rather, training teachers and parents who are in daily contact with students may be necessary to transform educational outcomes more substantively, as shown in the case of a teacher training in Alan et al., 2019.

References

Alan, Sule, Teodora Boneva, and Seda Ertac (2019). "Ever Failed, Try Again, Succeed Better: Results from a Randomized Educational Intervention on Grit". In: *The Quarterly Journal of Economics* 134.3, pp. 1121–1162.

Algan, Yann, Elise Huillery, and Corinne Prost (2018). "Trust, Cooperation and Autonomy:Towards a 21st Century School". In: *Note du Conseil d'Analyse Economique* 48.

Almlund, Mathilde, Angela L Duckworth, James J Heckman, and Tim Kautz (2011). "Personality Psychology and Economics". In: *Handbook of Economics of Education*. Vol. 4. Elsevier. Chap. Chapter 1, pp. 1–181.

Anderson, Michael L (2008). "Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects". In: *Journal of the American statistical Association* 103.484, pp. 1481–1495.

Athey, Susan and Stefan Wager (2019). "Estimating treatment effects with causal forests: An application". In: *Observational studies* 5.2, pp. 37–51.

Beffy, Magali and Laurent Davezies (2013). "Has the "Ambition Success Networks" Educational Program Achieved its Ambition?" In: *Annals of Economics and Statistics/Annales d'Economie et de Statistique*, pp. 271–293.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen (2013). "Inference on Treatment Effects after Selection among High-Dimensional Controls". In: *The Review of Economic Studies* 81.2, pp. 608–650.

Bénabou, Roland, Francis Kramarz, and Corinne Prost (2009). "The French zones d'éducation prioritaire: Much ado about nothing?" In: *Economics of Education Review* 28.3, pp. 345–356.

Bettinger, Eric, Sten Ludvigsen, Mari Rege, Ingeborg F. Solli, and David Yeager (2018). "Mind-set interventions are a scalable treatment for academic underachievement". In: *Journal of Economic Behavior & Organization* 146, pp. 1–15.

Boon-Falleur, Mélusine, Adrien Bouguen, Axelle Charpentier, Yann Algan, Élise Huillery, and Coralie Chevallier (2022). "Simple questionnaires outperform behavioral tasks to measure socio-emotional skills in students". In: *Scientific reports* 12.1, p. 442.

Boulay, Beth, Barbara Goodson, Rob Olsen, Rachel McCormick, Catherine Darrow, Michael Frye, Katherine Gan, Eleanor Harvill, and Maureen Sarna (2018). "The Investing in Innovation Fund: Summary of 67 Evaluations. Final Report. NCEE 2018-4013." In: *National Center for Education Evaluation and Regional Assistance*.

Bressoux, Pascal, Marc Gurgand, Nina Guyon, Marion Monnet, and Julie Pernaudet (2016). "Évaluation des Programmes de réussite éducative". In: *Paris: Institut des Politiques Publiques [en ligne]*.

Castillo, Marco, Paul J. Ferraro, Jeffrey L. Jordan, and Ragan Petrie (2011). "The today and tomorrow of kids: Time preferences and educational outcomes of children". In: *Journal of Public Economics* 95.11, pp. 1377–1385.

Cheung, Alan CK and Robert E Slavin (2016). "How methodological features affect effect sizes in education". In: *Educational Researcher* 45.5, pp. 283–292.

Claro, Susana, David Paunesku, and Carol S Dweck (2016). "Growth mindset tempers the effects of poverty on academic achievement". In: *Proceedings of the National Academy of Sciences* 113.31, pp. 8664–8668.

Davis, Jonathan MV and Sara B Heller (2020). "Rethinking the benefits of youth employment programs: The heterogeneous effects of summer jobs". In: *Review of economics and statistics* 102.4, pp. 664–677.

Dohmen, Thomas, Armin Falk, David Huffman, Uwe Sunde, Jürgen Schupp, and Gert G. Wagner (2011). "Individual Risk Attitudes: Measurement, Determinants, and Behavioral Consequences". In: *Journal of the European Economic Association* 9.3, pp. 522–550.

Duckworth, Angela L and Patrick D Quinn (2009). "Development and Validation of the Short Grit Scale (Grit–S)". In: *Journal of Personality Assessment* 91.2. PMID: 19205937, pp. 166–174.

Duckworth, Angela L, Abigail Quirk, Robert Gallop, Rick H Hoyle, Dennis R Kelly, and Michael D Matthews (2019). "Cognitive and noncognitive predictors of success". In: *Proceedings of the National Academy of Sciences* 116.47, pp. 23499–23504.

Duckworth, Angela L and Martin E P Seligman (2005). "Self-Discipline Outdoes IQ in Predicting Academic Performance of Adolescents". In: *Psychological Science* 16.12, pp. 939–944.

Duckworth, Angela L and David Scott Yeager (2015). "Measurement matters: Assessing personal qualities other than cognitive ability for educational purposes". In: *Educational researcher* 44.4, pp. 237–251.

Dweck, Carol S and David S Yeager (2019). "Mindsets: A view from two eras". In: *Perspectives on Psychological Science* 14.3, pp. 481–496.

Festinger, Leon (1954). "A theory of social comparison processes". In: *Human relations* 7.2, pp. 117–140.

Fryer Jr, Roland G (2017). "The production of human capital in developed countries: Evidence from 196 randomized field experiments". In: *Handbook of economic field experiments*. Vol. 2. Elsevier, pp. 95–322.

Galla, Brian M, Benjamin D Plummer, Rachel E White, David Meketon, Sidney K D'Mello, and Angela L Duckworth (2014). "The Academic Diligence Task (ADT): Assessing individual differences in effort on tedious but important schoolwork". In: *Contemporary educational psychology* 39.4, pp. 314–325.

Gerber, Jonathan P, Ladd Wheeler, and Jerry Suls (2018). "A social comparison theory meta-analysis 60+ years on." In: *Psychological bulletin* 144.2, p. 177.

Goldberg, Lewis R (1990). "An alternative "description of personality": the big-five factor structure." In: *Journal of personality and social psychology* 59.6, p. 1216.

Goldberg, Lewis R, John A Johnson, Herbert W Eber, Robert Hogan, Michael C Ashton, C Robert Cloninger, and Harrison G Gough (2006). "The international personality item pool and the future of public-domain personality measures". In: *Journal of Research in personality* 40.1, pp. 84–96.

Golsteyn, Bart HH, Hans Grönqvist, and Lena Lindahl (2014). "Adolescent time preferences predict lifetime outcomes". In: *The Economic Journal* 124.580, F739–F761.

Goux, Dominique, Marc Gurgand, and Eric Maurin (2017). "Reading enjoyment and reading skills: Lessons from an experiment with first grade children". In: *Labour Economics* 45. Field experiments in labor economics and social policies, pp. 17–25.

Guyon, Nina and Elise Huillery (2021). "Biased aspirations and social inequality at school: Evidence from french teenagers". In: *The Economic Journal* 131.634, pp. 745–796.

Heckman, James J, Jora Stixrud, and Sergio Urzua (2006). "The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior". In: *Journal of Labor economics* 24.3, pp. 411–482.

Heckman, James, Rodrigo Pinto, and Peter Savelyev (2013). "Understanding the mechanisms through which an influential early childhood program boosted adult outcomes". In: *American Economic Review* 103.6, pp. 2052–86.

Kraft, Matthew A (2020). "Interpreting effect sizes of education interventions". In: *Educational Researcher* 49.4, pp. 241–253.

Li, Yue and Timothy C Bates (2017). "Does mindset affect children's ability, school achievement, or response to challenge? Three failures to replicate". In: *Unpublished manuscript.[Available at: https://osf. io/preprints/socarxiv/tsdwy/download]*.

Macnamara, Brooke N and Alexander P Burgoyne (2022). "Do growth mindset interventions impact students' academic achievement? A systematic review and meta-analysis with recommendations for best practices." In: *Psychological Bulletin*.

Moffitt, Terrie E, Louise Arseneault, Daniel Belsky, Nigel Dickson, Robert J Hancox, HonaLee Harrington, Renate Houts, Richie Poulton, Brent W Roberts, Stephen Ross, et al. (2011). "A gradient of childhood self-control predicts health, wealth, and public safety". In: *Proceedings of the National Academy of Sciences* 108.7, pp. 2693–2698.

Outes-Leon, Ingo, Alan Sánchez, and Renos Vakis (2020). *The Power of Believing You Can Get Smarter: The Impact of a Growth-Mindset Intervention on Academic Achievement in Peru*.

Park, Daeun, Eli Tsukayama, Geoffrey P Goodwin, Sarah Patrick, and Angela L Duckworth (2017). "A tripartite taxonomy of character: Evidence for intrapersonal, interpersonal, and intellectual competencies in children". In: *Contemporary educational psychology* 48, pp. 16–27.

Paunesku, David, Gregory M Walton, Carissa Romero, Eric N Smith, David S Yeager, and Carol S Dweck (2015). "Mind-set interventions are a scalable treatment for academic underachievement". In: *Psychological science* 26.6, pp. 784–793.

Slavin, Robert E and Dewi Smith (2009). "The relationship between sample sizes and effect sizes in systematic reviews in education". In: *Educational evaluation and policy analysis* 31.4, pp. 500–506.

Sutter, Matthias, Martin G Kocher, Daniela Glätzle-Rützler, and Stefan T Trautmann (2013). "Impatience and uncertainty: Experimental decisions predict adolescents' field behavior". In: *American Economic Review* 103.1, pp. 510–31.

Tsukayama, Eli, Angela L Duckworth, and Betty Kim (2013). "Domain-specific impulsivity in school-age children". In: *Developmental Science* 16.6, pp. 879–893.

Wager, Stefan and Susan Athey (2018). "Estimation and inference of heterogeneous treatment effects using random forests". In: *Journal of the American Statistical Association* 113.523, pp. 1228–1242.

Walton, Gregory M (2014). "The new science of wise psychological interventions". In: *Current Directions in Psychological Science* 23.1, pp. 73–82.

Xu, Jianzhong and Hongyun Wu (2013). "Self-regulation of homework behavior: Homework management at the secondary school level". In: *The Journal of Educational Research* 106.1, pp. 1–13.

Yeager, David S and Carol S Dweck (2020). "What can be learned from growth mindset controversies?" In: *American psychologist* 75.9, p. 1269.

Yeager, David S, Paul Hanselman, Gregory M Walton, Jared S Murray, Robert Crosnoe, Chandra Muller, Elizabeth Tipton, Barbara Schneider, Chris S Hulleman, Cintia P Hinojosa, et al. (2019). "A national experiment reveals where a growth mindset improves achievement". In: *Nature* 573.7774, pp. 364–369.

Figures

Figure 1: Final Impacts on Final and Intermediary Outcomes

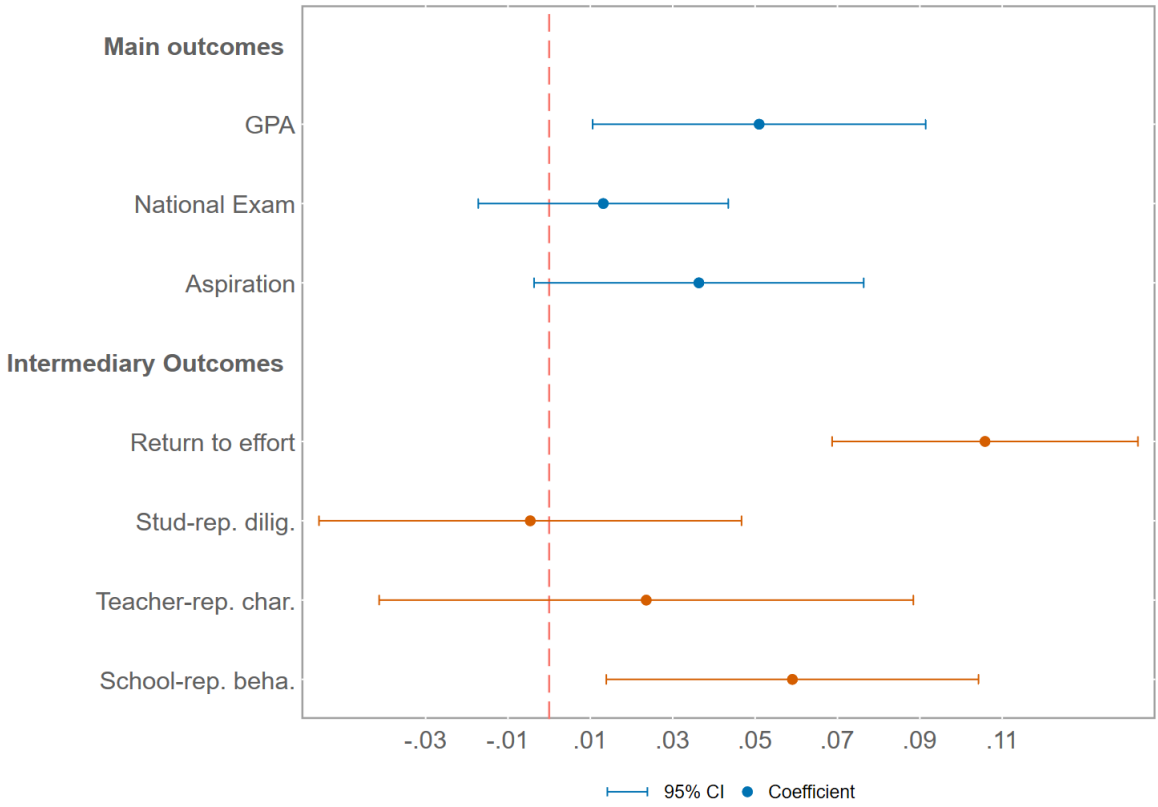
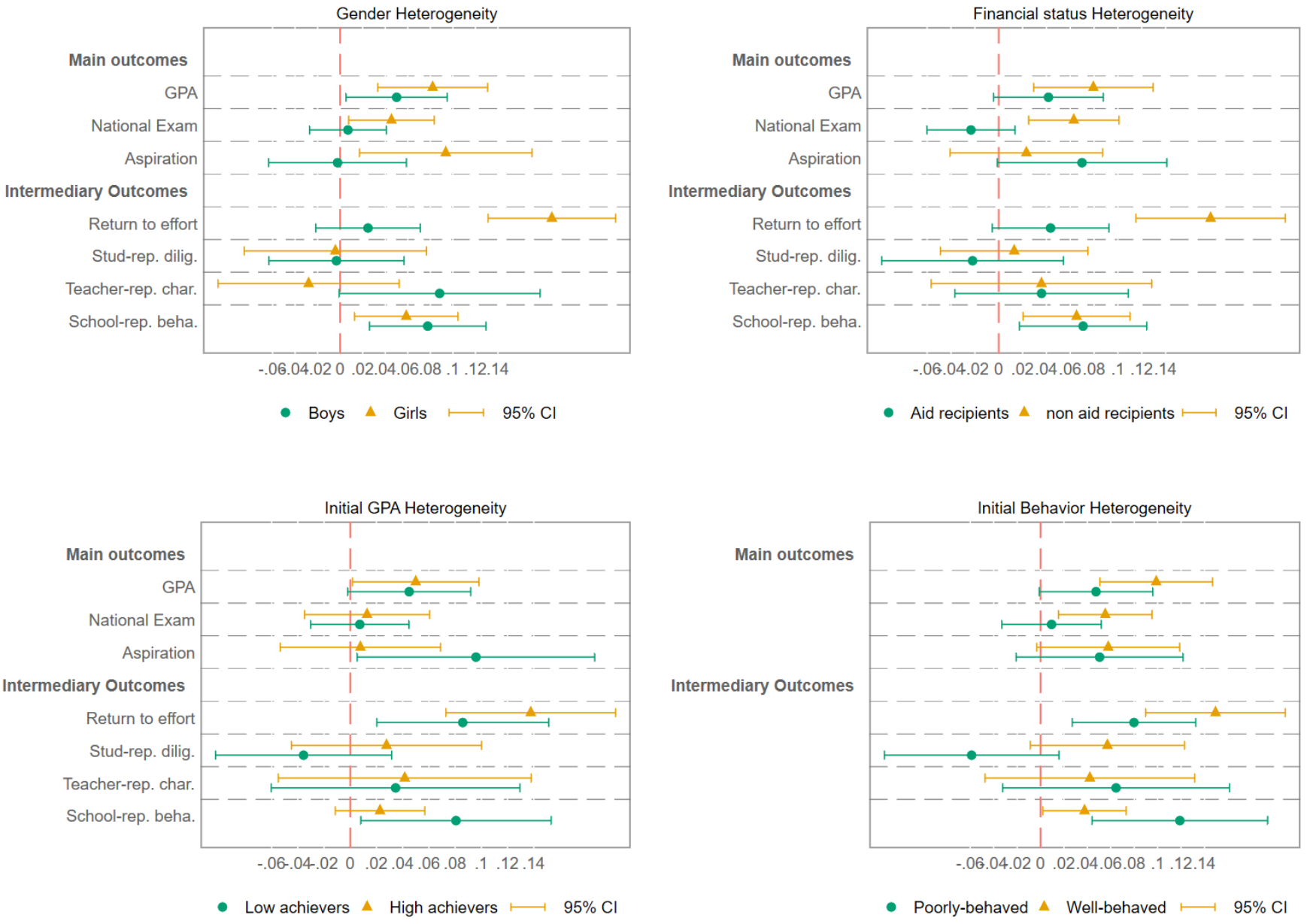


Figure 2: Heterogeneity in Final and Intermediary Outcomes



Tables

Table 1: Samples and Attrition

	Administrative data					Student Survey Data				Teacher Survey Data			
	G6	**G7**	**G8**	**G9**	**Nat. exam**	**G6**	**G7**	**G8**	**G9**	**G6**	**G7**	**G8**	**G9**
Actual													
Full	24,142	23,095	23,751	23,588	19,709	5,836	6,573	6,002	5,774	4,699	4,832	4,602	4,503
Control	11,914	11,330	11,817	11,645	9,934	2,868	3,215	2,973	2,805	2,868	3,215	2,973	2,805
Treatment	12,228	11,765	11,934	11,943	9,775	2,968	3,358	3,029	2,969	2,968	3,358	3,029	2,969
Expected													
Full	24,142	23,095	24,349	24,708	23,588	5,130	6,965	7,070	7,231	5,130	6,965	7,070	7,231
Control	11,914	11,330	12,070	12,079	11,645	2,525	3,437	3,514	3,570	2,525	3,437	3,514	3,570
Treatment	12,228	11,765	12,279	12,629	11,943	2,605	3,528	3,556	3,661	2,605	3,528	3,556	3,661
Attrition													
Control	0.00	0.00	0.02	0.04	0.15	0.05	0.07	0.16	0.22	0.22	0.32	0.35	0.42
T-C	0.00	0.00	0.01	0.02	0.03	-0.00	-0.01	-0.00	-0.03	0.01	-0.02	-0.00	-0.07
	(0.00)	(0.00)	(0.02)	(0.03)	(0.03)	(0.02)	(0.03)	(0.05)	(0.05)	(0.04)	(0.05)	(0.06)	(0.05)
Entry rate													
Control	.	0.070	0.165	0.218	.	.	.	.	.	.	.	.	.
T-C	.	-0.002	-0.004	0.004	.	.	.	.	.	.	.	.	.
		(0.005)	(0.009)	(0.011)	.								
Exit rate													
Control	.	0.118	0.151	0.198	.	.	.	.	.	.	.	.	.
T-C	.	-0.011	0.000	-0.011	.	.	.	.	.	.	.	.	.
		(0.011)	0.000	(0.011)									

This table shows sample sizes and attrition rates for the different data sources (administrative data, student survey and teacher survey). We first provide the number of observations in the full sample, control and treatment group (*Actual*), then the number of observations that we expected if no students attrited (*Expected*), and the comparison between the actual and the expected sample (*Attrition*) and its differential (*Differential*). We then provide the exist and entry rate each year and the treatment control differential. Regressions are controlled for school and cohort fixed effect, standard errors are robust and clustered at school*cohort level. * 10%, ** 5%, *** 1% significance level

Table 2: Participation to the EJ program

	Grade 6			Grade 7			Grade 8			Grade 9		
	Obs.	C	Impact	Obs.	C	Impact	Obs.	C	Impact	Obs.	C	Impact
Participated at least once	5,447	0.16	0.80***	5,998	0.19	0.73***	5,485	0.10	0.81***	5,490	0.04	0.79***
			(0.01)			(0.01)			(0.02)			(0.02)
# of sessions attended	5,244	0.21	2.46***	5,698	0.21	2.42***	5,334	0.11	2.43***	5,490	0.07	2.26***
			(0.03)			(0.03)			(0.04)			(0.07)
Made commitment	5,446	0.09	0.66***	5,998	0.11	0.59***	5,485	0.04	0.59***	5,490	0.02	0.52***
			(0.01)			(0.01)			(0.01)			(0.02)
Honored commitment	5,447	0.08	0.46***	5,998	0.09	0.44***	5,485	0.03	0.40***	5,490	0.02	0.34***
			(0.01)			(0.01)			(0.01)			(0.01)
Student sample size	24,142	11,914		23,095	11,330		23,751	11,817		23,588	11,645	
Number of clusters	194	97		194	97		190	95		186	94	

This table shows measures of students' participation and engagement in the *Energie Jeunes* program from Grade 6 to Grade 9. Participation measures are presented in rows. Column *Obs.* gives the number of observations, column *C* the mean of the variable in the control group and *Impact* the coefficient from the regression of the outcome on the treatment dummy controlling for school and cohort fixed effects. Standard errors are clustered at the school*cohort level and robust to heteroscedasticity. *10%, **5%, ***1% significance level

Table 3: Final Impacts and Intermediate Dynamics

	Final impact		**Dynamics**						**P-value**		
	Grade 9		**Grade 6**		**Grade 7**		**Grade 8**		**G6=G9**	**G7=G9**	**G8=G9**
	Obs.	Impact	Obs.	Impact	Obs.	Impact	Obs.	Impact			
Final outcomes											
GPA	19,330	0.05**	20,783	0.03***	21,443	0.03**	19,713	0.01	0.45	0.31	0.08
		(0.02)		(0.01)		(0.01)		(0.02)			
National exam	19,702	0.01		.		.		.	.	.	.
		(0.02)		.		.		.			
...French	19,660	0.03**		.		.		.	.	.	.
		(0.01)		.		.		.			
...Math	19,625	-0.01		.		.		.	.	.	.
		(0.02)		.		.		.			
Aspiration	5,497	0.04*		.		.		.	.	.	.
		(0.02)		.		.		.			
Intermediary outcomes											
Student mindset	5,485	0.11***		.	6,027	0.11***	5,496	0.11***	.	0.95	0.77
		(0.02)		.		(0.02)		(0.02)			
Self-assessed diligence	5,497	0.00	5,506	0.00	6,458	-0.05***	5,706	-0.01	0.66	0.15	0.83
		(0.03)		(0.02)		(0.02)		(0.02)			
Teacher-reported character	4,503	0.02	4,494	0.04	4,826	0.07**	4,596	0.06*	0.58	0.27	0.47
		(0.03)		(0.03)		(0.03)		(0.03)			
School-reported behavior	22,305	0.06***	22,074	-0.04***	22,449	0.01	22,445	0.01	0.00	0.10	0.13
		(0.02)		(0.01)		(0.02)		(0.02)			

This table presents the standardized impacts of the treatment in Grade 9 on our summary indices. Indices are presented in rows. Columns *Obs.* gives the number of observations, columns *Impact* the coefficients from the regressions of the outcomes on the treatment dummy. Regressions are controlled for school and cohort fixed effects as well as double-LASSO selected covariates. Standard errors in parenthesis, robust to heteroscedasticity and clustered at the school*cohort level. *10%, **5%, ***1% significance level

Table 4: Heterogeneous Treatment Effects - Pre-defined Subgroups

	Gender			**Non-aid recipients**			**Early GPA**			**Early behavior**		
	EJ	F	EJ*F	EJ	NR	EJ*NR	EJ	$\mathbb{1}_{g>p50}$	EJ* $\mathbb{1}_{g>p50}$	EJ	$\mathbb{1}_{b>p50}$	EJ*$\mathbb{1}_{b>p50}$
Participation												
at least once	0.77***	0.00	0.04***	0.76***	0.00	0.07***	0.77***	-0.01	0.06***	0.78***	0.01	0.03*
	(0.02)	(0.01)	(0.01)	(0.03)	(0.01)	(0.02)	(0.03)	(0.01)	(0.02)	(0.03)	(0.01)	(0.02)
sessions attended	2.21***	-0.03	0.10***	2.17***	0.00	0.19***	2.22***	-0.03	0.17***	2.20***	0.01	0.13***
	(0.07)	(0.02)	(0.04)	(0.08)	(0.02)	(0.05)	(0.08)	(0.02)	(0.05)	(0.08)	(0.02)	(0.05)
Final outcomes												
GPA	0.05**	0.19***	0.01	0.03	-0.05**	0.03	0.05**	0.97***	0.00	0.03	0.14***	0.05**
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
National exam	0.01	0.18***	0.01	-0.02	0.02	0.07***	0.00	0.99***	0.01	-0.01	0.12***	0.04*
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)
...French	0.03*	0.35***	0.01	0.02	0.02	0.01	0.02	0.83***	0.00	0.02	0.07***	0.01
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)
...Math	-0.01	-0.07***	0.00	-0.05***	0.00	0.08***	0.00	0.92***	-0.01	-0.03	0.15***	0.04
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)
Aspiration	0.00	0.12***	0.07	0.06*	0.08*	-0.04	0.09*	0.71***	-0.09	0.04	0.05	-0.01
	(0.03)	(0.04)	(0.04)	(0.03)	(0.04)	(0.04)	(0.04)	(0.05)	(0.06)	(0.03)	(0.04)	(0.04)
Intermediary outcomes												
Student mindset	0.03	-0.04	0.16***	0.04*	-0.05	0.14***	0.09***	0.25***	0.05	0.08***	0.00	0.06
	(0.02)	(0.04)	(0.03)	(0.02)	(0.05)	(0.04)	(0.03)	(0.04)	(0.05)	(0.03)	(0.04)	(0.04)
Self-assessed dilig.	0.00	-0.12**	-0.01	-0.03	-0.01	0.04	-0.03	0.13***	0.05	-0.06	0.10**	0.11***
	(0.03)	(0.05)	(0.05)	(0.04)	(0.05)	(0.05)	(0.03)	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)
Teacher-rep. char.	0.10**	0.17***	-0.14***	0.02	0.10**	0.03	0.04	0.52***	-0.01	0.06	0.29***	-0.05
	(0.04)	(0.05)	(0.05)	(0.04)	(0.04)	(0.05)	(0.05)	(0.05)	(0.06)	(0.05)	(0.05)	(0.06)
School-rep. beh.	0.08***	0.13***	-0.03*	0.06**	0.04**	-0.01	0.09**	0.36***	-0.07**	0.09**	0.51***	-0.06**
	(0.03)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.04)	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)

This table presents the heterogeneous impacts on final and intermediary outcomes (in SD) by gender (F), financial aid status (NR), early GPA ($\mathbb{1}_{g>p50}$) and early behavior ($\mathbb{1}_{b>p50}$). F is equal to 1 if the student is a girl, NR if non-aid recipient, $\mathbb{1}_{g>p50}$ if early GPA is above the median, and $\mathbb{1}_{b>p50}$ if early behavior is above the median, 0 otherwise. Columns *EJ* shows the treatment effect for students whose subgroup dummy is equal to 0. Columns EJ*F, EJ*NR, EJ*$\mathbb{1}_{g>p50}$, EJ*$\mathbb{1}_{b>p50}$ shows the additional treatment effect for students whose subgroup dummy is equal to 1. Regressions are controlled for school and cohort fixed effects, a set of double-LASSO selected covariates, and their interactions with the heterogeneity variables. Standard errors in parentheses are robust to heteroscedasticity and clustered at the school*cohort level. *10%, **5%, ***1% significance level

Table 5: Heterogeneous Treatment Effects - Generalized Random Forest

	CATE ranking - GPA				CATE ranking Nat. Exam			
	Obs	EJ	Q2	EJ*Q2	Obs	EJ	Q2	EJ*Q2
Participation								
at least once	5,490	0.76***	0.01*	0.04**	5,490	0.79***	0.01	0.01
		(0.03)	(0.01)	(0.02)		(0.02)	(0.01)	(0.02)
sessions attended	5,490	2.21***	0.03*	0.04	5,490	2.26***	0.01	0.03
		(0.09)	(0.02)	(0.05)		(0.07)	(0.02)	(0.05)
Final outcomes								
GPA	19,846	0.02	0.17***	0.07***	19,846	0.03	0.23***	0.03
		(0.02)	(0.03)	(0.02)		(0.03)	(0.02)	(0.03)
National exam	19,709	-0.01	0.22***	0.05**	19,709	-0.01	0.24***	0.05**
		(0.02)	(0.03)	(0.02)		(0.02)	(0.02)	(0.02)
...French	19,666	0.00	0.19***	0.05**	19,666	0.01	0.19***	0.04**
		(0.02)	(0.02)	(0.02)		(0.02)	(0.02)	(0.02)
...Math	19,632	-0.03	0.18***	0.05**	19,632	-0.05**	0.23***	0.07***
		(0.02)	(0.03)	(0.02)		(0.02)	(0.02)	(0.02)
Aspiration	5,504	0.04	0.18***	-0.02	5,504	-0.04	0.18***	0.07
		(0.03)	(0.05)	(0.05)		(0.04)	(0.05)	(0.05)
Intermediary outcomes								
Student mindset	5,492	0.08***	0.04	0.07*	5,492	0.08***	0.05	0.01
		(0.03)	(0.03)	(0.04)		(0.03)	(0.04)	(0.05)
Self-assessed diligence	5,504	0.11***	0.05	-0.18***	5,504	0.00	0.02	-0.05
		(0.04)	(0.05)	(0.05)		(0.04)	(0.04)	(0.05)
Teacher-reported character	4,503	0.07	0.26***	-0.10**	4,503	-0.03	0.23***	0.02
		(0.05)	(0.05)	(0.05)		(0.05)	(0.05)	(0.05)
School-reported behavior	23,158	0.06*	0.13***	0.00	23,158	0.04	0.14***	-0.02
		(0.03)	(0.02)	(0.02)		(0.03)	(0.02)	(0.02)

This table shows the heterogeneous treatment effects based on the Generalized Causal Forest method. Columns *CATE ranking - GPA* show treatment heterogeneity using Grade 9 GPA as the targeted outcome. Columns *CATE ranking - Nat.Exam* show treatment heterogeneity using the national exam score as the targeted outcome. Column EJ shows the treatment effect for students whose CATE is below the median, Q2 the difference on outcome between students below and above the median CATE, and EJ*Q2 the additional treatment effect for students whose CATE is above the median. All regressions are estimated using cohort and school fixed effect as well as a set of double-LASSO selected covariates. Standard errors are clustered at the school*cohort level. *10%, **5%, ***1% significance level.

Table 6: Balancing using the Causal Forest Median Indicator

	G9 GPA		**Nat. Exam**	
	Q1 mean	Q2	Q1 mean	Q2
Date of birth	2003.30	0.17***	2003.35	0.10***
		(0.01)		(0.01)
1=girls	0.49	0.04***	0.50	0.04***
		(0.01)		(0.01)
Financial aid (FA)	0.49	-0.09***	0.49	-0.12***
		(0.01)		(0.01)
FA level, max=4	1.02	-0.23***	1.02	-0.21***
		(0.02)		(0.02)
Blue collar parents	0.76	-0.03***	0.75	-0.15***
		(0.01)		(0.01)
active parents	0.47	0.07***	0.47	0.14***
		(0.01)		(0.01)
single parent	0.18	-0.07***	0.18	-0.14***
		(0.01)		(0.01)
# of children	2.45	-0.24***	2.44	0.33***
		(0.03)		(0.03)
Foreigner	0.15	-0.03***	0.14	0.04***
		(0.01)		(0.01)
Was held back	0.15	-0.16***	0.14	-0.14***
		(0.01)		(0.01)
repeated the grade	0.04	-0.02***	0.03	-0.02***
		(0.00)		(0.00)
Early GPA	5.41	0.28***	5.45	0.40***
		(0.02)		(0.02)
Early Behavior score	0.01	0.34***	0.04	0.21***
		(0.03)		(0.02)

This table shows the results of the regression of the dummy indicating that the student's CATE is above the median on baseline characteristics. In columns *G9 GPA*, the CATE is predicted using GPA as the targeted outcome ; in columns *Nat. Exam*, the national exam scores is the targeted outcome. *Q1 mean* shows the mean for students whose CATE is below the median CATE. *Q2 column* shows the difference between students above and below the median CATE. All regressions include cohort and school fixed effect. Standard errors are clustered at the school*cohort level. *10%, **5%, ***1% significance level

For Online Publication

Appendix Figures

Figure A1: Evaluation Design

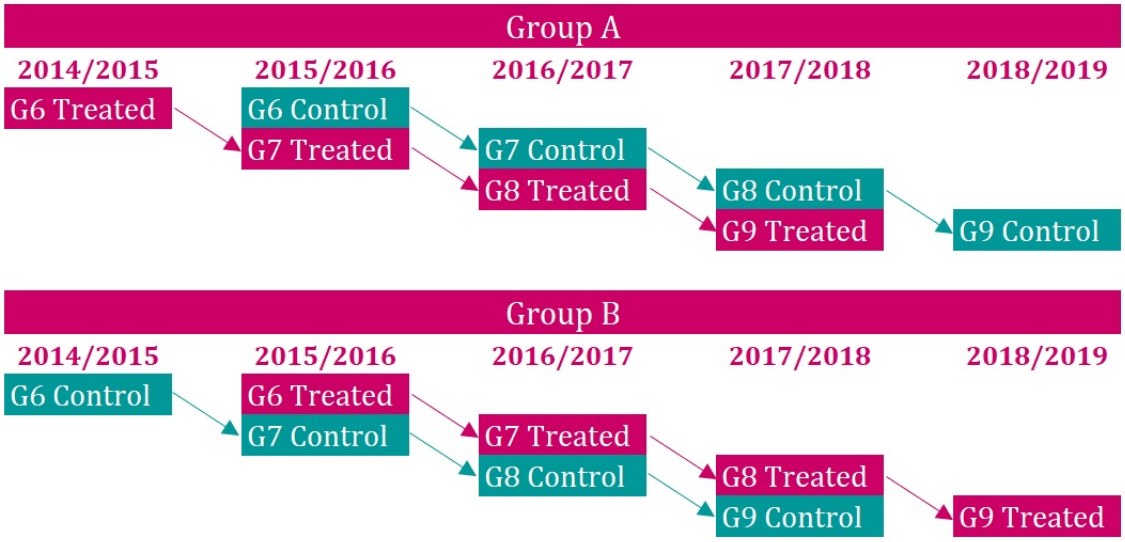


Figure A2: Content of the intervention

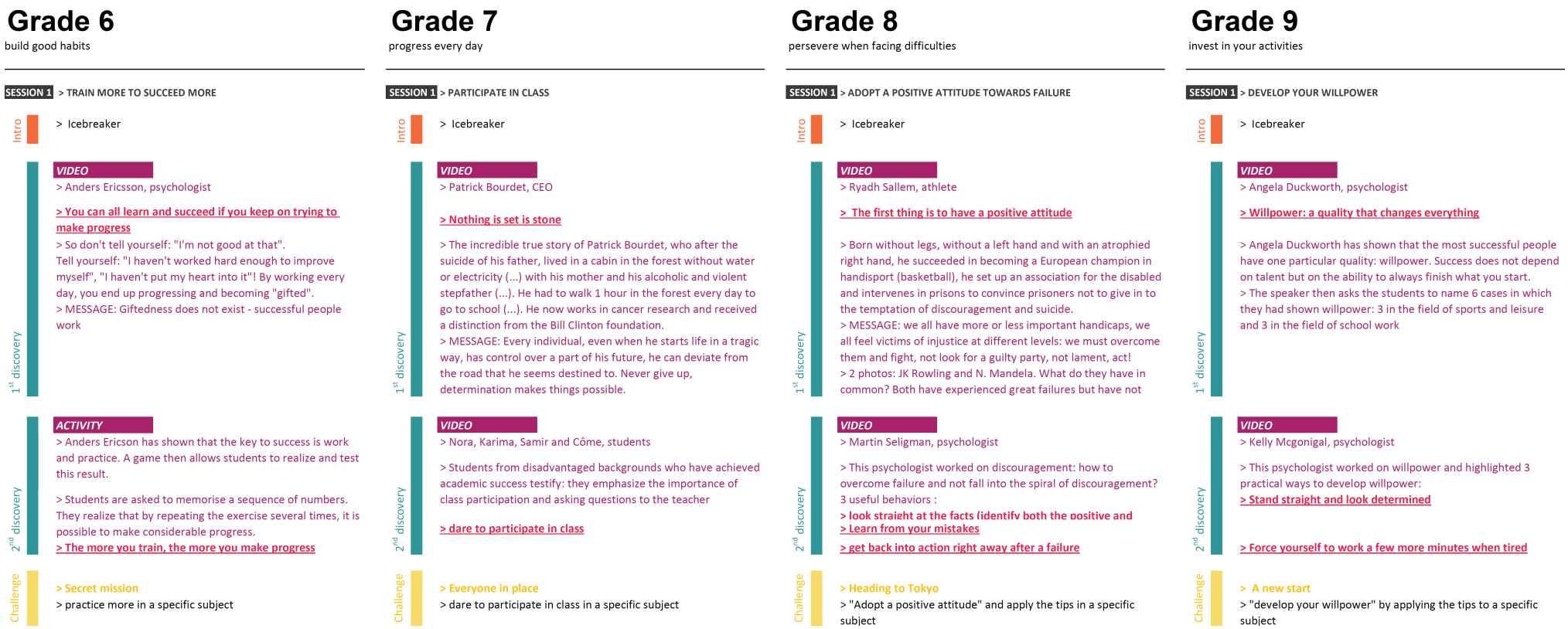


Figure A2: Content of the intervention - *Continued*

Grade 6

build good habits

SESSION 2 > THE POWER OF POSITIVE HABITS

Introduction

> **Review of session 1:** Do they remember what they learned in the first session? Did they rise to the "Secret Mission" challenge?
> **Reminder of the previous lessons:**
- You are all capable of learning and succeeding if you always try to progress
- The more you practice, the more you make progress

1st discovery

VIDEO
> The power of positive habits

> When you do something always at the same time, it becomes easy!

> Repeating a daunting and discouraging action every day at the same time and see how turns it into an automatism: it becomes easy!

2nd discovery

VIDEO
> The 3 winning habits

> It has been proven that some habits promote success. These include:

> Leaving on time every morning
> Tidying up one's things
> Going to bed before feeling sleepy

Commitment

> Choice of a commitment, written on a sticker
> **Self-diagnosis:** for each habit, is it something the student does: every day/almost never/it depends
> **Choice of a commitment:** record 1 habit among the 3 suggested habits (possibility of choosing another one)
> **Setting up a "plan B"**: think about potential obstacles and setting up a strategy to succeed in keeping the commitment despite obstacles
> **Choice of an older person** to ask for help if needed

Grade 7

progress every day

SESSION 2 > HOW TO MAKE PROGRESS EVERY DAY

Introduction

> **Review of session 1:** Do they remember what they learned in the first session? Did they rise to the "Everyone in place" challenge?
> **Reminder of the previous lessons**:
- Nothing is decided in advance
- Dare to participate in class

1st discovery

VIDEO
> Nora, Karima, Samir et Côme, students

> Successful students who come from a disadvantaged background explain what brought them to this point:
- Working at home on a more regular basis
- Not giving up even when it's difficult, when they are discouraged and feel like they are failing
- Daring to talk about one's difficulties
- Seeing school as an opportunity, a "ticket to somewhere else", to expand one's horizon beyond one's environment, to broaden one's aspirations ("the dream of travelling for me was a dream that I didn't have but was built thanks to school")

2nd discovery

VIDEO
> How to make progress every day

> Learning experts have identified four principles that make the difference between those who make progress and those who do not:

> Participate in class
> Dedicate a short time to work every day
> Focus 100% when working
> Training, practicing, repeating exercises, every day

Commitment

> Choice of a commitment, written on a sticker
> **Self-diagnosis:** for each habit, is it something the student does: every day/almost never/it depends
> **Choice of a commitment:** record 1 habit among the 4 suggested habits (possibility of choosing another one)
> **Setting up a "plan B":** think about potential obstacles and setting up a strategy to succeed in keeping the commitment despite obstacles
> **Choice of an older person** to ask for help if needed

Grade 8

persevere when facing difficulties

SESSION 2 > KEEP YOUR RESOLUTIONS

Introduction

> **Review of session 1:** Do they remember what they learned in the first session? Did they rise to the "Heading to Tokyo" challenge?
> **Reminder of the previous lessons:**
- The most important thing is to have a positive attitude
- Be reading to look straight at the facts
- Learn from one's mistakes

1st discovery

VIDEO
> Nora, communication manager

> Nora studied hard and became a communication manager after a very modest childhood. How did she do it? She dreamed of becoming a journalist, took friends as models and she discovered the joy of learning after a presentation on Mandela. She hung in there, did not get discouraged despite a few failures and finally succeeded.
> Message: Succeeding in school and then in one's professional like is possible! Students are asked to identify an asset that they and Nora have, and that made Nora's success possible (e.g., to take good students as role models, to get fully involved in homework, to work and go for it instead of getting discouraged if difficulties arise, to be motivated by

2nd discovery

VIDEO
> Changing one's habits, changing one's life

> When one makes a decision, it can be hard to stick to it. Research has shown that those who succeed are those who plan their activities and make them regular: for example, they may decide to do their homework every day at the same time. Those who succeed are also those who rely on friends who have the same goals, with whom they

> I listen to my teachers carefully.
> I do my homework at fixed times.
> I do my homework away from the TV or my phone
> I go to bed before 10pm
> I stay with friends who have good habits

Commitment

> Choice of a commitment, written on a sticker
> **Self-diagnosis:** for each habit, is it something the student does: every day/almost never/it depends
> **Choice of a commitment:** record 1 habit among the 4 suggested habits (possibility of choosing another one)
> **Setting up a "plan B":** think about potential obstacles and setting up a strategy to succeed in keeping the commitment despite obstacles
> **Choice of an older person** to ask for help if needed

Grade 9

invest in your activities

SESSION 2 > DO ONE THING AT A TIME AND DO IT THOROUGHLY

Introduction

> **Review of session 1:** Do they remember what they learned in the first session? Did they rise to the "New start" challenge?
> **Reminder of the previous lessons:**
- A quality that changes everything: willpower
- Be reading to look straight at the facts
- Be 100% concentrated from the very start of a task
- Forcing yourself to work a few more minutes when you're tired

1st discovery

VIDEO
> Mihaly Csikszentmihalyi, psychologist

> To understand what makes people happy, he asked volunteers to wear a belt that beeps at various times during the day: when the belt beeps, the volunteers had to: 1) record what they were doing + 2) record their happiness level from 0 to 10.
> Looking only at the 10/10, Professor Csikszentmihalyi realized that in all the happiness situations, people were doing only one activity at a time and doing it at 100% (dancing, playing, drawing, playing sports, doing their job...). People were completely concentrated and did not think about anything else.

> One thing at a time... but 100%!

2nd discovery

ACTIVITY
> "Concentrator"

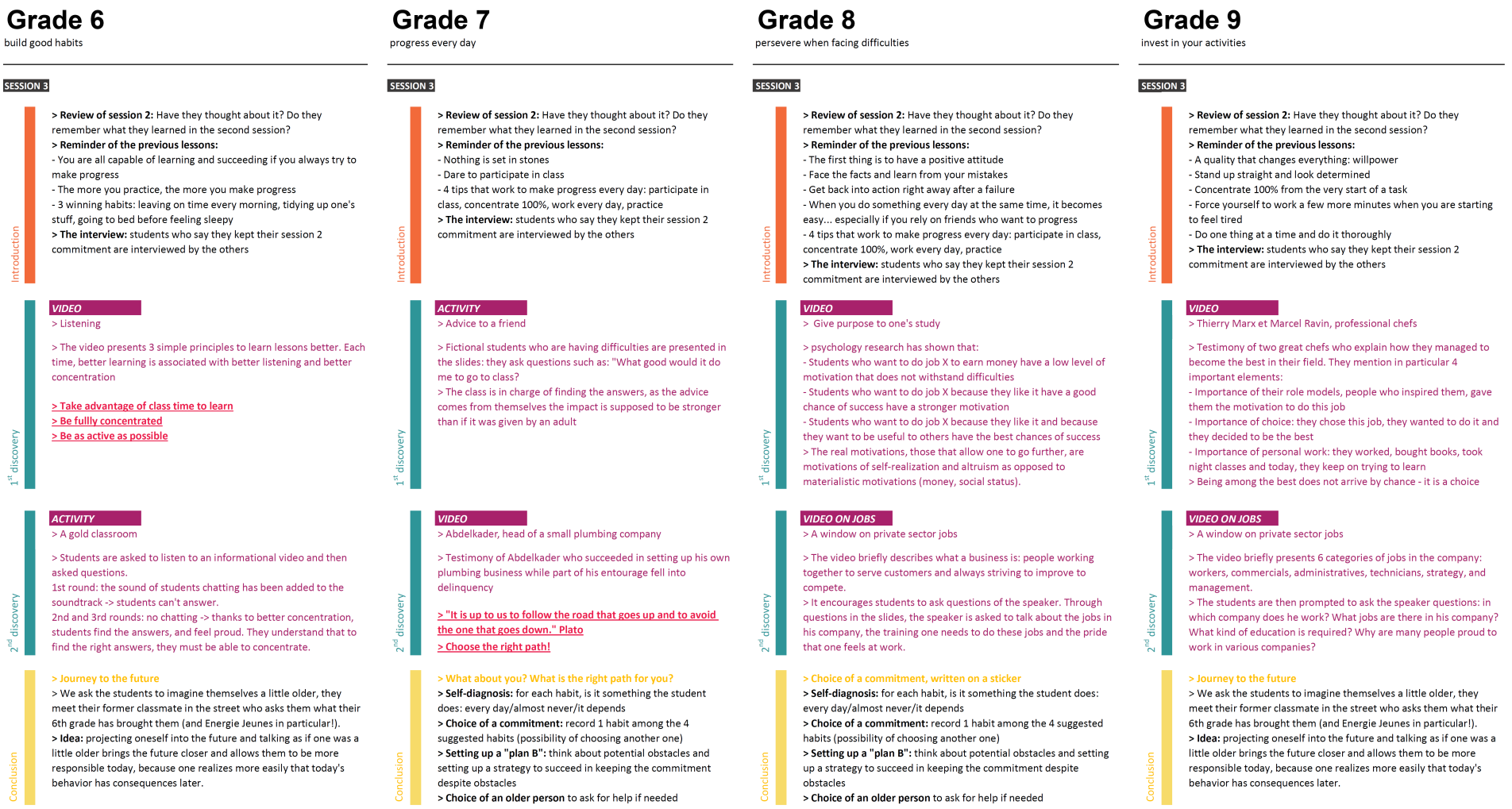
> Game in 2 rounds: a short documentary is shown. In the first round, the documentary is interrupted by questions -> students do not know how to answer. In the second round the documentary is not interrupted -> students manage to answer the questions and are proud of it. Students understand the importance of doing things 1 by 1 and 100%

> When I am in class, I concentrate fully on the lesson
> When I do my homework, I switch off the TV

Commitment

> Choice of a commitment, written on a sticker
> **Self-diagnosis:** for each habit, is it something the student does: every day/almost never/it depends
> **Choice of a commitment:** record 1 habit among the 4 suggested habits (possibility of choosing another one)
> **Setting up a "plan B":** think about potential obstacles and setting up a strategy to succeed in keeping the commitment despite obstacles
> **Choice of an older person** to ask for help if needed

Figure A2: Content of the intervention - *Continued*



Appendix Tables

Table A1: Sources and Items included in the Indices and Sub-Indices

Indice and sub-indices	Items	Source
GPA		
	Average grade French, Maths, Physics, Biology, History, Geography, Sports, Foreign languages, Design & Technology	Administrative data
Mindset		
Locus of control	Prob. success if from poor neighborhood	Guyon and Huillery (2021)
	Gap poor/wealthy neigborhood	
	Prob. success if parents without degree	
	Gap with and without college degree	
Growth Mindset	Intelligence is something that can't be changed	Claro et al. (2016)
	You can learn new things, not your intelligence	
	I prefer problems that I'll learn a lot from...	Li and Bates (2017)
	Prob. success if gifted but do not study hard	Guyon and Huillery (2021)
	Prob. success if study regularly	
	Prob. success if who is under-performing	
Self-assessed diligence		
Orderliness	I am organized and neat	Goldberg (1990)
	I work carefully, take one's time to get things right	
	I plan things ahead, think before acting	
	I am reliable, show up on time	
Grit	New ideas and projects sometimes distract me	Duckworth and Quinn (2009)
	Setbacks are not discouraging, I don't give up	
	I am obsessed with projects but quickly lose interest	
	I am obsessed with projects but quickly lose interest	
	I am a hard worker	
	I set a goal then change goal	
	I have problems focusing on long-term projects	
	I finish whatever is started	
	I am diligent, never give up	
School-work impulsivity	I forget things needed for class	Tsukayama et al. (2013)
	I interrupt other students	
	I say rude things	
	I lose things because I am messy (eg, desk, bedroom)	
	I lose temper at home or at school	
	I forget what teacher instructed	
	I let my mind wander instead of listening	
	I talk back to teacher or parent when I am upset	
Work discipline	I am always prepared	Goldberg et al. (2006)
	I get chores done right away	
	I start tasks right away	
	I get to work at once	
	I carry out plans	
	I waste my time	
	I find it difficult to get down to work	
	I need a push to get started	
	I have difficulty starting tasks	
	I postpone decisions	

Table A1: Sources and Items included in the Indices and Sub-Indices - *Continued*

Indice and sub-indices	Items	Source
Self-assessed diligence (cont.)		
Homework Manag.	I locate the materials needed for homework	Xu and Wu (2013)
	I find a quiet area	
	I remove things from table	
	I make enough space to work	
	I turn off the TV	
	I set priority and plan ahead	
	I keep track of what remains to be done	
	I remind myself of the remaining time	
	I tell myself to work more quickly when late	
	I find ways to make homework more interesting	
	I praise myself for good work/effort	
	I reassure myself about my abilities	
	I try not to be bothered with previous mistakes	
	I try to pay attention to what needs to be done	
	I try to calm down	
	I tell myself that I can do it	
	I daydream during homework	
	I start conversations unrelated to work	
	I play around during homework	
	I stop homework to eat or drink	
hours of homework		
	Time spent on homework yesterday?	Authors
	Time spent on homework day before yesterday?	
Teacher-reported character		
Achievement	Finishes whatever I begin	Park et al. (2017)
	Comes to class prepared	
	Works independently with focus	
	Gets to work right away rather than procrastinating	
	Remembers and follows directions	
	Believes that effort will improve her future	
	Tries very hard even after experiencing failure	
	Actively listens to others	
Intelligence	Shows enthusiasm	
	Invigorates others	
	Actively participates	
	Asks and answers questions to deepen understanding	
	Is eager to explore new things	
Social	Keeps her temper in check	
	Remains calm even when criticized or otherwise provoked	
	Gets over frustrations and setbacks quickly	
	Is polite to adults and peers	
	Demonstrates respect for the feelings of others	
	Allows others to speak without interruption	
	Is able to find solutions during conflicts with others	
	Recognizes and shows appreciation for other	
	Recognizes and shows appreciation for her opportunities	
School-reported behavior		
	1/2 day absenteism	
	number of recorded lateness	Administrative data
	disciplinary actions	
	sanctions	

The tables gives the composition of each index and sub-index as well as their sources (references).

Table A2: Comparison between EJ Sample and other French schools

	EJ sample	EJ sample - PE schools	EJ sample - Pub. schools	EJ sample - all schools
Panel A: Schools' characteristics				
Public school (%)	1.00	0.00	0.00	0.24***
		(0.00)	(0.00)	(0.01)
Priority education (%)	0.79	-0.21***	0.60***	0.65***
		(0.04)	(0.04)	(0.04)
# Students per school	528.05	18.8	32.5**	49.3***
		(14.5)	(13.1)	(13.1)
# Teachers per school	46.08	1.89	7.16***	7.92***
		(1.16)	(1.09)	(1.09)
# Students per class	23.19	0.64***	-1.74***	-2.02***
		(0.20)	(0.19)	(0.19)
Remoteness index	-2.07	-1.06***	-2.24***	-2.10***
		(0.14)	(0.06)	(0.06)
Panel B: Teachers' characteristics				
Years of service	5.89	-0.06	-1.38***	-1.25***
		(0.25)	(0.25)	(0.25)
$<$ 2 years in school (%)	0.35	0.01	0.05***	0.05***
		(0.01)	(0.01)	(0.01)
Part-time (%)	0.06	0.00	-0.03***	-0.03***
		(0.00)	(0.00)	(0.00)
Female (%)	0.6	0.00	-0.04***	-0.05***
		(0.01)	(0.01)	(0.01)
Permanent (%)	0.91	0.01	-0.03***	0.01*
		(0.01)	(0.01)	(0.01)
Agrégés (%)	0.04	0.00	-0.01***	0.00
		(0.00)	(0.00)	(0.00)
Turnover (%)	0.18	-0.01	0.03**	0.03**
		(0.01)	(0.01)	(0.01)

Table A2: Comparison between EJ Sample and other French schools - *Continued*

	EJ sample	EJ sample - PE schools	EJ sample - Pub. schools	EJ sample - all schools
Panel C: Students' characteristics				
G6 ever held back (%)	0.12	0.00	0.04***	0.05***
		(0.01)	(0.01)	(0.01)
G6 National test	229.57	0.05	-1.10***	-1.27***
		(0.09)	(0.08)	(0.08)
G9 National test	7.04	-0.10	-1.22***	-1.46***
		(0.10)	(0.09)	(0.09)
Female (%)	0.49	0.00	0.00	0.00
		(0.00)	(0.00)	(0.00)
With financial aid (%)	0.49	-0.02	0.20***	0.24***
		(0.02)	(0.02)	(0.02)
Living in a poor area (%)	0.42	0.05*	0.32***	0.33***
		(0.03)	(0.03)	(0.03)
Social position index (%)	0.83	0.02*	-0.18***	-0.21***
		(0.01)	(0.01)	(0.01)
High SES (%)	0.17	0.03**	-0.13***	-0.17***
		(0.01)	(0.01)	(0.01)
Low SES (%)	0.78	-0.02	0.12***	0.16***
		(0.01)	(0.01)	(0.01)
Post-G9 enrollment:				
Academic high school (%)	0.59	0.02**	-0.05***	-0.07***
		(0.01)	(0.01)	(0.01)
Technical high school (%)	0.35	-0.01	0.07***	0.10***
		(0.01)	(0.01)	(0.01)
G9 held back (%)	0.03	0.00	0.01**	0.00*
		(0.00)	(0.00)	(0.00)

This table shows the differences between our experimental sample of middle schools and the population of French middle schools. Column *EJ sample* provides the mean for the experimental sample. Columns *EJ sample - PE schools*, *EJ sample - pub. schools* and *EJ sample - all schools* show the differences between our sample and schools in Priority Education, Public schools, and all French middle schools respectively. Standard errors are reported in parentheses and are clustered at the school level. The remoteness index is computed using the proportion of students living in a rural commune, the range of courses offered around the school, the distance of the school from sports and cultural facilities (the higher the index value, the more remote the school is). *Agrégés* are qualified teachers who have passed the high-level competitive *Aggregation* exam. *G6 and G9 national tests* are the average scores obtained in mathematics and French at the national tests administered at the beginning of Grade 6 and at the end of Grade 9, respectively. Coefficients are standardized using the mean and standard deviation of all French public middle schools. The social position index is created from family environment information: parental education, income, cultural practices, housing conditions. The higher the index value, the more favorable the family background is. *10%, ** 5%, *** 1% significance level

Table A3: Baseline Balancing

	G6 sample		**G7 sample**		**G8 sample**		**G9 sample**		**Nat. ex. sample**		Nat.Ex.
	C	T-C	C	T-C	C	T-C	C	T-C	C	T-C	- G9
Panel A: Full Sample											
Date of Birth	2,003	-0.002	2,003	0.004	2,003	0.004	2,003	0.000	2003.32	0.00	0.03
	[0.656]	(0.005)	[0.667]	(0.005)	[0.686]	(0.005)	[0.710]	(0.006)	[0.68]	(0.00)	(0.02)
Female	0.490	0.002	0.493	0.003	0.489	0.008	0.489	0.007	0.50	0.01	0.01***
	[0.500]	(0.005)	[0.500]	(0.005)	[0.500]	(0.005)	[0.500]	(0.006)	[0.50]	(0.01)	(0.00)
Single parent fam.	0.177	-0.005	0.179	-0.004	0.188	-0.009**	0.191	-0.011**	0.18	-0.01*	-0.01***
	[0.382]	(0.005)	[0.384]	(0.004)	[0.391]	(0.004)	[0.393]	(0.005)	[0.39]	(0.01)	(0.00)
# of children	2.388	-0.009	2.387	-0.016	2.401	0.001	2.448	0.008	2.46	-0.01	0.02
	[1.686]	(0.028)	[1.689]	(0.025)	[1.666]	(0.025)	[1.627]	(0.023)	[1.61]	(0.02)	(0.02)
Financial aid	0.504	0.013	0.512	-0.004	0.506	-0.001	0.491	0.000	0.49	0.00	-0.00
	[0.500]	(0.008)	[0.500]	(0.005)	[0.500]	(0.006)	[0.500]	(0.005)	[0.50]	(0.01)	(0.01)
Blue collar	0.764	-0.011**	0.765	-0.009*	0.770	-0.010**	0.766	-0.011**	0.76	-0.01**	-0.01*
	[0.425]	(0.005)	[0.424]	(0.005)	[0.421]	(0.005)	[0.423]	(0.005)	[0.43]	(0.00)	(0.01)
Foreigner	0.125	0.004	0.133	-0.001	0.152	-0.003	0.153	0.004	0.14	0.00	-0.01
	[0.331]	(0.004)	[0.340]	(0.003)	[0.359]	(0.004)	[0.360]	(0.004)	[0.35]	(0.00)	(0.00)
Was held back	0.180	-0.001	0.172	-0.004	0.167	-0.004	0.155	-0.005	0.14	-0.01	-0.01***
	[0.384]	(0.004)	[0.378]	(0.004)	[0.373]	(0.004)	[0.362]	(0.004)	[0.35]	(0.00)	(0.00)
Early GPA	13.18	-0.002	13.31	0.022	13.35	0.027	13.48	0.045	13.6	0.03	0.13***
	[2.674]	(0.041)	[2.586]	(0.040)	[2.571]	(0.038)	[2.527]	(0.039)	[2.46]	(0.04)	(0.04)
Early behavior score	-0.008	0.023*	-0.009	0.021	-0.008	0.010	0.001	0.001	-0.02	0.00	-0.02***
	[0.644]	(0.014)	[0.614]	(0.013)	[0.622]	(0.013)	[0.618]	(0.013)	[0.58]	(0.01)	(0.01)

Table A3: Baseline Balancing - *Continued*

	G6 sample		**G7 sample**		**G8 sample**		**G9 sample**		Panel A
	C	T-C	C	T-C	C	T-C	C	T-C	-B
Panel B: Student reported survey									
Date of Birth	2,003	-0.019**	2,003	0.009	2,003	0.013	2,003	0.014	0.05*
	[0.625]	(0.009)	[0.656]	(0.007)	[0.651]	(0.008)	[0.692]	(0.011)	(0.03)
Female	0.497	0.001	0.489	0.017	0.501	-0.007	0.507	-0.006	0.02*
	[0.500]	(0.011)	[0.500]	(0.011)	[0.500]	(0.012)	[0.500]	(0.011)	(0.01)
Single parent family	0.175	0.001	0.174	-0.019***	0.192	-0.022**	0.182	-0.012	-0.01
	[0.380]	(0.008)	[0.379]	(0.007)	[0.394]	(0.008)	[0.386]	(0.009)	(0.01)
# of children	2.425	0.010	2.358	-0.009	2.387	-0.057*	2.502	-0.031	0.05*
	[1.683]	(0.042)	[1.690]	(0.034)	[1.661]	(0.032)	[1.573]	(0.035)	(0.03)
Financial aid	0.523	0.013	0.506	-0.007	0.509	-0.001	0.514	-0.004	0.02**
	[0.500]	(0.011)	[0.500]	(0.007)	[0.500]	(0.010)	[0.500]	(0.010)	(0.01)
Blue collar	0.759	0.011	0.762	-0.009	0.763	-0.001	0.756	0.001	-0.01
	[0.428]	(0.008)	[0.426]	(0.008)	[0.426]	(0.008)	[0.430]	(0.008)	(0.01)
Foreigner	0.122	0.003	0.127	0.001	0.134	0.001	0.135	0.017**	-0.02**
	[0.327]	(0.007)	[0.333]	(0.006)	[0.341]	(0.005)	[0.342]	(0.007)	(0.01)
Was held back	0.160	0.015*	0.158	-0.007	0.159	-0.017**	0.149	-0.012	-0.01
	[0.367]	(0.008)	[0.365]	(0.006)	[0.365]	(0.008)	[0.356]	(0.008)	(0.01)
Early GPA	13.37	-0.060	13.48	0.034	13.52	0.015	13.73	0.040	0.25***
	[2.500]	(0.064)	[2.538]	(0.061)	[2.515]	(0.066)	[2.429]	(0.072)	(0.05)
Early behavior score	-0.053	0.040**	-0.043	0.011	-0.054	0.017	-0.044	0.001	-0.05***
	[0.521]	(0.019)	[0.571]	(0.015)	[0.529]	(0.019)	[0.527]	(0.019)	(0.01)

Table A3: Baseline Balancing - *Continued*

	G6 sample		G7 sample		G8 sample		G9 sample		Panel A
	C	T-C	C	T-C	C	T-C	C	T-C	-C
Panel C: Teacher reported survey									
Date of Birth	2,003	-0.013	2,003	0.029***	2,003	0.023**	2,003	-0.003	0.06*
	[0.627]	(0.010)	[0.649]	(0.008)	[0.644]	(0.011)	[0.694]	(0.012)	(0.03)
Female	0.495	0.015	0.489	0.024**	0.492	0.019	0.515	0.006	0.03**
	[0.500]	(0.013)	[0.500]	(0.011)	[0.500]	(0.015)	[0.500]	(0.012)	(0.01)
Was held back	0.177	-0.002	0.165	-0.008	0.192	-0.023**	0.188	-0.017*	-0.00
	[0.382]	(0.010)	[0.371]	(0.009)	[0.394]	(0.010)	[0.391]	(0.009)	(0.01)
Single parent family	2.428	0.041	2.349	0.030	2.415	-0.062	2.505	-0.036	0.06
	[1.683]	(0.049)	[1.691]	(0.044)	[1.653]	(0.040)	[1.582]	(0.040)	(0.04)
# of children	0.523	-0.001	0.517	-0.013	0.505	0.015	0.515	-0.007	0.02*
	[0.500]	(0.013)	[0.500]	(0.009)	[0.500]	(0.012)	[0.500]	(0.012)	(0.01)
Financial aid	0.761	0.007	0.761	0.001	0.763	0.001	0.765	-0.002	-0.00
	[0.427]	(0.009)	[0.427]	(0.009)	[0.425]	(0.010)	[0.424]	(0.010)	(0.01)
Blue collar	0.129	0.001	0.129	-0.003	0.144	-0.001	0.133	0.016*	-0.02**
	[0.335]	(0.009)	[0.335]	(0.007)	[0.351]	(0.007)	[0.340]	(0.008)	(0.01)
Foreigner	0.166	0.013	0.163	-0.020***	0.166	-0.029***	0.147	-0.006	-0.01
	[0.372]	(0.009)	[0.369]	(0.007)	[0.372]	(0.011)	[0.354]	(0.009)	(0.01)
Early GPA	13.33	0.009	13.46	0.055	13.47	0.016	13.70	0.090	0.22***
	[2.501]	(0.074)	[2.515]	(0.066)	[2.540]	(0.075)	[2.421]	(0.091)	(0.06)
Early behavior score	-0.039	0.025	-0.030	0.005	-0.037	0.013	-0.038	0.000	-0.04**
	[0.541]	(0.022)	[0.597]	(0.015)	[0.560]	(0.022)	[0.561]	(0.020)	(0.02)

This table shows pre-treatment differences between the treatment and the control group. Under column *C*, we provide the mean and standard deviation of the variable in the control group, and under column *T-C* we show the coefficient of the regression of the outcome on the treatment variable, controlled for school and cohort fixed effects. Standard error in parentheses, robust to heteroscedasticity and clustered at the school*cohort level. *(-)* indicates that the value was inverted so that a higher value means a better outcome. *10%, ** 5%, *** 1% significance level

Table A4: Impacts on Sub-Indices

	Grade 9	**Grade 6**	**Grade 7**	**Grade 8**
Student mindset				
Locus of Control	0.06***	.	0.07***	0.09***
	(0.02)	.	(0.02)	(0.02)
Growth Mindset	0.10***	.	0.09***	0.09***
	(0.02)	.	(0.02)	(0.02)
Self-assessed diligence				
Orderliness	-0.02	0.03	0.00	-0.01
	(0.02)	(0.02)	(0.02)	(0.02)
Grit	-0.02	-0.08***	-0.07***	-0.06***
	(0.03)	(0.02)	(0.02)	(0.02)
Impulsivity (-)	0.06**	0.04*	-0.01	0.03
	(0.02)	(0.03)	(0.02)	(0.03)
Work discipline	0.01	0.00	-0.05**	-0.04**
	(0.02)	(0.02)	(0.02)	(0.02)
Homework manag.	-0.03	-0.01	-0.05**	-0.03*
	(0.02)	(0.03)	(0.02)	(0.02)
Hours of homework	0.01	.	-0.04	0.02
	(0.03)	.	(0.03)	(0.03)
Teacher-reported character				
Intellectual character	0.04	0.03	0.07**	0.03
	(0.03)	(0.02)	(0.03)	(0.03)
Achievement character	0.03	0.05*	0.05*	0.03
	(0.03)	(0.03)	(0.03)	(0.03)
Social character	0.01	0.05	0.07***	0.06*
	(0.04)	(0.03)	(0.03)	(0.03)
School-reported behavior				
1/2 day abs. (-)	0.02	-0.09***	-0.01	-0.01
	(0.02)	(0.03)	(0.02)	(0.02)
Lateness (-)	0.04	0.00	-0.01	-0.07**
	(0.03)	(0.02)	(0.02)	(0.03)
Disciplinary actions (-)	0.04	-0.02	0.00	0.00
	(0.04)	(0.02)	(0.02)	(0.02)
Sanctions (-)	0.04	0.00	-0.03	0.03
	(0.03)	(0.02)	(0.02)	(0.03)

Same as Table 3 for sub-indices. *(-)* indicates that the value was inverted so that a higher value means a better outcome. *10%, **5%, ***1% significance level.

Table A5: Treatment Effects by Length of Exposure

	Full sample		**4 years sample**		**$\leq$ 3 years**		**4 v.$\leq$ 3y**
	Obs.	Impact	Obs.	Impact	Obs.	Impact	P-value
Final outcomes							
GPA	19,330	0.05**	15,271	0.05**	4,059	0.04	0.04
		(0.02)		(0.02)		(0.03)	(0.03)
National exam	19,702	0.01	16,125	0.01	3,577	0.00	0.00
		(0.02)		(0.02)		(0.03)	(0.03)
... French	19,660	0.03**	16,094	0.03**	3,566	0.01	0.01
		(0.01)		(0.02)		(0.02)	(0.02)
... Math	19,625	-0.01	16,071	-0.01	3,554	-0.03	-0.03
		(0.02)		(0.02)		(0.03)	(0.03)
Aspiration	5,497	0.04*	4,501	0.04*	996	0.01	0.01
		(0.02)		(0.02)		(0.05)	(0.05)
Intermediary outcomes							
Student mindset	5,485	0.11***	4,493	0.10***	992	0.06	0.06
		(0.02)		(0.02)		(0.05)	(0.05)
Self-assessed diligence	5,497	0.00	4,501	0.00	996	-0.04	-0.04
		(0.03)		(0.03)		(0.06)	(0.06)
Teacher-rep. character	4,503	0.02	3,711	0.01	792	0.07	0.07
		(0.03)		(0.03)		(0.07)	(0.07)
School-rep. behavior	22,305	0.06***	17,387	0.05**	4,918	0.09**	0.09**
		(0.02)		(0.02)		(0.04)	(0.04)
Years of exposure	3.50		4.00		1.74		

Same as Table 3 but treatment effects are presented for varying length of exposure. Columns *Full Sample* reproduce the results from Table 3. Columns *4-year sample* give the impacts for children who were exposed to the treatment during 4 years. Columns ≤ 3 *years* give the impacts for children who were exposed to the treatment at most 3 years. Columns *4 v.*≤ 3 *years* give the p-values of the comparison between these two samples. *10%, **5%, ***1% significance level.

Table A6: Final Impacts and Intermediate Dynamics - Using Weights to Correct for Attrition

	Final impact		**Dynamics**						**P-value**		
	Grade 9		**Grade 6**		**Grade 7**		**Grade 8**		**G6=G9**	**G7=G9**	**G8=G9**
	Obs.	Impact	Obs.	Impact	Obs.	Impact	Obs.	Impact			
Final outcomes											
GPA	19,330	0.05**	20,783	0.03***	21,443	0.03**	19,713	0.01	0.45	0.31	0.08
		(0.02)		(0.01)		(0.01)		(0.02)			
National exam	19,702	0.01		.		.		.	.	.	.
		(0.02)		.		.		.			
...French	19,660	0.03**		.		.		.	.	.	.
		(0.01)		.		.		.			
...Math	19,625	-0.01		.		.		.	.	.	.
		(0.02)		.		.		.			
Aspiration	5,497	0.04*		.		.		.	.	.	.
		(0.02)		.		.		.			
Intermediary outcomes											
Student mindset	5,485	0.11***		.	6,027	0.11***	5,496	0.11***	.	0.95	0.77
		(0.02)		.		(0.02)		(0.02)			
Self-assessed diligence	5,497	0.00	5,506	0.00	6,458	-0.05***	5,706	-0.01	0.66	0.15	0.83
		(0.03)		(0.02)		(0.02)		(0.02)			
Teacher-reported character	4,503	0.02	4,494	0.04	4,826	0.07**	4,596	0.06*	0.58	0.27	0.47
		(0.03)		(0.03)		(0.03)		(0.03)			
School-reported behavior	22,305	0.06***	22,074	-0.04***	22,449	0.01	22,445	0.01	0.00	0.10	0.13
		(0.02)		(0.01)		(0.02)		(0.02)			

Same as Table 3 but with inverse probability weights to correct for attrition and provide population estimates. *10%, **5%, ***1% significance level.