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]

July 21, 2022

Abstract

This article shows that a four-year mindset intervention in disadvantaged middle schools led to a 0.07 standard deviation increase in GPA, associated with more optimistic beliefs, improved school behavior, higher educational and professional aspirations, and higher exam test scores. International empirical benchmarks reveal that the intervention is highly cost-effective. However, the program benefits less those who may need it the most, especially less-disciplined students. Moreover, the effect size remains small despite sustained exposure over four years, which provides first evidence that the impact of mindset interventions quickly plateaus and therefore may not have the potential to fundamentally transform education outcomes.

JEL classification: I24

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 project SOWELL. Adrien Bouguen also acknowledges financial support from the German Research Foundation (DFG) project SFB 884 during his stay at the University of Mannheim. The project was approved by the Institutional Review Boards 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, de la Jeunesse et des Sports, DEPP: axelle.charpentier@education.gouv.fr

[4]Sciences Po, Paris: yann.algan@sciencespo.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 later educational or professional outcomes (Almlund et al., 2011; Castillo et al., 2011; Dohmen et al., 2011; Angela Lee Duckworth and Seligman, 2005; Golsteyn et al., 2014; Moffitt et al., 2011; Sutter et al., 2013), and Heckman et al. (2006) first established causality from non-cognitive skills to schooling decisions. In the wake of these results, many governments have encouraged a shift of educational priorities to promote the development of socioemotional 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 has also launched a multi-million pound push to improve character education, and in particular the capacity to strive for and succeed at long-term and higher-order goals.[1] But beyond correlational findings, the question is whether these character skills can indeed be taught, and whether such training has the potential to have a strong downstream impact on behavior and academic outcomes. Current scientific evidence suggests that character training programs can have positive effects at low cost. However, these effects are typically small and one important question is whether more intense character interventions may meet policymakers' high hopes in terms of academic performance and social inequality at school. Assessing whether longer and more intense exposure amplifies the effect is therefore of critical importance.

In this paper, we test the impact of *Energie Jeunes*, a program conducted in French disadvantaged middle schools to stimulate effort by helping adolescents develop a sense of self-efficacy and agency and view effort and hard work as a normal part of academic success, encouraging feelings of confidence that help bounce back from failure, and driving adolescents to seek out and engage with challenges rather than avoid them. The program is built on three components: first, the growth mindset component teaches students that the brain is highly plastic and grows stronger and smarter when it experiences rigorous and 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

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

a behavioral change that derives from the change in mindset. In this paper, we refer to the program as a 'mindset intervention' since it primarily works on perceptions and beliefs, with operational tools to translate these beliefs into behavioral change. The immediate goal of the program is to encourage students to reconsider their beliefs about their potential, and to view academic ability not as fixed but as something that can grow in response to sustained effort and healthy work habits. The ultimate goal is to increase students' motivation and discipline in order to improve their academic and life prospects. This is particularly important in a country like France, where many students demonstrate low perseverance and a weak internal locus of control, and where students from low socioeconomic backgrounds develop biased perceptions of their academic potential and future chances of success (Algan et al., 2018; Guyon and Huillery, 2020).

The novelty of the program is that it runs from Grade 6 to Grade 9, offering repeated exposure over four years. Each year, students and their homeroom teacher participated in three one-hour in-class sessions presented by two external facilitators. In addition to measuring the impacts of these yearly interventions on student performance and the mechanisms of their impacts, this paper also asks whether sustain exposure to small and low-cost mindset interventions amplifies the effect, or if the effect saturates after a certain level of exposure. If the interventions are complements, then the optimal educational policy would be to package several interventions together to raise their impact at modest cost. However, the mindset interventions may be substitutes, in which case the potential gains are much less promising. This paper therefore asks whether routinizing mindset interventions has any chance of generating an economically significant change in student performance.

The experiment was implemented in 97 disadvantaged middle schools located all over the country, which are representative of the population of Priority Education[2] middle schools in France. In each school, two cohorts of about five classes each participated in the experiment: students who entered Grade 6 in September 2014, and students who entered Grade 6 in September 2015. We randomly assigned one cohort in each school to either treatment or control. In half of the schools, students in the 2014 cohort benefited from the program during four years (from sixth grade to ninth grade[3]), while students in the 2015 cohort did not, and *vice-versa* in the other half of schools. Our sample consists of nearly a thousand classes and 23,000 students, which confers a high degree of external validity to this experiment.

We collected outcome measures from four sources: school administrative data, a survey administered to a representative sample of students (seven students per class), a survey administered to a sample of teachers (two teachers per class), and the results of the national exam administered at the end of middle school. The administrative data includes behavior at school (absenteeism, lateness, and disciplinary sanctions),

[2]disadvantaged schools that benefit from extra support from the State

[3]Middle school lasts 4 years in France from sixth grade to ninth grade

and grade point average (GPA). The teacher survey provides a measure of students' character as demonstrated in class. Finally, the student survey assesses perceived return to effort through growth mindset and locus of control questions, as well as self-reported diligence (orderliness, grit, school-work impulsivity, work discipline, and homework management). The student survey also asks about educational and professional aspirations in Grade 9. For each student, we collected four waves of outcomes, one per year spent in middle school. Finally, we collect the results of the national end-of-middle school exam which provides an anonymous and externally-graded assessment of students' academic knowledge. The richness of our data and its different points of view (national exam, school administration, teachers, and students) provide a unique opportunity to understand the precise channels through which a mindset intervention can change school outcomes.

We find that the intervention increased GPA by 0.07 standard deviation (hereafter, SD) at the end of middle school (ITT impacts) which corresponds to an impact of about 0.09 SD for a student who participated in at least one Energie Jeunes session. While the impact is approximately similar for low- and high-achievers, we find strong heterogeneity according to gender, socioeconomic status, and baseline behavior—i.e., absenteeism, lateness and school discipline. In fact, the impact appears immediately in Grade 6 and is stronger for girls (0.08 SD in Grade 9), non-aid recipients (0.08 SD), and students whose baseline behavior was relatively good (0.10 SD), whereas it is only in Grade 9 that boys and aid recipients start to academically benefit from the intervention, and for poorly-behaved students the effect remains insignificant. The effect on academic performance is confirmed by a 0.03 SD positive impact at the national end-of-middle school exam, with the same pattern of results as those found on GPA (larger and only significant for girls, non-aid recipients, and originally better behaved students). Since the national exams are anonymous and externally graded, this result proves without ambiguity that the program improved academic performances.

We next turn to the mechanisms by which the program improved academic achievement. As expected, the program made students more optimistic about the possibility of improving their intelligence and academic abilities through effort, and led them to attribute more weight to effort relative to innate talent or external constraints. The effect size on a summary index of perceived return to effort is about 0.04-0.05 SD in each grade. In line with the impact on GPA, the increase in perceived return to effort is about twice as big for girls, financial aid non-recipients, and well-behaved students than for their counterparts.

Importantly, this change in perceptions is generally associated with better behavior at school. Effect sizes on teacher-reported student character are 0.04-0.05 SD in Grades 7 and 8. More specifically, students are more disciplined in their school-work, more dynamic and enthusiastic about learning, and less impulsive, than in the

control group. Severe violations of school rules (absenteeism, lateness, insolence, disrespect, and violence) also decreased by 0.04 SD in Grade 9. For the initially less-disciplined students, i.e., boys, aid recipients, low-achievers and poorly-behaved students, there is a possible association between impacts on discipline and impacts on academic achievement. It is also worth noting that students do not report more time spent on homework, so the positive impact of GPA comes from improved quality (i.e., diligence, concentration, and discipline) rather than quantity of schoolwork.

Interestingly, self-reported diligence does not show a similar improvement as teacher- and school-reported measures. On the contrary, the program worsens some aspects of self-reported diligence for most students, in particular regarding self-reported grit, work discipline and homework management. Since teachers and school registers show that actual behavior improved, this result indicates that the program changed the reference point against which students compare their own character.

Finally, ninth graders in the treatment group are 2 percentage points (hereafter, pp) more likely to aspire to academic high school (rather than technical high school), and 2 pp more likely to aspire to a medium- or a high-skilled job (rather than a low-skilled job). The rise in aspirations is concentrated in girls, aid recipients, low achievers, and well-behaved students, whose choice of academic high school and medium- or high-skilled jobs increased by 3-6 pp. Overall, our results show that changing students' beliefs about their capacity to improve and returns to effort leads to significant gains in their academic behavior, performance, and aspirations.

This paper provides important new insights on mindset interventions. First, it improves the understanding of why mindset interventions affect school performance by providing evidence on a rich range of mechanisms and final outcomes, which is often lacking in the literature. A subsidiary benefit of the richness of our data is that it allowed us to uncover that self-reported behavioral measures are prone to reference-point biases, which may lead diligent students to not perceive themselves as such. Future studies in this area should thus use observational instead of self-reported measures. More generally, this paper is the first to test the full theory of change, going from students' beliefs and self-perceptions to time spent on homework, in-class attitude, behavior at school, educational and professional aspirations, and finally GPA and national exam scores, which is unique in the literature.

Second, this study was conducted in a natural setting including the largest number of schools, students, and facilitators ever involved in an experiment targeting non-cognitive skills, and with multiple points of view (students, teachers, school administration). Most existing papers use short-term indicators exterior to the education system, whereas we use indicators internal to the education system, those which determine educational paths. For these two reasons, this study offers particularly relevant and generalizable results to guide education policy.

Third, this paper improves our understanding of who benefits from a mindset in-

tervention and when the effects emerge. Our paper tests effect heterogeneity along four important dimensions: gender, socioeconomic background, academic performance, and school behavior. While existing papers often focus on baseline academic performance only, we show that the main driver of heterogeneity is baseline school behavior, and to a lesser extent gender and socioeconomic background. We also show that mindset interventions eventually benefit everyone, but that the effects are quicker and larger on students who show better academic attitudes and performance to begin with. Mindset programs in disadvantaged schools may thus help reduce the gap between these schools and the national population, while widening within-school inequality in academic performance.

Finally, we evaluate a mindset intervention that is sustained over four years, instead of a short one-shot intervention as in previous studies (Bettinger et al., 2018; Cohen et al., 2009; Outes-Leon et al., 2020; Paunesku et al., 2015; Yeager et al., 2019).[4] This sustained intervention over four years had small but robust effects, which can be interpreted in two ways. One the one hand, if we pit our results against the broader literature on the impact of education interventions in high-income countries, this intervention clearly has a high cost-efficiency ratio. For instance, the median effect of the randomized trials that evaluated promising 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 0.07 SD on GPA or 0.03 SD on exam test scores for a total cost of €65 (i.e., $75) per student. On the other hand, the dynamic of impacts throughout middle school indicates that the effects quickly materialized for girls, non-financial aid recipients and initially better behaved students in Grade 6, but then plateaued. For instance, the heterogeneous effect on initially better behaved students already reached 0.08 SD by the end of Grade 6 while it only attained

[4]Cohen et al., 2009 randomly assigned 385 high school students to a series of brief writing assignments focusing on a self-affirming value, and find that this intervention increased school grades of Black students by 0.24 grade points two years after the intervention. Paunesku et al., 2015 delivered one-hour growth-mindset and sense-of-purpose interventions through online modules to 1,594 students in 13 high schools. Each intervention raised students' semester GPA by about 0.04 SD, concentrated among the 30% of the students at highest risk of dropping out of high school. Bettinger et al., 2018 tested a growth mindset intervention focused on math in a small sample of 385 high school students in Norway. Two 45-minute online sessions led to a 0.19 SD increase on an independent algebra task administered two weeks after the intervention. At a larger scale, Yeager et al., 2019 randomized 13,000 ninth graders from 65 schools in the US into a growth mindset intervention or a placebo intervention. The growth mindset intervention was a short computer-based intervention delivered in class across two half-hour sessions. This very short intervention increased pupils' GPA by 0.05 SD at the end of the school year, and benefited the low-achieving students more. Finally, Outes-Leon et al., 2020 show that a growth mindset intervention in Peru increased math test scores by 0.05 SD, providing the first evidence of the success of growth mindset stimulation in a developing country.

0.1 SD by the end of middle school, after four years of program's exposure. Boys, aid-recipients and the students with initially lower GPA eventually benefit from the program in Grade 9, but the initially poorly-behaved students still do not. This paper thus provides first evidence that mindset interventions performed by external actors, although highly cost-effective, may not have the potential to be truly transformative. It implies that involving other actors such as parents or teachers may be necessary to transform educational outcomes more substantively (Alan, Boneva, et al., 2019; Alan and Ertac, 2018).[5]

The rest of the paper is structured as follows: Section 2 presents the intervention and the evaluation design. Section 3 presents the sampling strategy and data. Section 4 verifies that the experimental protocol was conducted in accordance with scientific standards. Section 5 presents the impact on GPA, and Section 6 explores the mechanisms explaining this impact, such as students' mindset, behaviors, and aspirations. Section 7 concludes.

2 Experimental Design

2.1 Content of the Intervention

The goal of the *Energie Jeunes* program is to improve 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[6]) created in 2009. It consists of three 55-minute class interventions per year during the four years of middle school, i.e., from Grade 6 (about 11-year-old students) to Grade 9 (about 14-year-old students). Every school year, the NGO selects and trains two external facilitators who are responsible for conducting the interventions in class in the presence of homeroom teachers. Facilitators are either senior managers from the private sector, retirees from the private sector, or young adults volunteering in a civic service mission for one year. During the *Energie Jeunes* sessions, the facilitators present slides, play videos and follow a standardized scripted text. In order to develop children's school motivation and perceived return to effort, the program essentially focuses on three components: the growth mindset of intelligence (Dweck and Yeager, 2019), the internal locus of control (Angela L Duckworth et al., 2019) and diligence. The educational content of the program is thus based on recent research in psychology (Walton, 2014).

[5]Alan, Boneva, et al., 2019 studied the impact of a one-day teacher training in 52 Turkish primary schools in Istanbul that encourages teachers to foster students' growth mindset and goal-setting by adopting appropriate teaching practices and by delivering 12 two-hour sessions of a curriculum in class. It demonstrated a large positive impact on effort exerted on a behavioral task as well as on standardized test scores, which persisted after 2.5 years (0.23 SD gain on the math test, no effect on the verbal test).

[6]The NGO is also called *Energie Jeunes*

As a typical mindset intervention, the program conveys the message that the human brain is highly plastic, that intelligence is not fixed and that there are long-term benefits to working hard and consistently on challenging tasks. The program also emphasizes that setbacks and challenges are normal, that they constitute opportunities to learn, and that they should not be interpreted as signs of low innate abilities. According to the NGO, this optimistic and positive view of failures is critical to trigger behavioral change because students with fixed mindsets tend to believe that trying hard or asking for help signals low ability. 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).

Relatedly, locus of control is also a central concept that is highlighted throughout the four-year program. During the sessions, 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 strong external constraints. For example, one of the videos features the story of a handicapped person who became an Olympic medallist; another video features a man from a very poor family 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. Growth-mindset and internal locus of control are related constructs but the former impacts students' perceived innate ability to succeed, while the latter impacts students' perception of the situational or contextual factors that are deemed necessary to succeed.

Finally, the program provides operational tools to sustain behavioral change through increased diligence. As students become aware that they have some amount of control over their success, they need concrete tips to materialize their intentions into actions. This component of the program thus includes advice to routinize effort, increase concentration, build healthy work habits, and minimise distractions. During the sessions, students are encouraged to consider both mini-cases and their own experience and to reflect on ways to strengthen their effort and perseverance. For instance, facilitators provide ways to make effort more productive, such as keeping one's cellphone away when doing homework or being attentive in class. They also discuss potential sources of resistance or discouragement and strategies to overcome them, like being bold enough to resist peer pressure when one tries to be attentive in class. Students are also asked to make a yearly commitment during the second session that they record in their individual journal. Examples of such commitments include: stop chatting in class, improve one's math average grade, or leave one's cellphone outside the bedroom when doing homework. During the third session, students assess whether they were able to honour their commitment and analyze the reasons for their success or failure. This part of the program is designed to help

students close the intention-to-action gap and reflect on the importance of sustained effort in achieving one's goals.

For more details on the content of the program, we present the exact material used in the twelve session in Appendix H. For each grade, Figure H1 reports the messages, videos, and activities of the first session which takes place between November and January, Figure H2 the second session which takes place between January and March, and Figure H3 report the content of the third session which takes place between March and May.

2.2 Evaluation Design

With the support of the Ministry of Education, the NGO contacted schools (some of which had already implemented the program) and offered them the opportunity to participate in a randomized experiment, in which the upcoming Grade 6 cohorts of school year 2014/2015 and 2015/2016 were to be randomly assigned to a treatment or a control group. The schools willing to participate signed a contract with our research team to commit to authorizing data collection for five years (2014-2019).

Our objective was to recruit about 100 schools into the program. In September 2014, 97 middle schools, located in seven different regional school districts,[7] agreed to participate in the program. Two cohorts of students per school were included and followed throughout the duration of middle school: the cohort of students who entered Grade 6 in September 2014, and the cohort of students who entered Grade 6 in September 2015. Within each school, one cohort was assigned to the treatment group and benefited from the program during the full duration of middle school (4 years), while the other cohort, assigned to the control group, received no intervention at all. The experimental design is represented in Appendix Figure A1.

This design relies on the assumption that the risk of spillover across cohorts is small enough for our design to guarantee internal validity. Such spillovers could come from three sources, which we all consider infrequent: friends or siblings enrolled in two consecutive years; teachers who attend the program, teach students in two consecutive years, and convey the messages to control students; and students who fail a school year and get retained. Note that the proportion of students who get retained is extremely limited: 0.5% in Grades 6, 7, and 8, and 2% in Grade 9.[8] Although minimal, such spillovers would lead to an attenuation bias, i.e., run again finding a treatment effect.

On the upside, this design has four advantages: first, since the randomization is conducted within school at the cohort level, all schools included in our sample benefited from the *Energie Jeunes* program at some point in time: some schools

[7]The schools are located in the regional school districts (called *académies* in French) of Aix-Marseille, Créteil, Amiens, Lille, Lyon, Paris and Versailles

[8]Data from the Ministry of Education: MENJ-DEPP, Note d'Information 19.46, November 2019, available online at education.gouv.fr/statistiques.

in 2014 (Group A), others in 2015 (Group B). Second, cohorts who had started to benefit from the program before the experiment (i.e., before school year 2014/2015) continued to benefit from the program after the beginning of the experiment. These two features considerably facilitated schools' willingness to participate in the experiment and to collaborate with the research team on data collection. Third, this design allowed us to use within-school (and within-cohort) variations in treatment assignment, which significantly increases statistical precision compared to a school-level randomization. Fourth, cohort-level randomization considerably limits spillovers compared to a class-level or an individual randomization.

3 Data

3.1 Sampling Strategy

Schools In France, priority education schools (i.e., schools receiving additional support from the Ministry of Education) represent about 20% of middle schools, non-priority public schools 60%, and private schools the remaining 20%.[9] The *Energie Jeunes* program targets public disadvantaged middle schools, mostly in priority education. Our sample includes 97 middle schools (and 194 cohorts) at baseline (Grade 6) that volunteered to be part of the experiment in seven (out of 25[10]) regional school districts. A large majority (79.4%) of the sample is located in a priority education zone.

Appendix Table B1 presents the characteristics of the schools in our sample compared to the populations of priority education schools, public schools, and all French schools. Table B1 shows that the *Energie Jeunes* sample is slightly more advantaged than the average priority education schools, whereas much more disadvantaged than the average public schools and, *a fortiori*, the average French schools, which also include private schools. Column *EJ Sample - PE Schools* shows, for instance, that students in our sample have slightly larger classes (+0.64 students per class), are from slightly higher social and economic backgrounds (+2.6 pp high-SES) and are slightly more likely to go to an academic high school (+2.1 pp) than the average priority education students. Despite these small differences, *Energie Jeunes* schools perform similarly to the average priority education school on the national tests in Grade 6 and Grade 9. Likewise, teachers in *Energie Jeunes* schools are also fairly similar to teachers in the average priority education school. Conversely, *Energie Jeunes* schools are significantly more disadvantaged than other middle schools in France. For instance, the proportion of students from a high-SES background is half as large in *Energie Jeunes* schools as in the rest of French middle schools, the proportion of financial aid beneficiaries is twice as large, and students in *Energie*

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

[10] In metropolitan France, i.e., excluding overseas France.

Jeunes schools perform much worse on the G6 and G9 national tests (1.3-1.5 SD lower). In view of these comparisons, we conclude that the *Energie Jeunes* sample is composed of disadvantaged schools with performance strikingly lower than the rest of French middle schools. Our results are therefore likely to generalize well to the population of students in priority education areas.

Classes As described above, our study includes two cohorts of students and all the classes[11] in these two cohorts took part in the experiment, except for the special-need classes ("Segpa"), which were not targeted by the program.[12] In Grade 6, our sample contained 1,026 Grade 6 classes. In Grade 7, we tracked 995 classes, 985 classes in Grade 8, and 983 classes in Grade 9, all equally distributed between the control and the treatment groups. The slow decay in the number of classes included in our sample over time is essentially driven by the fact that a few middle schools dropped out of the experiment (see *infra*, section 4.3).[13]

Students - Administrative data We collect student administrative data in all classes each year. The resulting samples vary slightly from year to year because a few schools dropped out of our sample and also because of student exits and entries. Our full student sample is therefore a panel dataset[14] that includes between 23,000 and 24,000 students, equally distributed between the control and the treatment group (Table 1, first panel, first column).

Students - Survey data Every year, we randomly selected seven students per class to answer the student and teacher questionnaires. We conducted the random selection every year so the students' sub-samples were independent from one year to the next and representative of the full population. Both student and teacher questionnaires focused on these selected students; the selected students received the student questionnaire while their teachers received the teacher questionnaire to provide information about them.[15]

[11] In French middle schools, students are assigned to classes at the beginning of the school year. All students in the same class basically follow the same courses at the same time, except for a restrictive set of elective courses (typically languages).

[12] "Segpa" classes account for only 3% of middle school students at the national level.

[13] Grade 6: 521 treatment and 505 control classes; Grade 7: 504 treatment and 491 control classes; Grade 8: 494 treatment and 491 control classes; Grade 9: 493 treatment and 490 control classes.

[14] It is 4-year panel for students who stayed 4-years i.e., 17,566 students in Grade 9 or 74.5% of the Grade 9 sample.

[15] The way we approached the teachers varied over the study period. In Grade 6, only the teacher who was present in class when the research team visited the school received the teacher questionnaire. In Grade 5, we invited two teachers per class to answer the questionnaire online. In the following years, we distributed paper questionnaires in teachers' mailboxes and provided the school with a letter asking the school to send the responses to the research team a few weeks after the visit. Teachers who were present on the day of the visit, those who had more hours of instruction with the class, and those who were more willing to participate in research, were more likely to respond to the questionnaire. Importantly, while the selection of teachers may have varied

3.2 Data Sources

All data were collected every year from 2014-2015 to 2017-2018 for the first cohort, and from 2015-2016 to 2018-2019 for the second cohort.

The first source of data is administrative school registers, which provide students' grade point average (GPA), the yearly number of absences, lateness, sanctions[16], and disciplinary actions[17]. The administrative data also provide the students' socioeconomic characteristics, including gender, year of birth, country of birth, parental occupation, whether the parents are employed or unemployed, whether the family is one- or two-parents, and whether the family receives need-based financial aid from the State. The administrative data are available for all students in the sample.

The second source of data is a student questionnaire. The research team administered the student questionnaire on digital tablets to the sub-sample of seven randomly selected students per class. The student questionnaire comprehensively followed our theory of change, using instruments validated in the psychology literature. The first outcome of interest is perceived return to effort, captured by two components: (i) fixed *versus* growth mindset, measured using the short version of the standard instrument validated by Claro et al., 2016, as well as the questions on growth mindset used by Li and Bates, 2017 and three questions used by Guyon and Huillery, 2020; and (ii) external *versus* internal locus of control, measured using four questions capturing the perceived importance of family and social factors in academic success, developed by Guyon and Huillery, 2020. Our second outcome of interest is self-reported diligence, captured by six components: (i) Orderliness, measured using the Big Five Inventory developed by Goldberg, 1990; (ii) Grit, measured using the Short-Grit Scale developed by Angela Lee Duckworth and Quinn, 2009; (iii) School-work impulsivity, measured using the Domain-Specific Impulsivity Scale for Children developed by Tsukayama et al., 2013; (iv) Work discipline, measured using the International Personality Item Pool developed by Goldberg et al., 2006; (v) Homework management, measured using the Homework Management Scale developed by Xu and Wu, 2013; and (vi) Hours of homework, measured by the amount of time spent on doing homework in the two days before the survey. Appendix Table C1 provides the item composition of all of these measures. In addition to the non-cognitive measures, we include a measure of Grade 9 students' educational and professional aspirations. We ask students which job they aspire to and which type of high school they would prefer to enroll in (academic or technical school). Both measures capture how the *Energie Jeunes* program affects ambition, aspirations and projection into the future. Finally, every year, we measure partici-

across time, it was similar in both treatment and control classes, so any resulting bias is the same in both groups and does not affect the internal validity of the results.

[16]This mainly reflects the number of hours of detention. Sanctions for minor misconduct may also include writing an essay, confiscation of cellphones, letters to parents, etc.

[17]This includes disciplinary hearings for serious offenses—violent behavior, for instance.

pation in the program, at the very end of the student questionnaire. Note that the initial student survey also 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., 2020). 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.

The third source of data is the teacher survey, which contained questions on the same students as those taking the student survey. The teacher questionnaire includes questions on students' character—we administered a French version of the Character Report Card developed by Park et al., 2017. The Character Report Card evaluates three main dimensions: social character (predicting less peer conflict and greater popularity), intellectual character (predicting greater participation in class), and achievement character (predicting higher report card grades).[18]

Finally, we collect data from the national exam administered at the end of middle school, exactly in the last two days in Grade 9. The exam is externally and anonymously graded. Even though the exam is not required to enroll in high school, virtually all ninth graders take it (97% in our sample). The exam includes tasks in maths, French, history-geography, sciences, and an oral examination based on a personal project.[19]

3.3 Outcomes of Interest

3.3.1 Take-Up and Adherence

We use four measures of participation, all collected yearly at the end of the student questionnaire: one binary variable 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 during the school year (in theory it should be three), whether or not the student made a commitment, and whether or not she honored her commitment.

[18]In 2018 and 2019, we also implemented an Information Session experiment targeting only ninth graders. In each school, the research team invited ninth graders to an information session providing advice on how to be well-organized and get support for life in high school. The objective was to use the participation in this information session as a real-life behavioral measure of students' motivation and grit. However, only 33% of control cohorts and 45% of treated cohorts accepted the invitation to the Information Session experiment. As a result, we do not report the results in this paper but we can provide them upon request.

[19]The students receive a grade that aggregate both the exam score and continuous assessment based on GPA in Grade 9. Since we already measure GPA separately, we only use the exam scores here.

3.3.2 Perceived Return to Effort

The immediate objective of the program is to increase perceived return to effort among the students. In order to avoid inference issues due to multiple hypothesis testing, we favor summary indices. We construct a summary index of perceived return to effort combining all questions related to the growth mindset of intelligence, which shows the extent to which a student perceives that her cognitive ability and performance can grow through dedicated effort, and all questions related to the internal locus of control, which shows perceived importance attributed to socioeconomic constraints in academic success. The index is constructed as a weighted mean of the standardized items listed in Appendix Table C1. 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 items that are highly correlated with each other receive less weight, while items that are uncorrelated and thus represent new information receive more weight. We use the same aggregation method for all indices.

To better understand the potential channels of impacts, we also use two separate sub-indices: a summary index of growth mindset on the one hand, and a summary index of internal locus of control on the other hand, constructed following the same methodology. Note that the growth-mindset and locus of control questions were not included in the questionnaire administered to sixth graders, so we constructed the indices from Grade 7 on.

3.3.3 Self-reported Diligence

As the intervention is meant to develop motivation and effort to achieve short-term and long-term goals, we hypothesize that students will increase their own diligence, i.e., develop steady application, careful work involving long-term effort, conscientiousness, determination, and perseverance. We construct a summary index aggregating all items reflecting diligence as perceived by students themselves: self-perceived orderliness, grit, school-work impulsivity, work discipline, homework management[20], and hours of homework[21]. The Self-reported diligence index is constructed as a weighted mean of all the standardized items listed in Appendix Table C1, following the same methodology as described above.

To better understand the potential channels of impacts, we also used separate indices for self-perceived orderliness, grit, school-work impulsivity, work discipline, homework management, and hours of homework.

[20]Not included in the Grade 6 index as this sub-index was not collected then.
[21]Not included in the Grade 6 index as this sub-index was not collected then.

3.3.4 Teacher-reported Character

Teachers' view on students' character is crucial for our study, as it provides a third-party evaluation. In a context in which the intervention may affect self-perception, a third-party evaluation has the advantage of not being affected by the intervention.[22] 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 C1).

To better understand the potential channels of impacts, we also used separate indices for the three main factors measured by the Character Report Card: achievement character, intellectual character, social character (Park et al., 2017). Items loading on social character included items from the gratitude, optimism, social intelligence, and interpersonal self-control scales. Items loading on intellectual character included items from the zest and curiosity scales. Items loading on achievement character included items from the grit, optimism, curiosity, and schoolwork self-control scales.

3.3.5 School-reported Behavior

If the intervention can change perceived return to effort and character, we hypothesize that students will be more likely to respect school rules. We used a summary index of the respect of school rules aggregating information from school administrative registers: 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 (e.g., expulsion from class or detention), and number of disciplinary actions (e.g., temporary or permanent expulsion from the school) over the school year. Sanctions are given by the teachers, school supervisors or yard duty monitors when students are disrespectful or disruptive, while disciplinary actions are taken by a disciplinary board for more severe misconducts such as violence. Our school-reported behavior summary index was constructed using the same methodology as described above.

Compared to the teacher reported character index, this index provides a different and complementary measure of behavior: first, it is a more objective measure as it is purged of potential declaration biases from the teachers[23]. Second, it captures violations of important school rules in and out of the classroom, whereas teacher-reported character captures finer variations in work attitude in class.

[22]The fact that some treatment teachers attended the Energie Jeunes sessions is unlikely to affect the reliability of this measure in our opinion. First, teachers who attended the session do not directly participate in the intervention. Second, it would be a cause of concern only if teachers would modify their perception of their treated students: It would require them to remember which of their students received the program when we survey them (i.e. at the end of the school year), which seems very unlikely. Last, the respondent teachers are not necessarily the ones who responded to our teacher questionnaire: since middle school students have in average nine teachers every years, any declaration bias would be highly diluted.

[23]although we do not believe this is a major issue, see footnote 22

3.3.6 Educational and Professional Aspirations

In France, there are two main types of high schools: technical high schools (including a two-year track and a three-year track)[24] and academic high schools[25]. Because students in academic high schools have higher average GPA, more years of higher education, and higher wages on the job market (Guyon and Huillery, 2020), we use a dummy equal to 1 if the student aspires to go to an academic high school to indicate a "better" outcome, *versus* 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 aspirations at the end of ninth grade . We coded the answers using the National Institute for Statistics and Economic Studies (INSEE) job classification. INSEE assigns each job to one of the following categories: farmers, craftsmen and storekeepers, manual labourers, low-skilled office workers, intermediate occupations, and high-skilled occupations. We then constructed three dummies corresponding to increasing levels of job qualification: a dummy equal to 1 if the student aspires to a job in the farmers, craftsmen and storekeepers[26], manual labourers or low-skilled office workers categories ("low-skilled job"); a dummy equal to 1 if the student aspires to a job in the intermediate occupations category, e.g., nurse, primary school teacher, or accounting officer ("medium-skilled job"); and a dummy equal to 1 if the student aspires to a high-skilled occupation, e.g., lawyer, doctor, journalist, or computer programmer ("high-skilled job"). Note that students who aspire 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 fourth category for the students who answered that they do not know ("no aspiration").

Finally, we created a synthetic score of aspirations combining two components: a dummy indicating that the student aspires to an academic high school, and a dummy indicating that the student aspires to a medium- or high-skilled job. The aspiration summary index was constructed based on these two dummies using the same methodology as described above.

3.3.7 Academic Performance

The ultimate objective of the intervention is to increase academic performance. We used the GPA—average grades from all major academic courses— collected each year for all students. In France, teachers use a 0-20 point scale and grade the students based on predefined expected competences.This grading system contrasts with systems in other countries where students are graded *on the curve* using percentiles or a predefined distribution. We aggregated and standardized the GPA score using

[24]Here we group together "lycée professionnel", "CAP" and "apprentissage".

[25]In French: "lycée général et technologique".

[26]Except students who aspire to become entrepreneurs, whom we assigned to the high-skilled job category.

the same methodology as described above. GPA are internally graded by the middle school teachers.

In addition to GPA, we used standardized scores at the national end-of-middle school exam, which is anonymous and externally graded, to confirm that the increase in GPA is not due to a change in teacher grading. Both indicators measure academic performance and complement each other. On the one hand, GPA is our preferred measure because it is the key parameter used by the education system to assign students to high school. GPA is also a more comprehensive measure of students' performance as it averages all the grades received in all subjects over the year whereas the end-of-middle school exam is a one shot measure, hence noisier. On the other hand, the end-of-middle school exam provides an objective complementary measure of academic performance purged from any potential grading biases due to students' improved behavior, which is important to adequately interpret our findings.

4 Validation of the Experimental Protocol

In this section, we verify that the experimental protocol is successfully implemented and derive the estimation strategy.

4.1 Balance Checks

The validity of the experiment protocol is based on the pre-treatment comparability of the treatment and control students. We present balance checks in Appendix D for each survey year. Appendix Table D1 concerns the administrative data sample, Appendix Table D2 the student survey sample, and Appendix Table D3 the teacher survey sample. Most characteristics are well balanced and we do not see any consistent differences across grades between the treatment and the control groups. We detect some significant but small differences for some samples and some grades, which is expected due random variations. Also, our high detection power makes small differences significant. The only consistent difference is the proportion of blue collar families in the administrative data sample, which is 1 percentage point lower in the treatment than in the control group throughout. Overall, the differences are small and represented a small proportion of the tested differences in means.. In the remainder of the paper, we thus favor the average differences between the treatment and control groups without covariates. Yet, as a robustness test, we conduct an analysis using the baseline characteristics presented in the balancing Table (see Table D4).

4.2 Student Entries and Exits

Apart from data collection issues, our sample is subjects to students entering and exiting the surveyed schools. Since our sample is always balanced in all survey rounds, these entries and exists do not appear to have affected the validity of the randomization (see *supra*). Table 1 describes the effects of exists and entries on our sample. First, every year, new students registered in our schools after Grade 6 (entries). Mechanically, those who registered in the treated cohort did not receive the full intervention but only part of it. Nevertheless, we include all new students in the intention-to-treat analysis. As shown in the last panel of Table 1 (*Entry rate*), students who entered the sample after Grade 6[27] represent, in the control group, 7% of Grade 6 students in Grade 7, 16.5% in Grade 8, and 21.8% in Grade 9. As shown in Table 1, these proportions are very well balanced between treatment and control groups. Second, some students changed schools over the study period (exits). As we could not collect outcomes for these students, they are not included in the study sample. Students who left the sample after Grade 6 represent, in the control group, 11.8% of Grade 6 students in Grade 7, 15.1% in Grade 8, and 19.8% in Grade 9.[28] As shown in Table 1 (*Exits rate*), these proportions are fairly similar to the entry rates and are very well balanced between treatment and control groups. The fact that the rate of entries and exists are very similar in the treatment and control groups, together with the fact that our data are balanced in each survey round (see Table D1 D2 and D3), confirms that entries and exists did not modify the validity of the original random assignment.

4.3 Survey Attrition

In addition to students entering and exiting, our data collection suffer from regular survey attrition. Regarding administrative data, we had no attrition in 2015 and 2016. Two schools refused to let us collect administrative data in 2017, four schools in 2018, and four schools in 2019. Hence, attrition rates for the administrative data are 0 among sixth and seventh graders, 2.5% among eighth graders, and 4,5% among ninth graders. Attrition remained minimal and balanced across the control and treatment groups, as shown in Table 1, first four columns.

Regarding student survey data, a few schools refused to administer the student questionnaire due to the logistical burden it imposed on them. This did not happen

[27] we considered as a new entry only students who entered in Grade 5 or later. Students who enter during the sixth Grade's school year are therefore not considered as new entries.

[28] While we observe the students who enter our sample, we do not perfectly observe which students exit our sample. We therefore infer the exists using the sample sizes of two consecutive survey rounds and the known number of entries. This leads to some measurement errors (for instance if a school did not report entry date, or entry dates are inconsistent) forcing us to drop some schools from the analysis of exists. The rates are still comparable between treatment and control however.

in 2015, but it happened in one school in 2016, in four schools in 2017, in nine schools in 2018, and in 10 schools in 2019. In addition, we failed to administer the questionnaire in a few classes due to organizational issues (e.g., one teacher was absent so students left school early, before the time of the survey), although this remained very marginal. As a consequence, attrition rates for the student survey increased from 5.9% among sixth graders to 21.2% among ninth graders (Table 1, second four columns). Again, attrition was balanced across the treatment and control groups so it did not affect the internal validity of our results. Moreover, the comparison between Appendix Table D1 and Appendix Table D2 showed that the student survey sample is very similar to the full sample, i.e., attrition at the student survey did not affect the external validity of the results either.

Regarding teacher survey data, attrition rates were higher, from 22.9% among sixth graders to 38.6% for ninth graders. This is due to the difficulty of finding teachers available to answer the questionnaire when our research assistants were on site. However, attrition rates were statistically similar across the treatment and control groups, and the characteristics of students in the teacher survey sample are similar to those of students in the full sample (see Appendix Tables D1 and D3). Therefore, attrition affects the precision of the estimates, but not their internal and external validity.

Finally, attrition rates for the national exam test scores was 15.7%, not significantly different across the control and treatment groups (Table 1). Appendix Tables D1 (column National exam sample) also shows how balanced the sample of ninth graders with valid national exam results is. In this sample, the treatment group had 1 percentage point fewer blue collar families (as was the case for all the administrative data in this study), and 1 percentage point more girls than the control group. Overall the treatment and control groups are thus very similar.

4.4 Estimation Strategy

We use intention-to-treat estimates, meaning that data were analyzed for all students enrolled in a school-cohort randomized to an experimental condition and whose outcome data could be collected. As noted above, 25% of Grade 9 students registered schools after Grade 6 and received only partial treatment. Note that a treatment-on-the-treated analysis yields the same conclusions but produces larger effect sizes.

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

$$Y_{iscj} = \alpha_j + \beta_j T_{sc} + \theta_s + \theta_c + \epsilon_{iscj} \quad (1)$$

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, 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, the small decay being the consequence of the school-level attrition already discussed.

As intended in our registered pre-analysis plan,[29] we test heterogeneous effects according to gender, socioeconomic status, baseline academic performance, and baseline school behavior. For each characteristic, we add a dummy indicating the subgroup in the model, as well as the interaction between this dummy and the treatment dummy and with school and cohort fixed effects. The dummies defining the subgroups are: whether the student is a female, whether she does not receive financial aid (hereafter, "non-recipients" *versus* "aid recipients"), whether her academic performance at baseline was above the median[30] (hereafter, "high-achievers" *versus* "low achievers"), and whether her school behavior at baseline was above the median[31] (hereafter, "well-behaved students" *versus* "poorly-behaved students"). Note that the number of observations decreases when we analyze impact heterogeneity depending on baseline academic performance and baseline behavior, because the sample is restricted to students enrolled in the school in Grade 6 for whom baseline measures are available. The correlations between the characteristics defining the subgroups are: 0.03 between girls and aid recipients, 0.08 between girls and high-achievers, 0.12 between girls and well-behaved students, 0.15 between non-recipients and high-achievers, 0.15 between non-recipients and well-behaved students, and 0.27 between high-achievers and well-behaved students. These correlations are small enough so that each subgroup analysis captures a different factor of heterogeneity.

Finally, we check the robustness of the results to multiple hypothesis testing issues by pooling all grades in the estimation (hereafter, "stacked" sample). The estimated β is the average intention-to-treat effect throughout middle school, i.e., the difference between a student in a treatment cohort and a student from the same school and grade but in the control cohort, whatever their grade. This specification has the advantage of reducing the risk of spurious effects due to multiple hypothesis testing and it mechanically increases statistical power since the dataset is now made of about four times more clusters. The drawback is that it assumes the same treatment effect whatever the grade, while one could expect a cumulative effect of the treatment, i.e., growing impacts over time. In what follows, we favor the grade-by-grade analysis, which informs us about the dynamics of impacts, and use

[29] See AEA RCT Registry Number AEARCTR-0000376.

[30] Academic performance at baseline is GPA during the first three months in Grade 6.

[31] School behavior at baseline is a summary index based on z-scores of the following variables measured in the three first months in Grade 6: the number of half-day absences, the number of days the student was late, the number of sanctions, and the number of disciplinary actions.

the stacked sample as a robustness check and a synthetic view of impact heterogeneity. We check the robustness of our results grade-by-grade by controlling for baseline imbalance. We use all baseline characteristics presented in Table D1- D1 and re-estimate our results in Table D4. The results are slightly reduced by about 0.01 SD but are not fundamentally different from the specification without control variables, we therefore keep the results without control variables as our preferred specification.[32]

4.5 Participation and Adherence to the treatment

Before examining the impacts, we first present our measures of the participation in the treatment. Table 2 shows that there are large differences in take-up between the treatment and the control groups. Based on survey data, participation among control students was between 3.6% and 18.7%, which may be due to mistakes in program implementation or to students misinterpreting the question. In the treatment group, 83-95% of students declared that they participated in at least one session, representing a 73-81 percentage point increase. Students in the treatment group report having attended 2.5 more sessions than 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: while 75% of sixth graders made a commitment, it falls to 54% of the ninth graders. The proportion of students who declared that they had honored their commitment is lower and decreases over time, from 54% of sixth graders down to 35% of ninth graders. Although these statistics are purely descriptive and should not be taken at face value, it seems that a majority of students played the game (but do not always succeed at it).

Tables E1 and E2 show participation by subgroup using the stacked sample. Overall, girls are 4.2 pp more likely than boys to take up the intervention, aid recipients are 4.5 pp less likely to participate than non-recipients, high-achievers are 7.3 pp more likely to participate than low-achievers, and well-behaved students are 4.5 pp more likely to participate than poorly-behaved students. The take-up is thus positively correlated with predictors of better school outcomes. Finally, older students seem more difficult to engage in the program than younger ones. One explanation for these findings may be that older students, boys, aid recipients, low-achievers, and poorly-behaved students are more likely to be absent and miss parts of the intervention. They may also be less interested in school activities in general.

[32]Note that controlling for a finite number of unbalanced characteristics is controversial Athey and Imbens (2017) and sometimes not recommended Freedman (2008). Since the imbalances are of small magnitude, we favor our original specification with adjustment for school and cohort fixed effects only.

5 Impact on Academic Performance

5.1 Grade Point Average

Table 3 shows the impact of the program on academic achievement in each grade. The four-year intervention increased GPA by 0.07 SD in Grade 9, which represents 5% of the Priority Education *versus* national average achievement gap.[33] With an average differential take-up rate evaluated at 79% in Grade 9 (See Table **??**), the local average treatment effect is about 27% higher i.e. +0.09 SD[34]. The impact on GPA in previous grades is smaller, but still significant: 0.03 SD in Grade 6, 0.04 SD in Grade 7, and 0.03 SD again in Grade 8. As shown in Appendix Figure G1, GPA in Grade 9 is 0.77 SD lower than in Grade 6. The program thus reduced the natural decreasing trend in GPA over the course of middle school by about 10%.

Figure 1 provides interesting insights on the dynamics of the impact by subgroups (corresponding point estimates are reported in Tables 4 and 5). The impact is immediately positive and significant for girls (0.05 SD, 0.07 SD, 0.06 SD and 0.08 SD in Grades 6-9 respectively), as well as for aid non-recipients (0.05 SD, 0.04 SD, 0.03 SD and 0.08 SD in Grades 6-9 respectively). The impact for well-behaved students is especially large in every grade (0.08 SD, 0.07 SD, 0.12 SD and 0.10 SD in Grades 6-9 respectively). In contrast, we find no significant impact from Grade 6 to Grade 8 on boys, aid recipients, and poorly-behaved students. The impact only materializes three years after the beginning of the program (in Grade 9) for males (+0.05 SD), aid-recipients (+0.04 SD), low-achievers (+0.05 SD) and to a lesser extent poorly-behaved students (non significant +0.03 SD). Tables E1 and E2 provide heterogeneity analysis using the stacked sample to protect against multi-hypothesis testing issues. It confirms that the impact on GPA is similar for low- and high-achievers, but is significantly differential according to gender, financial aid status and baseline behavior: the overall difference in impact is 0.05 SD between girls and boys, 0.03 SD between aid recipients and non-recipients students, and 0.07 SD between well- and poorly-behaved students. The subgroup differences in impact and its dynamics indicate that some students are immediately sensitive to the messages conveyed by *Energie Jeunes*, while other groups are more resistant at first but eventually find benefits to the intervention.

[33] According to Table B1, the achievement gap between Priority Education schools and the national average in Grade 9 is equal to 1.4 SD.

[34] This is the LATE of participating in at least one session of Energie Jeunes over the four years of college. The effect to participating in the whole program (12 sessions over four years) is likely larger. Yet, since most of our impacts materialize during the first year of intervention (in Grade 6), we do not expect the LATE of participating in all sessions to be much larger than 0.09 SD.

5.2 National end-of-middle school exam

One could be worried that the GPA results only reflects an Hawthorn effect whereby the teachers of treated students voluntarily overrate students' performance in response to the awareness of being observed or treated (through for instance grading leniency, giving easier assignments or rewarding better class behavior). We believe this is not the case for two main reasons. First, teachers had no vested interest in the Energie Jeunes program. The program only took place a couple of hours a year and only one teacher actually attended the sessions. The majority of teachers had no idea about the program and they did not know whether their classes were in the treatment or control group. Second, our results at the national end-of-middle school exam suggest the contrary. The national end-of-middle school exam is anonymous and externally graded and therefore captures improved academic performances more objectively than GPA does. While GPA may incorporate behavioral aspects (i.e., being late on submitting a home assignment or grading leniency for well-behaved students), the national exam better isolates the learning component of GPA. On the downside, national end-of-middle school exam is low-stake, which does not affect the academic trajectory as assignments to high schools is determined before the exam. Conversely to the national exam, GPA is a key measure of academic performance in France and is taken into account by the education system to determine academic trajectories. In that sense, our results on GPA are of primary importance. We therefore consider the national exam results as an independent measure of academic performance and a tool to confirm that the GPA impacts are not driven by teachers' grading behavior.

We report the results at the national exam in Table 3. It shows that treatment students scored higher at the exam (+0.03 SD, significant at 10 %). Figure 6, Table 8, and Table 9 show that girls, well-behaved students, and non-aid recipients gained more (respectively +0.05, +0.06 and +0.07 SD) and that there is no significant impact on boys, aid recipients, and poorly-behaved students—the difference between subgroups being significant for gender, financial aid and baseline behavior. These results show the exact same patterns as the one observed for GPA, which rules out the possibility that the effect on GPA comes from a change in teacher grading. Tables 8 and 9 also show the impact by discipline: treatment students received better grades at the French exam (+0.05 SD), the history-geography exams (+0.03 SD) and at the oral exam (+0.03 SD), but not in maths or biology.

Overall, the national end-of-middle school exam confirms that the intervention improved students' academic performance. Since our GPA results are concomitant with improved behaviors in class and at school (see Sections 6.3 and 6.4 below), the result on the national exam also rules out the possibility that the impact on GPA is solely due to teachers rewarding students who exhibit better behavior in class. The positive impacts on the national exam shows that the program improved aca-

demic performance without ambiguity. This is an important finding since academic performance is a tangible measure of human capital accumulation.

5.3 Magnitude and Cost-effectiveness

Are the magnitudes of these impacts substantively meaningful? As mentioned in the introduction, large-scale education interventions in high-income countries often fail or have fairly small effects. Cheung and R. E. 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 effects decrease when smaller targeted programs are taken to scale (R. Slavin and Smith, 2009). Impressive effects from small and non-representative samples often fail to replicate when programs are expanded to larger and more representative populations. 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* is at the 50^{th} percentile of the distribution of effect sizes if we consider scores at the national exam (0.03 SD), and at the 70^{th} percentile if we consider GPA scores (0.07 SD). But *Energie Jeunes* is much more cost-effective than the typical intervention: while the average cost of programs at the 50th percentile is $882 per pupil, *Energie Jeunes* is only €65 (approximately $75) per pupil, hence more than ten times cheaper. The fact that this experiment was conducted on a large number of students ($\approx$ 24,000) and on schools that are fairly representative of the population of priority education schools in France confers more importance to our results.

6 Mechanisms

In this section, we explore the mechanisms through which the intervention may have increased GPA. We follow our pre-registered theory of change to show the associated impacts, but we do not claim that any particular mechanism explains the impact on GPA as these are only associations. We also consider these impacts as outcomes in their own right, as they may improve students' welfare and translate into better long-term life outcomes.

6.1 Perceived return to effort

Table 3 shows positive impacts of the program on perceived return to effort in each grade for which the index is available, ranging from 0.04 to 0.05 SD. Students are

generally prone to update their perceptions and beliefs regarding their chances of success at school. Even though every subgroups' perceived return is positively affected, the malleability of perceptions varies according to gender, socioeconomic background, and baseline behavior. Figure 2 and Table 4 show that girls consistently respond more than boys to the intervention. The stacked sample analysis in Appendix Table E1 confirms that girls are more responsive overall than boys in terms of changing their perceptions; the overall effect in middle school is 0.02 SD for boys and 0.06 SD for girls (the difference between boys and girls is significant at the 1% level). Similarly, Figure 2 shows that aid non-recipients' and well-behaved students' mindset responds more than their counterparts' mindset (see the corresponding coefficients in Tables 4 and 5). Appendix Tables E1 and E2 show that when we stack all grades together, we find effect sizes approximately twice as large for non-recipients and well-behaved students as for aid recipients and poorly-behaved students respectively. Note that, in contrast, the malleability of perceptions does not depend much on baseline academic performance (Table 5, and Appendix Tables E2 and F3).

As shown in Appendix Figure G2, perceived return to effort slightly increases over time in adolescents (significant +3.2 % SD over three years), i.e., adolescents naturally become slightly more aware of the role of their own effort relative to fate or external circumstances. Interestingly,aid non-recipients in the control group tend to perceive lower returns to effort than their counterparts, and the program was able to reduce this gap. Conversely, girls in the control group tend to perceive higher returns to effort than boys, so the program increased this gap even further. Except for girls, it is interesting that impacts on mindset are larger for students whose baseline growth mindset and internal locus of control are lower.

The sub-indices analysis in Appendix Tables F1-F4 shows that the positive impacts concern both the growth mindset of intelligence and the internal locus of control. We see a clear association between impacts on GPA and impacts on growth mindset: the students who reacted the most in terms of GPA—girls, non-recipients, and well-behaved students – are the ones reacting significantly more to our growth mindset measures. The impacts on the locus of control seem also stronger for aid non-recipients and well-behaved students, although the estimates on the interaction term are less precise and not significant. This suggests that perceived return to effort is a possible pathway to explain the impact on GPA. Yet, the fact that boys, aid-recipients, initially lower GPA and poorly-behaved students also improved their perceived return while did not experience GPA gains (except in Grade 9) raise the concern that increasing perceived return to effort may not be sufficient to induce behavioral changes for more fragile students.

6.2 Self-reported diligence

Table 3 shows null impacts of the intervention on self-reported diligence, and even a small negative impact in Grade 7. The subgroup analysis in Tables 4 and 5 shows that the negative impact in Grade 7 concerns mostly girls and poorly-behaved students (see Figure 3). The differential response according to gender and baseline behavior is confirmed when we stack all grades together in Tables E1 and E2, so it is unlikely a spurious effect.

Examining the components of the self-reported diligence score provides interesting insights on what happened regarding self-assessment. Results are reported in Appendix Tables F1-F4, in which we use the stacked dataset for the sake of simplicity. For all students except well-behaved ones, the intervention had an overall negative impact on self-reported grit, of -0.02 to -0.05 SD depending on the subgroup. We also find a negative impact on self-reported homework management and self-reported work discipline for aid recipients and for high-achievers (Appendix Tables F2 and F3). The intervention thus led students to revise their diligence in general and grit in particular downwards, with the notable exception of the well-behaved.

How should we interpret these negative effects on diligence? One possible explanation may be that these effects reflect a true deterioration in students' behavior due to possible reactance to the *Energie Jeunes* messages. Students who feel heavily pressured to adopt a diligent attitude may resist and strengthen the opposite attitude. An alternative explanation is that these messages attracted students' attention to their own diligence deficiencies, changing the reference point against which they compare their own character. Students may have realized that they lack the combination of tenacity and passion in school work, which is a strong focus of the intervention. They may have become more self-critical in their assessment of how well they concentrate and refrain from distraction when they do homework, which is another strong focus of the program. The intervention may thus have worked as a developing path of actual deficiencies in diligence. To disentangle actual changes in diligence *versus* changes in reference point, it is crucial to cross-reference the views of students and teachers, as we do below.

Finally, we see no impact of the intervention on self-reported time spent doing homework, which holds for all groups of students (see last component of self-reported diligence in Appendix Tables F1-F4). This means that the change in perceived return to effort did not motivate students to work more (extensive margin). This result comes as a surprise since, as noted above, we do not see improvement in the intensive margin either (i.e. grit, concentration, discipline, orderliness), at least in the view of students themselves. Thanks to the data collected from teachers and schools, we will show that the action goes through the intensive margin with improved attitudes in class and behavior at school.

6.3 Teacher-reported Character

Table 3 shows that the intervention had a positive impact on student character as perceived by their teachers. The effect is 0.03 SD (non-significant) in Grade 6, 0.05 SD in Grade 7, 0.04 SD in Grade 8, and 0.02 SD (non-significant) in Grade 9. When we stack all grades together to avoid multiple hypothesis testing, we find an overall significant effect of 0.04 SD (Appendix Table E1).

Figure 4 shows the impacts by subgroups (corresponding coefficients are reported in Tables 4 and 5). The picture that emerges is that all groups improved in character sometime during the course of middle school, except low-achieving students. The largest point estimates are found in well-behaved students, although the difference between well-behaved and poorly-behaved is not always significant.

Looking at sub-indices in Appendix Tables F1-F4 reveals that boys and poorly-behaved students tend to gain more in social character (i.e., gratitude, optimism, social intelligence, and interpersonal self-control), while girls gain more in intelligence character (zest and curiosity) and in achievement character (grit, optimism, curiosity, and schoolwork self-control). Low-achieving students do not gain in any dimension of character, whereas high-achieving and well-behaved students gain in all dimensions of character, with effect sizes ranging between 0.05 SD and 0.08 SD. Note that, in the view of teachers, students' character deteriorates over the course of middle school, especially for boys in Grade 9 (Appendix Figure G4). The intervention thus moderated the natural deterioration of students' character, in particular for boys in Grade 9, when the deterioration is more pronounced.

These results provide key information to interpret the negative impact of the intervention on poorly-behaved students' and girls' self-reported diligence noted in the previous section. The program led poorly-behaved students and girls to revise their own diligence downwards, but their actual behavior did not worsen—rather, it improved, according to teachers. The program changed the reference point against which students compare their own character, leading some students to be more severe in their self-assessment. It is interesting that girls and poorly-behaved students are more sensitive to this awareness effect than their counterparts. These results also indicate that the association teacher-reported behavior and GPA is not clearly established: sometimes both outcomes do increase in parallel (in G7 and G8 on the full sample and in G9 for boys), sometimes they do not (like in G6 and G9 on the full sample). This suggests that GPA and teacher-reported behavior, while evaluated by the same teachers, capture different dimensions of school performance.

6.4 School-reported Behavior

In contrast to teacher-reported character which measures work attitude in class, school-reported behavior measures violations of major rules such as being absent,

late, disrespectful, insolent, or violent. Table 3 shows that the intervention had a positive impact on these rule violations once the whole intervention was delivered. Surprisingly, the initial response of the students was negative in Grade 6, then null in Grades 7 and 8, and finally positive in Grade 9. These results suggest some initial reactance in younger students, which dissipates quickly in Grade 7, but it takes as long as four years for the intervention to eventually reduce major discipline issues.

Figure 5 (and corresponding Tables 4-5) provides insightful clarifications. First, groups who benefit more are those whose baseline behavior at school is worse: low-achievers and poorly-behaved students, whose score is approximately 0.3-0.35 SD below the score of high-achievers and well-behaved students, benefited a lot from the interventions (+0.06 SD and +0.07 SD respectively). The impact on school behavior varies less according to gender and financial aid status, as does school behavior itself. By improving low-achievers' and poorly-behaved students' discipline, the intervention thus reduced the behavior gap between these students by about 20%.

Second, groups whose behavior improved the most in Grade 9 are those who counter-reacted the most in the first place in Grade 6: negative coefficients in Grade 6 are stronger in boys, aid recipients, low-achievers and poorly-behaved, the same students who show the largest improvement in Grade 9 (+0.05 SD, +0.05 SD, +0.06 SD and +0.07 SD respectively). The good news is that reactance is not always fatal and does not preclude later behavioral improvement—which takes time. This result highlights the value of a four-year intervention to assess the subtle mechanisms of such program.

Appendix Figure G5 shows that school-reported behavior worsens over the course of middle school for all subgroups. Violations of major school rules such as absences, insolence, and disrespect, deteriorate substantially in Grade 9. It is important that the intervention worked as a moderator of this natural deterioration, in particular among students whose behavior is more problematic to start with.

Overall, impacts on discipline are not fully aligned with impacts on GPA, as they seem stronger in different subgroups, and not always concomitant.[35] However, increases in GPA become larger and more often significant for boys, aid recipients, low-achievers and poorly-behaved students in Grade 9, exactly when improvements in behavior materialize. We can thus speculate that, for less-disciplined students, a reduction in major discipline issues is a prerequisite for academic achievement. Our paper only shows associations between impacts on different outcomes, so the evidence is only suggestive.

[35]The fact that behavior and GPA are concomitant could also be interpreted as teachers rewarding students who improved their behavior (see 5.2). The fact that we do find positive and consistent impacts on the national exam shows that this cannot be the only mechanisms. The program did have real impacts on academic performances.

6.5 Aspirations

We find a positive impact of *Energie Jeunes* on students' aspirations, both educational and professional. As shown in Table 3, the aspiration score is 0.04 SD higher in the treatment group compared to the control group, which means that the intervention increased aspirations in the direction of higher degrees and skills.

Table 6 shows the detailed impacts by education and job categories. We find a 2 pp decrease in the proportion of students who aspire to technical high school, and a corresponding increase in the proportion of students who aspire to academic high school (although the latter estimate is not significant at conventional levels). Regarding jobs, we find a 2.2 pp decrease in the proportion of students who aspire to a low-skilled job, and a corresponding increase in the proportion of students who aspire to a medium-skilled job. Since the effects on having no aspirations and high-skilled job aspirations are both close to zero, it is likely that job aspirations moved from low to medium-skilled.[36]. Note that these two impacts are internally consistent since academic high school is meant to lead to higher education, and medium-skilled jobs typically require higher education while low-skilled jobs do not.

Tables 6 and 7 also provide the impacts by subgroup. The impacts appear heterogeneous along all dimensions. The students whose aspiration score responds most to the intervention are girls (+0.08 SD), aid recipients (+0.06 SD), low-achievers (+0.11 SD), and well-behaved students (+0.08 SD). We find no impact on the aspirations of their counterparts. For the four subgroups whose aspirations respond, the patterns are similar: aspiration to technical high school decreases by 3 to 7 pp, mostly in favor of academic high school, and aspiration to low-skilled jobs decreases by 4 to 6 pp, mostly in favor of medium-skilled jobs. For girls, we see a mix of medium- and high-skilled jobs, and for well-behaved students, we see an increase in the proportion of students who do not know what job they aspire to[37].

The picture that emerges regarding aspirations provides clear evidence that the intervention led some students to increase their academic and professional ambition. Regarding girls and well-behaved students, this effect is in line with quite large and sustained impacts on GPA. The GPA gains may be the cause of higher aspirations, or vice versa. Regarding aid recipients and low-achievers, gains in GPA are more modest but still positive and significant in Grade 9; again, both impacts are consistent and may work in both directions. Since higher aspirations are predictors of greater academic improvement in the future (Guyon and Huillery, 2020), the positive impact on aspirations can be viewed as a positive outcome *per se*, which may

[36]To confirm our result, we use a multinomial logit regression model (results not shown here). Using low-skilled job aspiration as the base category, we find that the relative risk ratio of being in the medium-skilled job category increases by 18% significant at 1% while the effect on other categories (no aspiration or high-skilled aspirations) are positive but not significant. This confirms our finding that aspirations moved slightly from low to medium skilled jobs.

[37]Note that in the aspiration score, the "I do not know which job I aspire to" category is loaded negatively.

encourage more effort and therefore higher school outcomes later on.

7 Conclusion

This paper presents the impacts of a large-scale mindset intervention in disadvantaged schools. The intervention was efficient at developing a growth mindset, at increasing students' an internal locus of control, and improving school-related behavior. The intervention has an impact on academic and professional aspirations, GPA, and students' scores at a national exam. The impacts take more time to materialize for students whose baseline behavior and performance is lower, such as boys, aid recipients, and poorly-behaved students. Well-behaved students, non-recipients, and girls change their mindset more profoundly in response to the program, which may explain why they benefit more.

More generally, this paper shows that mindset interventions such as *Energie Jeunes* 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 (e.g., 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). Most of these papers mention negative effects due to stigmatization of beneficiary students (or schools) as an explanation for the absence of impact. These policies may actually activate stereotypes and reinforce self-confidence issues in students, as well as pessimistic anticipations in teachers and parents. 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 program runs for all four years of middle school, with an intensity that is higher than that used in prior studies. The fact that the effects of the program materialize quickly (Grade 6) and plateau despite sustained exposure also raises doubt that multiple low-cost mindset interventions would produce much larger treatment effects. It rather implies that involving other actors such as parents or teachers may be necessary to transform educational outcomes more substantively.

References

Alan, Sule, Teodora Boneva, and Seda Ertac (Feb. 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.

Alan, Sule and Seda Ertac (2018). "Fostering Patience in the Classroom: Results from Randomized Educational Intervention". In: *Journal of Political Economy* 126.5, pp. 1865–1911.

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 Lee Duckworth, James 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 Guido W Imbens (2017). "The econometrics of randomized experiments". In: *Handbook of economic field experiments*. Vol. 1. Elsevier, pp. 73–140.

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.

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, Melusine, Adrien Bouguen, Axelle Charpentier, Yann Algan, Elise Huillery, and Coralie Chevallier (July 2020). *Measuring non-cognitive skills in schools: simple questionnaires outperform behavioral tasks*.

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.

Cohen, Geoffrey L., Julio Garcia, Valerie Purdie-Vaughns, Nancy Apfel, and Patricia Brzustoski (2009). "Recursive Processes in Self-Affirmation: Intervening to Close the Minority Achievement Gap". In: *Science* 324, pp. 400–403.

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

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

Freedman, David A (2008). "On regression adjustments to experimental data". In: *Advances in Applied Mathematics* 40.2, pp. 180–193.

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.

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 (2020). "Biased aspirations and social inequality at school: Evidence from french teenagers". In: *The Economic Journal* forthcoming.

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.

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].*

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

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, 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.

Figure 1: Impact on GPA

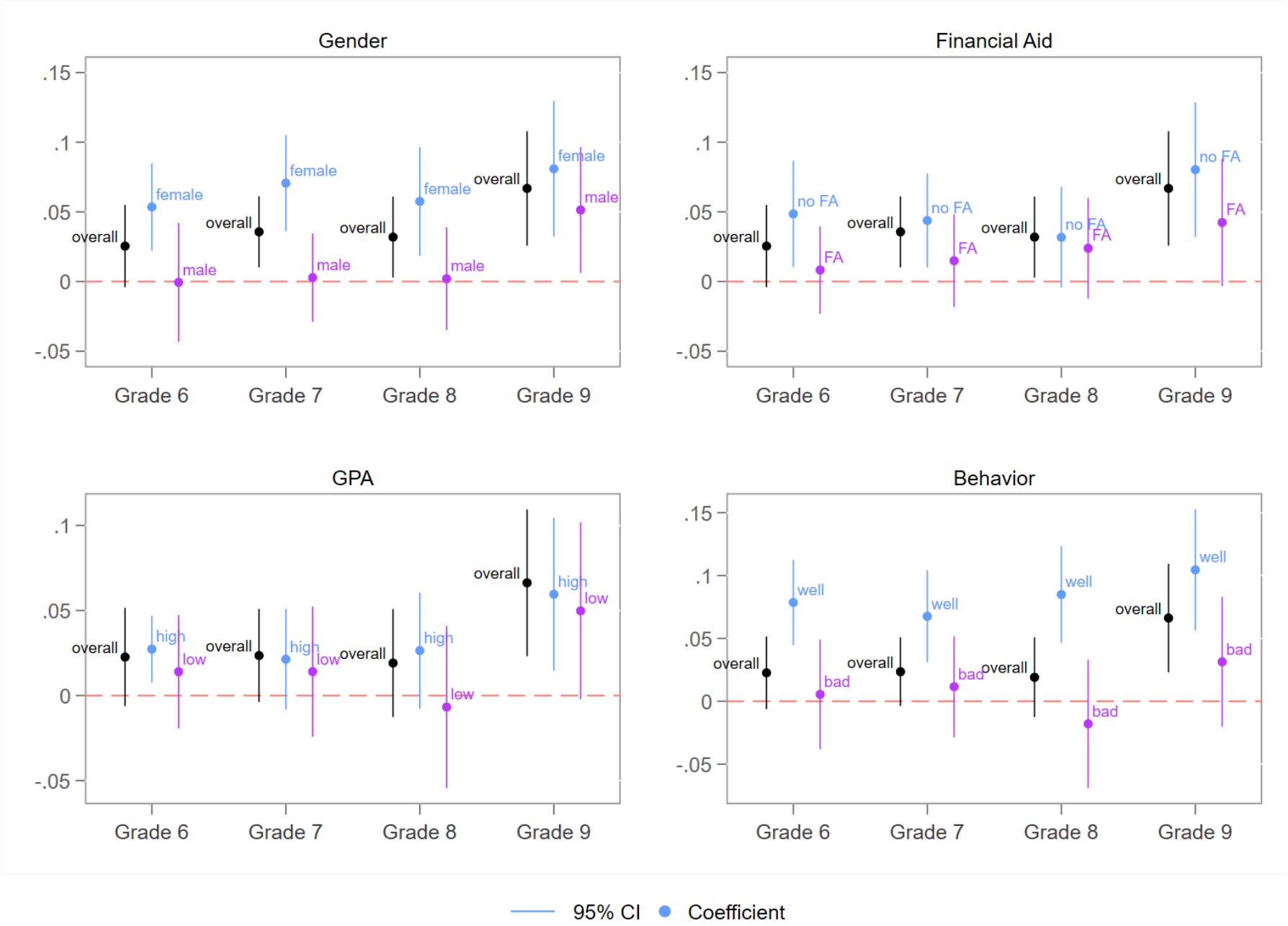


Figure 2: Impact on Perceived Return to Effort

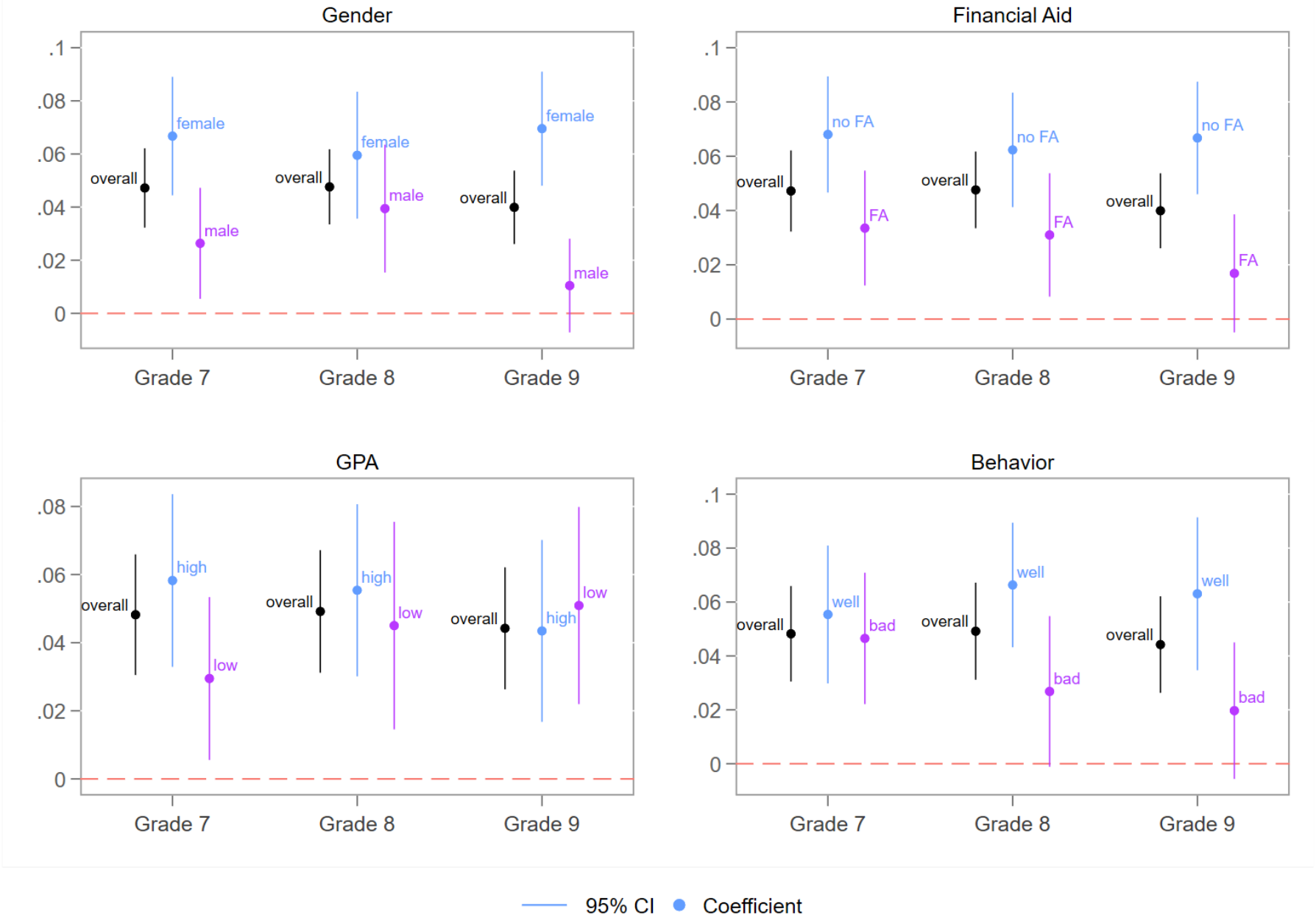


Figure 3: Impact on Self-reported Diligence

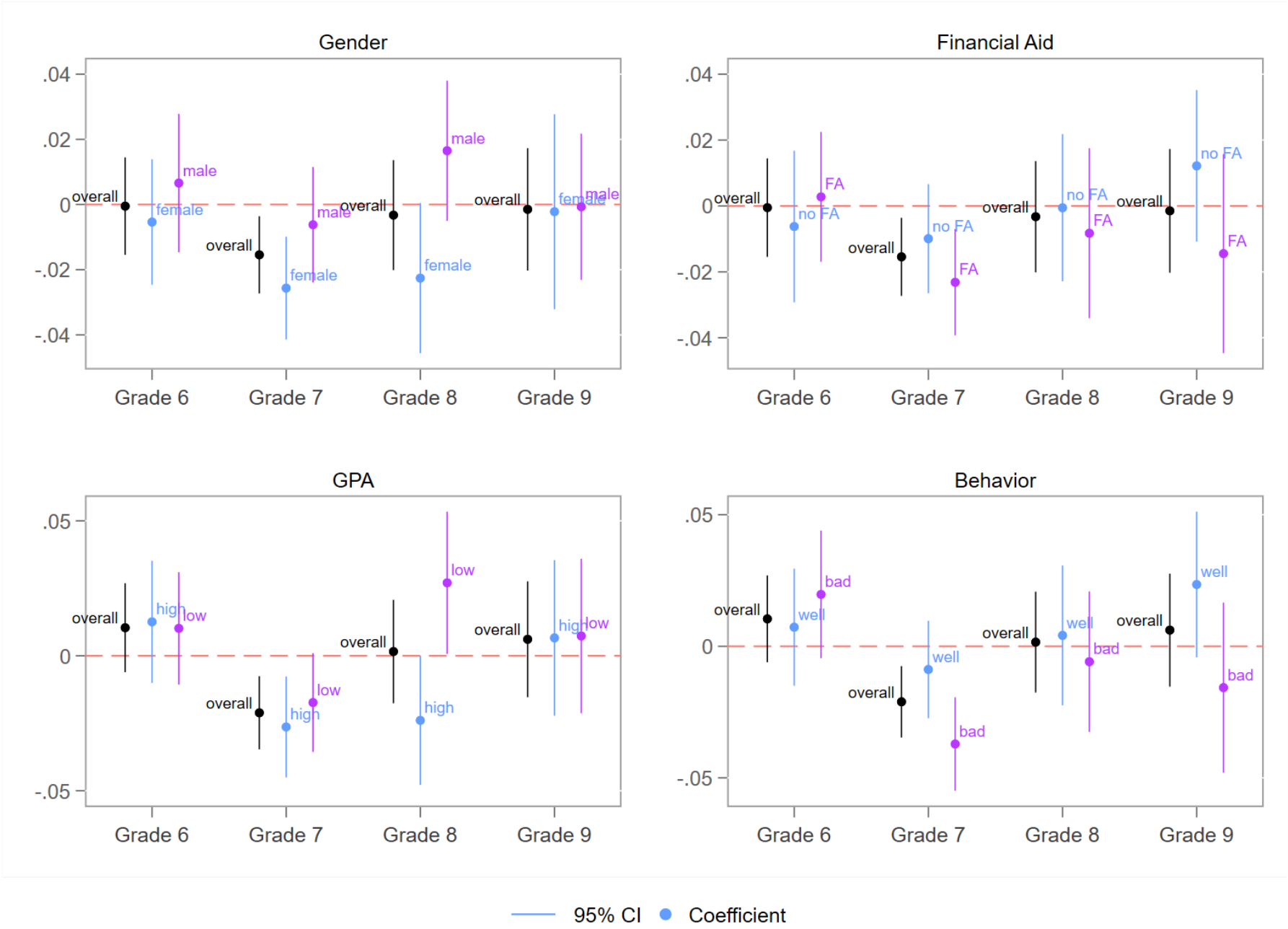


Figure 4: Impact on Teacher-reported Character

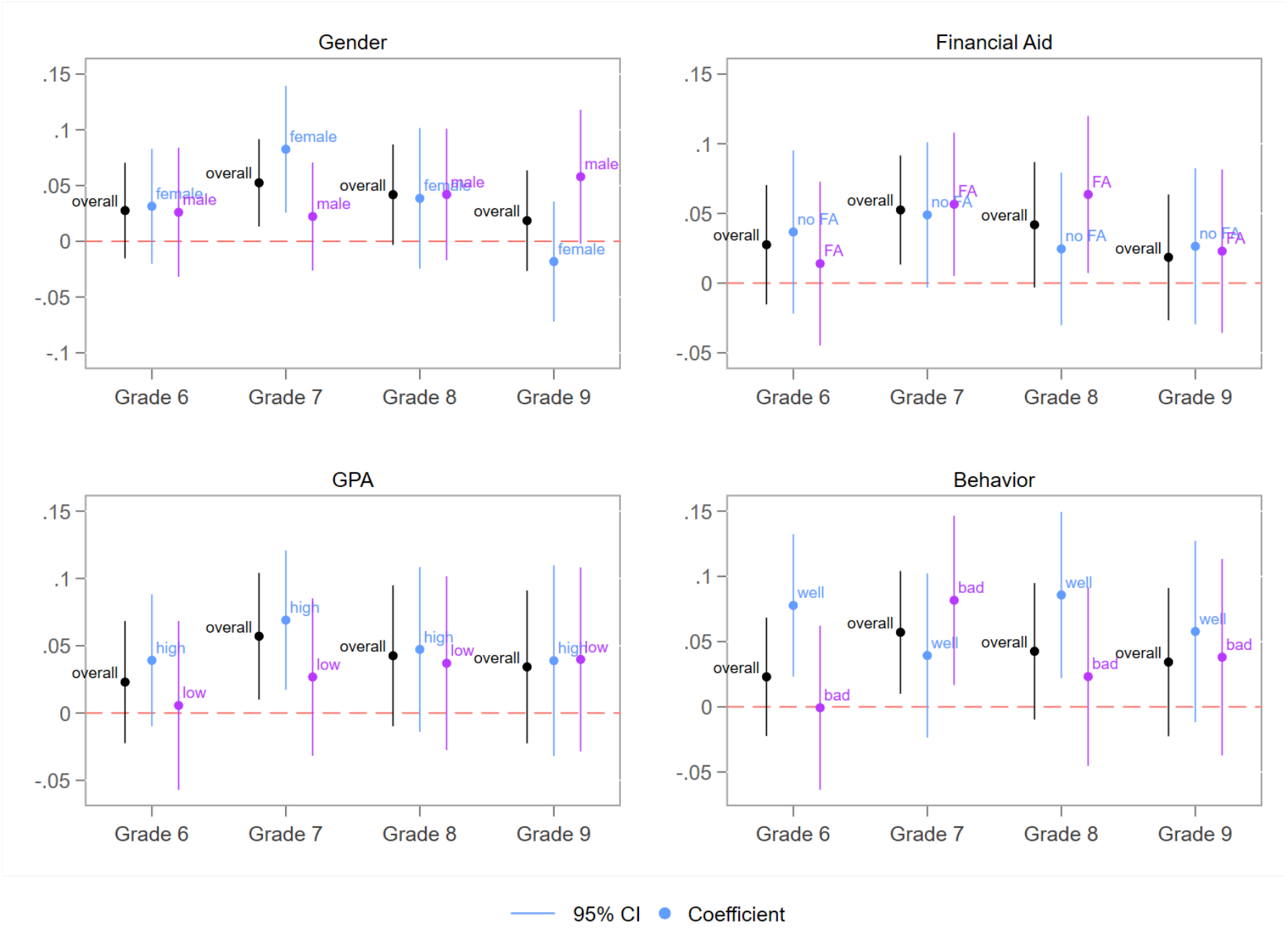


Figure 5: Impact on School-reported Behavior

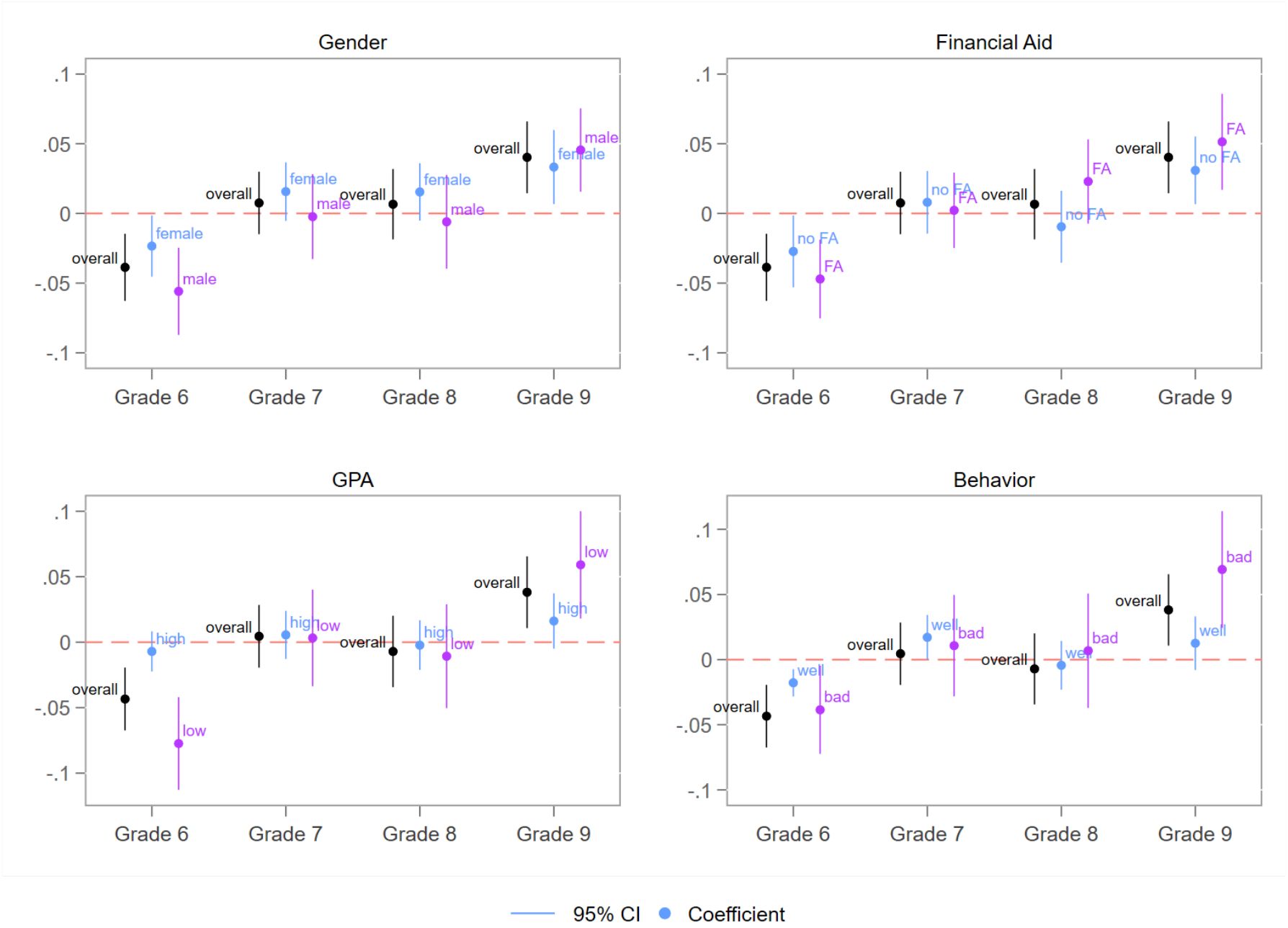
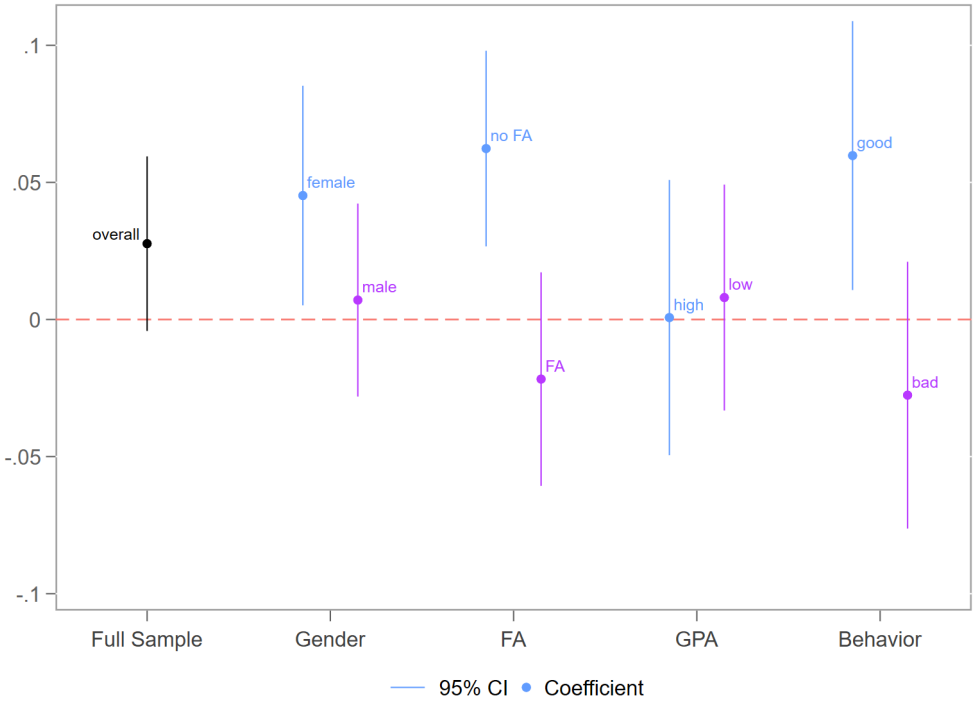


Figure 6: Impacts on the National End-of-Middle School Exam



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	19,885	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	10,031	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	9,854	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	0.14	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	0.04	-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)									

The table shows sample sizes and attrition rates for the different surveys (administrative, student and teacher). 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%

Table 2: Participation and Engagement in 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	0.06	0.65***	5,490	0.03	0.61***
			(0.01)			(0.01)			(0.02)			(0.02)
Honored commitment	5,447	0.08	0.46***	5,998	0.09	0.44***	5,485	0.05	0.43***	5,490	0.02	0.39***
			(0.01)			(0.01)			(0.01)			(0.02)
Student sample size	24,142	11,914		23,095	11,330		23,751	11,817	11,934	23,588	11,645	
Number of clusters	194	97		194	97		190	95	95	186	94	

The 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: EJ Impacts on Summary Indices

	Grade 6		**Grade 7**		**Grade 8**		**Grade 9**	
	Obs.	Impact	Obs.	Impact	Obs.	Impact	Obs.	Impact
GPA	20,783	0.03*	21,443	0.04***	19,713	0.03**	19,330	0.07***
		(0.01)		(0.01)		(0.01)		(0.02)
National exam score	.	.	.	.	.	.	19,197	0.03*
		.		.		.		(0.02)
Return to effort	.	.	6,027	0.05***	5,496	0.05***	5,485	0.04***
	.	.		(0.01)		(0.01)		(0.01)
Student-reported diligence	5,506	0.00	6,458	-0.02**	5,706	-0.00	5,497	0.00
		(0.01)		(0.01)		(0.01)		(0.01)
Teacher-reported character	4,494	0.03	4,826	0.05***	4,596	0.04*	4,503	0.02
		(0.02)		(0.02)		(0.02)		(0.02)
School-reported behavior	22,074	-0.04***	22,449	0.01	22,445	0.01	22,305	0.04***
		(0.01)		(0.01)		(0.01)		(0.01)
Aspiration	.	.	.	.	.	.	5,497	0.04*
		.		.	.	.		(0.02)
Observations	22,662		22,905		23,266		22,688	
Clusters	188		194		190		186	

The table presents the standardized impacts of the treatment from Grade 6 to 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 and standard errors, given below in parenthesis, are robust to heteroscedasticity and clustered at the school*cohort level. The sample is composed of students with non-missing gender and financial aid status.
*10%, **5%, ***1% significance level

Table 4: Heterogeneous Impacts by Gender and Financial Aid Status

	Grade 6			Grade 7			Grade 8			Grade 9		
	EJ	F	EJ*F	EJ	F	EJ*F	EJ	F	EJ*F	EJ	F	EJ*F
Panel A : Gender												
GPA	-0.00	0.30***	0.05**	0.00	0.33***	0.07***	0.00	0.35***	0.06**	0.05**	0.32***	0.03
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)
Return to effort	.	.	.	0.03**	0.04**	0.04***	0.04***	0.05***	0.02	0.01	-0.01	0.06***
	.	.	.	(0.01)	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)
Stud-rep. dilig.	0.01	0.04***	-0.01	-0.01	0.00	-0.02*	0.02	-0.03***	-0.04***	-0.00	-0.04**	-0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)
Teacher-rep. char.	0.03	0.16***	0.01	0.02	0.09***	0.06*	0.04	0.07*	-0.00	0.06*	0.16***	-0.08**
	(0.03)	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)
School-rep. beha.	-0.06***	0.18***	0.03***	-0.00	0.16***	0.02	-0.01	0.15***	0.02*	0.05***	0.13***	-0.01
	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)
	EJ	nFA	EJ*nFA	EJ	nFA	EJ*nFA	EJ	nFA	EJ*nFA	EJ	nFA	EJ*nFA
Panel B : FA												
GPA	0.01	0.31***	0.04*	0.01	0.28***	0.03	0.02	0.24***	0.01	0.04*	0.06*	0.04
	(0.02)	(0.03)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.03)	(0.02)
Return to effort	.	.	.	0.03***	-0.03*	0.03**	0.03***	-0.01	0.03*	0.02	-0.01	0.05***
	.	.	.	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)
Stud-rep. dilig.	0.00	0.02	-0.01	-0.02***	-0.01	0.01	-0.01	-0.01	0.01	-0.01	-0.00	0.03
	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
Teacher-rep. char.	0.01	0.12***	0.02	0.06**	0.14***	-0.01	0.06**	0.13***	-0.04	0.02	0.11***	0.00
	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
School-rep. beha.	-0.05***	0.11***	0.02	0.00	0.10***	0.01	0.02	0.12***	-0.03***	0.05***	0.10***	-0.02
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)
Observations	22,662			22,905			23,266			22,688		
Clusters	188			194			190			186		

The table presents the heterogeneous impacts on summary indices (in SD) by gender (panel A) and financial aid status (Panel B) from Grade 6 to Grade 9. In Panel A, EJ gives the impacts for male students, F the difference between female and male students in the control group, and EJ*F their interaction. In Panel B, EJ gives the impacts for financial aid recipients, F the difference between non-recipients and aid recipients in the control group, and EJ*nFA their interaction. Regressions are controlled for school and cohort fixed effects, and their interactions with the heterogeneity variable (F and nFA respectively). Standard errors in parentheses are robust to heteroscedasticity and clustered at the school*cohort level. The sample only includes students with non-missing gender and FA status.
*10%, **5%, ***1% significance level

Table 5: Heterogeneous Impacts by Baseline GPA and Behavior

	Grade 6			Grade 7			Grade 8			Grade 9		
	EJ	$\mathbb{1}_{g>p50}$	EJ*$\mathbb{1}_{g>p50}$	EJ	$\mathbb{1}_{g>p50}$	EJ*$\mathbb{1}_{g>p50}$	EJ	$\mathbb{1}_{g>p50}$	EJ*$\mathbb{1}_{g>p50}$	EJ	$\mathbb{1}_{g>p50}$	EJ*$\mathbb{1}_{g>p50}$
Panel A: GPA												
GPA	0.01	1.50***	0.01	0.01	1.35***	0.01	-0.01	1.27***	0.03	0.05*	1.04***	0.01
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.02)	(0.02)
Return to effort	.	.	.	0.03**	0.02	0.03*	0.05***	0.03	0.01	0.05***	0.09***	-0.01
	.	.	.	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)
Stud-rep. dilig.	0.01	0.09***	0.00	-0.02*	0.06***	-0.01	0.03**	0.07***	-0.05***	0.01	0.04***	-0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)
Teacher-rep. char.	0.01	0.61***	0.03	0.03	0.50***	0.04	0.04	0.45***	0.01	0.04	0.40***	-0.00
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.04)
School-rep. beha.	-0.08***	0.33***	0.07***	0.00	0.32***	0.00	-0.01	0.29***	0.01	0.06***	0.30***	-0.04***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
	EJ	$\mathbb{1}_{b>p50}$	EJ*$\mathbb{1}_{b>p50}$	EJ	$\mathbb{1}_{b>p50}$	EJ*$\mathbb{1}_{b>p50}$	EJ	$\mathbb{1}_{b>p50}$	EJ*$\mathbb{1}_{b>p50}$	EJ	$\mathbb{1}_{b>p50}$	EJ*$\mathbb{1}_{b>p50}$
Panel B : Behavior												
GPA	0.01	0.70***	0.07***	0.01	0.68***	0.06**	-0.02	0.62***	0.10***	0.03	0.52***	0.07***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.03)
Return to effort	.	.	.	0.05***	0.01	0.01	0.03*	-0.03	0.04**	0.02	0.02	0.04**
	.	.	.	(0.01)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)
Stud-rep. dilig.	0.02	0.09***	-0.01	-0.04***	0.04***	0.03**	-0.01	0.04***	0.01	-0.02	0.04**	0.04*
	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Teacher-rep. char.	-0.00	0.39***	0.08**	0.08**	0.38***	-0.04	0.02	0.28***	0.06	0.04	0.27***	0.02
	(0.03)	(0.03)	(0.04)	(0.03)	(0.04)	(0.04)	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)
School-rep. beha.	-0.04**	0.55***	0.02	0.01	0.37***	0.01	0.01	0.34***	-0.01	0.07***	0.33***	-0.06***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Observations	21,114			18,825			17,465			15,773		
Clusters	183			185			184			180		

Same as Table 4 but for baseline GPA and baseline behavior score. GPA heterogeneity is computed using $\mathbb{1}_{g>p_{50}}$ which takes 1 when the student scored in the GPA top half at baseline (i.e. Grade 6 first quarter of), 0 otherwise. Similarly, behavior score heterogeneity is computed using $\mathbb{1}_{b>p_{50}}$ which takes 1 when the student scored in the top half of the behavior score at baseline (i.e. Grade 6 first quarter).

Table 6: Impacts on Aspirations in Grade 9 by Gender and Financial Aid Status

	Full Sample			Gender Heterogeneity			FA Heterogeneity		
	Obs.	C	Impact	EJ	F	EJ* F	EJ	nFA	EJ*nFA
Professional Aspirations									
High skill	5,372	0.28	-0.00	-0.02	-0.04**	0.03	-0.00	0.02	-0.00
			(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)
Medium skill	5,372	0.20	0.02***	0.01	0.09***	0.02	0.04***	0.04**	-0.04**
			(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)
Low skill	5,372	0.34	-0.02*	-0.00	-0.05**	-0.03*	-0.05***	-0.05**	0.05**
			(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
No aspiration	5,372	0.18	0.01	0.01	-0.01	-0.01	0.01	-0.00	-0.00
			(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)
Educational Aspirations									
Academic High School	5,497	0.69	0.02	-0.00	0.08***	0.04*	0.02	0.06***	0.00
			(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Technical High School	5,497	0.31	-0.02*	0.00	-0.08***	-0.05**	-0.03*	-0.07***	0.01
			(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Repeat G9	5,497	0.00	0.00	-0.00	-0.00	0.01*	0.01***	0.01*	-0.01***
			(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Index									
Aspiration Score (SD)	5,497	-0.00	0.04*	-0.00	0.15***	0.08*	0.06*	0.12***	-0.04
			(0.02)	(0.03)	(0.04)	(0.04)	(0.03)	(0.04)	(0.04)
Observations	5,767	2,799							
Clusters	169	83							

The table presents the treatment impacts on the aspiration outcomes and the heterogeneous impacts by gender and financial aid status using the same specification and notations as in Tables 2 and 4. The sample is restricted to the students with non-missing gender and financial status.

*10%, **5%, ***1% significance level

Table 7: Impacts on Aspirations in Grade 9 by Baseline GPA and Behavior

	Full Sample			GPA heterogeneity			Behavior heterogeneity		
	Obs.	C	Impact	EJ	$\mathbb{1}_{g>p_{50}}$	EJ*$\mathbb{1}_{g>p_{50}}$	EJ	$\mathbb{1}_{b>p_{50}}$	EJ*$\mathbb{1}_{b>p_{50}}$
Professional Aspirations									
High skill	3,932	0.29	-0.01	0.02	0.05	-0.04	0.01	0.06***	-0.02
			(0.01)	(0.02)	(0.08)	(0.03)	(0.02)	(0.02)	(0.03)
Medium skill	3,932	0.20	0.03***	0.02*	-0.10***	0.01	0.03***	0.03**	0.00
			(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)
Low skill	3,932	0.32	-0.03**	-0.04*	-0.14***	0.03	-0.02	-0.08***	-0.04
			(0.01)	(0.02)	(0.03)	(0.03)	(0.02)	(0.02)	(0.03)
No aspiration	3,932	0.19	0.01	-0.00	0.19***	0.00	-0.02	-0.02	0.06**
			(0.01)	(0.01)	(0.05)	(0.02)	(0.02)	(0.02)	(0.02)
Educational Aspirations									
Academic High School	3,995	0.72	0.02	0.06***	0.28***	-0.06**	-0.01	0.10***	0.07***
			(0.01)	(0.02)	(0.06)	(0.03)	(0.02)	(0.02)	(0.02)
Technical High School	3,995	0.27	-0.02	-0.06***	-0.28***	0.07***	0.01	-0.10***	-0.06**
			(0.01)	(0.02)	(0.06)	(0.03)	(0.02)	(0.02)	(0.03)
Repeat G9	3,995	0.00	0.00	0.01***	-0.00	-0.01***	0.01*	-0.00	-0.01**
			(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)
Index									
Aspiration Score (SD)	3,995	0.05	0.04*	0.11***	0.26***	-0.10*	0.03	0.20***	0.05
			(0.02)	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.05)
Observations	4,190	2,047							
Clusters	158	79							

Same as Table 6 for baseline GPA and behavior heterogeneity. GPA heterogeneity is computed using $\mathbb{1}_{g>p_{50}}$ which takes 1 when the student scored in the GPA top half at baseline (i.e. Grade 6 first quarter of), 0 otherwise. Similarly, behavior score heterogeneity is computed using $\mathbb{1}_{b>p_{50}}$ which takes 1 when the student scored in the top half of the behavior score at baseline (i.e. Grade 6 first quarter).

Table 8: Impacts on National End-of-Middle School Exam by Gender and Financial Aid Status

	Full Sample		Gender Heterogeneity			FA Heterogeneity		
	Obs.	Impact	EJ	F	EJ*F	EJ	nFA	EJ*nFA
National exam score, st	19,193	0.03*	0.01	0.10	0.04*	-0.02	0.95***	0.08***
		(0.02)	(0.02)	(0.08)	(0.02)	(0.02)	(0.03)	(0.02)
... *French score, st*	19,660	0.04***	0.02	0.41***	0.04*	0.01	0.79***	0.05**
		(0.01)	(0.02)	(0.10)	(0.02)	(0.02)	(0.09)	(0.02)
... *Maths score, st*	19,625	-0.00	-0.01	-0.24***	0.02	-0.05**	1.08***	0.08***
		(0.02)	(0.02)	(0.05)	(0.02)	(0.02)	(0.04)	(0.02)
... *Hist.geo. score, st*	19,572	0.03*	0.01	0.09	0.04**	-0.02	0.76***	0.08***
		(0.02)	(0.02)	(0.09)	(0.02)	(0.02)	(0.06)	(0.02)
... *Biology score, st*	19,549	0.01	-0.01	-0.06**	0.03	-0.03	0.81***	0.07***
		(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)
... *Oral exam score, st*	19,647	0.03	0.02	0.07***	0.02	-0.00	0.52***	0.06**
		(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.12)	(0.02)

The Table presents the impacts of the national end-of-middle school exam ("Brevet") by gender and financial aid status using the same specification and notations as in Tables 2 and 4. In addition to providing whether the student attended, pass or received an honor, we provide the final score (composed of both G9 GPA and the the results at the exam) as well as the results at the overall exam ("Exam score") and the results at the french and maths exam. Scores are standardized. Standard errors in parentheses are robust to heteroscedasticity and clustered at the school*cohort level.
*10%, **5%, ***1% significance level

Table 9: Impacts on National End-of-Middle School Exam by Baseline GPA and Behavior

	Full Sample		GPA Heterogeneity			Behavior Heterogeneity		
	Obs.	Impact	EJ	$\mathbb{1}_{g>p_{50}}$	EJ*$\mathbb{1}_{g>p_{50}}$	EJ	$\mathbb{1}_{b>p_{50}}$	EJ*$\mathbb{1}_{b>p_{50}}$
National exam score, st	14,060	0.01	0.01	1.21***	-0.01	-0.03	0.50**	0.09***
		(0.02)	(0.02)	(0.11)	(0.03)	(0.02)	(0.22)	(0.03)
... *French score, st*	14,311	0.03	0.02	0.87***	0.01	-0.00	0.12	0.06**
		(0.02)	(0.02)	(0.07)	(0.03)	(0.03)	(0.33)	(0.03)
... *Maths score, st*	14,292	-0.02	-0.01	1.53***	-0.01	-0.05**	0.82***	0.08***
		(0.02)	(0.02)	(0.23)	(0.03)	(0.02)	(0.24)	(0.03)
... *Hist.geo. score, st*	14,260	0.01	-0.01	0.83***	0.02	-0.00	0.24***	0.02
		(0.02)	(0.02)	(0.05)	(0.03)	(0.02)	(0.08)	(0.03)
... *Biology score, st*	14,258	-0.01	-0.02	0.99***	0.01	-0.03	0.40***	0.06*
		(0.02)	(0.02)	(0.04)	(0.03)	(0.02)	(0.10)	(0.03)
... *Oral exam score, st*	14,325	0.02	0.02	0.49***	-0.02	-0.02	0.38***	0.08***
		(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.09)	(0.03)

Same as Table 8 by baseline GPA and baseline behavior score. GPA heterogeneity is computed using $\mathbb{1}_{g>p_{50}}$ which takes 1 when the student scored in the GPA top half at baseline (i.e. Grade 6 first quarter of), 0 otherwise. Similarly, behavior score heterogeneity is computed using $\mathbb{1}_{b>p_{50}}$ which takes 1 when the student scored in the top half of the behavior score at baseline (i.e. Grade 6 first quarter).