

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

Elise Huillery¹, Adrien Bouguen², Axelle Charpentier³, Yann Algan⁴,
Coralie Chevallier⁵

January 17, 2021

Abstract

This article provides experimental evidence of the impact of a four-year intervention aimed at developing students' growth mindset and internal locus of control in disadvantaged middle schools. We find a 0.07 standard deviation increase in GPA, associated with a change in students' mindset, improved behavior as reported by teachers and school registers, and higher educational and professional aspirations. International empirical benchmarks reveal that the intervention is at least ten times more cost-effective than the typical educational intervention. However, while reducing between-school inequality when targeted to disadvantaged schools, the program benefits less to more fragile students, therefore increasing within-school inequality.

JEL classification: I24

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

First and foremost, we are deeply indebted to the *Energie Jeunes* team, who made the evaluation possible and put in endless effort and collaboration with the research team: Philippe Korda, Françoise and Denis Yon, Léa Combette, and many others. We thank the Ministry of Education for their collaboration, in particular Isabelle Robin and Fabienne Rosenwald, who were central in the implementation of the project. We are grateful to Alice Danon, Anaëlle Solnon, Thiphaine Forzy and Thomas Gesmond for outstanding project management, and to Clémence Gleizes, Tereza Varejkova, Cloé Sarrabia, Olivier Frances, Sara Mouhoud, Nathan Viltard, and Clémence Pichon for excellent research assistance, as well as many field assistants who collected the data in the schools every year from 2015 to 2019. Finally, and most importantly, we gratefully acknowledge the staff, teachers, and pupils of the schools for the time and information they shared with us. This research is supported by a grant from 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. All errors are our own. 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 and the unique identifying number is AEARCTR-0000376.

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

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

³Ministère de l'Éducation nationale, de la Jeunesse et des Sports, DEPP: axelle.charpentier@education.gouv.fr

⁴Sciences Po, Paris: yann.algan@sciencespo.fr

⁵Ecole normale Supérieure - PSL, Inserm, Paris, Laboratoire de neurosciences cognitives et computationnelles: coralie.chevallier@gmail.com

1 Introduction

What does it take for children to succeed academically and be better prepared for life? Beyond cognitive abilities, a large number of articles show 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). How adolescents perceive themselves, and 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 salaries. In the past few years, many governments have encouraged a shift of educational priorities to promote the development of non-cognitive 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 Beyond correlational findings, the question then is whether these character skills can indeed be taught, and whether such training has a downstream impact on behavior and academic outcomes.

In this paper, we test the impact of *Energie Jeunes*, a program conducted in French disadvantaged middle schools to stimulate effort by raising the perceived return of schooling among adolescents. Every school year from Grade 6 to Grade 9, students participated in three one-hour in-class sessions presented by two external facilitators. The goal of the program is to increase perceived return to effort by developing a growth mindset of intelligence and an internal locus of control. 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. In addition, the intervention emphasizes the role of effort and encourages students to develop an internal locus of control by downplaying the importance of external constraints linked to family background, teachers or peers. The immediate goal of the program is to induce 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. 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 (Algan et al., 2018).

The experiment was implemented in 97 disadvantaged middle schools located all over the country, which are representative of the population of Priority Education²

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

²disadvantaged schools that benefit from extra support from the State

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 Grade 6 to Grade 9), 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 three sources: school administrative data, a survey administered to a representative sample of students (seven students per class), and a survey administered to a sample of teachers (two teachers per class). The administrative data includes behavior at school (absenteeism, lateness, and disciplinary sanctions), 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. The richness of our data and its different points of view (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. The positive impact on GPA increases slightly over time but is present in all grades, meaning that the effect on academic achievement materializes quickly. 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), 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. Overall, the effect sizes are impressive in view of the program's limited intensity (12 in-class one-hour sessions total in four years) and cost (€65—i.e., \$75—per pupil). The most recent literature on the impact of education interventions in high-income countries indicates that large-scale interventions often fail or have fairly small effects (Cheung and R. E. Slavin, 2016; Fryer Jr, 2017). For instance, the median effect of the randomized trials that evaluated promising interventions in disadvantaged secondary schools funded by the US federal government's i3 initiative is 0.03 SD, although these interventions cost several thousand

US dollars per student (Boulay et al., 2018). The effect of *Energie Jeunes* is twice as big and comes at a considerably smaller cost. A review of 242 evaluations of educational interventions find a median effect size of 0.07 SD, as *Energie Jeunes*, for an average cost of \$882 per pupil—more than ten times its cost (Kraft, 2020). In view of this literature, the cost-effectiveness of the *Energie Jeunes* program is outstanding. The messages conveyed by the program are thus a powerful and relevant lever for improving education in disadvantaged schools.

We next turn to the mechanisms by which the program improved academic achievement. First, the program has made students more optimistic about the possibility of improving their intelligence and academic abilities through effort, and has 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, non-recipients, and well-behaved students than for their counterparts.

Importantly, this change in perceptions resulted in 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 schoolwork, more dynamic and enthusiastic about learning, and less impulsive, than in the control group. School-reported behavior (absenteeism, lateness, and discipline) also improved by 0.04 SD in Grade 9. For the less-disciplined students, i.e., boys, aid recipients, low-achievers and poorly-behaved students, there is a clear 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 rather than quantity of schoolwork (i.e., diligence, concentration, and discipline).

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 provide evidence 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.

This paper provides important insights on mindset interventions. First, such

interventions eventually benefit everyone, but the effects are quicker and larger on students who show better academic attitudes and performance to begin with (especially well-behaved students). 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. Second, although the literature often focuses on baseline academic performance, it seems that the main driver of the heterogeneity is the baseline school behavior, and to a lesser extent gender and socioeconomic background. Third, there is a clear association between GPA impacts and improvement in student’s mindset, which suggests that most of the effect originates from a shift in beliefs and perceptions. Fourth, impacts of mindset interventions extend into areas other than pure academic achievement, like attitude towards schoolwork, discipline, and aspirations, which ultimately have a positive impact academically and beyond *per se*. Finally, on the methodological side, this paper shows that self-reported behavioral measures are problematic because they are prone to reference point bias, and more diligent students may not always perceive themselves as such.

This paper adds to the literature on the causal impact of non-cognitive skills on educational outcomes. Causal evidence at large scale relating non-cognitive skills to school outcomes remains scarce. Heckman et al., 2006 was the first attempt to establish causality from non-cognitive skills to wages by correcting non-cognitive skills for schooling and family background effects. Since then, there have been a few rigorous randomized controlled trials assessing the impacts of interventions targeting non-cognitive skills. 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. Alan and Ertac, 2018 show that an intervention that trains teachers to develop students’ forward-looking behavior, implemented in 37 primary schools in Istanbul, increased students’ patience and disciplinary conduct at school three years after the intervention. 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). 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.

The paper makes several contributions to this literature. First, we evaluate a sustained intervention over all four years of middle-school rather than a one-shot program. This larger window of intervention is critical to assess the overall impact of the program as well as its heterogeneous effects. We find that developing a growth mindset and an internal locus of control improves school performance in adolescents but, for the most fragile students, the impacts only materialize after four years of sustained intervention. The paper provides evidence that sustained interventions are sometimes necessary to reach the more fragile and have larger impact sizes, as it may trigger a positive dynamic of cumulative impacts.

Second, the paper improves the understanding of why such light interventions affect school performance, by providing rich and precise evidence on the channels of impacts, which is often lacking in the literature. Thanks to the large window of intervention over the course of middle school, this paper is the first to measure and 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, which is unique in the literature. Our large sample provides the opportunity to track these mechanisms along four important dimensions, namely gender, socioeconomic background, academic performance, and school behavior. This paper thus improves our understanding of who benefits from a mindset intervention, how students benefit from it and when the effects emerge.

Third, our study is the first to show impacts of a mindset intervention not only on performance at school (GPA), but also on life choices (aspirations). Most papers present impacts in the short-term, or on indicators exterior to the education system. Beyond short-term impacts, our study shows impacts on GPA throughout middle school and on the choice of high-school track, an outcome that is likely to affect a range of long-term professional and social outcomes.

Fourth, this study was done 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 point of views of evaluations (students, teachers, school administration) that turn out to be critical for the overall assessment of

the effects of mindset programs. As a consequence, it offers highly reliable and generalizable results to guide education policy.

Finally, this paper contributes to the literature on social inequalities in education. It is well-known that socioeconomic disadvantage can depress students' academic achievement through multiple mechanisms, including reduced peer and school quality (Chetty, Friedman, et al., 2011; Chetty, Hendren, et al., 2016; Goux and Maurin, 2007), lower parental inputs (Avvisati et al., 2013), and lower inherited ability (Black et al., 2005). This paper explores a different and complementary channel of transmission, namely students' beliefs and mindset. It is closely related to the results of Guyon and Huillery, 2020, who show that students from low socioeconomic backgrounds develop biased perceptions of their academic potential and future chances of success, leading to sub-optimal educational aspirations. This paper shows that light interventions can change these beliefs and buffer the detrimental effects of social stereotypes on academic achievement, reducing the achievement gap between disadvantaged and advantaged schools by 6%. However, while between school inequality might be reduced when targeted to disadvantaged schools, the program benefits more to academically and socially advantaged students, therefore increasing within school inequality.

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 NGO³) 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. 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

³The NGO is also called *Energie Jeunes*

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 two character traits: the growth mindset of intelligence developed by the psychologist Carol Dweck (Dweck and Yeager, 2019) and the internal locus of control (Angela L Duckworth et al., 2019). The educational content of the program is thus based on recent research in psychology (Walton, 2014).

As a typical growth-mindset intervention, the program conveys the message that the human brain is highly plastic and provides arguments against the common belief that ability is innate. Specifically, the program stresses 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 on 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 supports (e.g., videos, class discussions, slides, and case studies).

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-emphasize 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.

During the sessions, students are also invited to consider both mini-cases and their own experience and to reflect on ways to strengthen their effort and perseverance. For instance, facilitators discuss concrete strategies to decrease the cost of efforts, such as setting up a daily routine to do one's homework or getting a full night's sleep to be more attentive in class. They also provide ways to make effort more productive, such as keeping one's cellphone away when doing homework or being attentive in class. Finally, they 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.

Finally, students are 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.

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 terms of statistical precision, our power calculations indicated that 100 schools led to a minimum detectable effect of 0.10 SD (ignoring the reduced variance in the outcome due to school and cohort fixed effects). In September 2014, 97 middle schools, located in seven different regional school districts,⁴ 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 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 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

⁴The schools are located in the regional school districts (called *académies* in French) of Aix-Marseille, Créteil, Amiens, Lille, Lyon, Paris and Versailles.

a class-level or an individual randomization. We assumed that the risk of spillover between friends across cohorts or siblings enrolled in the same middle school was small enough for our design to guarantee internal validity.

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%.⁵ 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 25⁶) 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 (+0.638 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 Jeunes* schools perform much worse on the G6 and G9 national tests (1-1.3 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.

⁵<https://www.education.gouv.fr/reperes-et-references-statistiques-1316>

⁶In metropolitan France, i.e., excluding overseas France.

Classes As described above, our study includes two cohorts of students and all the classes⁷ in these two cohorts took part in the experiment, except for the special-need classes (“Segpa”), which were not targeted by the program.⁸ 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 4.2).⁹

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 dataset¹⁰ 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.¹¹

3.2 Data Sources and Description

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.

⁷In 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).

⁸“Segpa” classes account for only 3% of middle school students at the national level.

⁹Grade 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.

¹⁰It 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.

¹¹The 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 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.

The first source of data is administrative school registers, which provide students' grade point average (GPA), the yearly number of absences, lateness, sanctions¹², and disciplinary actions¹³. 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 participation 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

¹²This 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.

¹³This includes disciplinary hearings for serious offenses—violent behavior, for instance.

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).¹⁴

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.

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

¹⁴In 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.

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 construct 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 management¹⁵, and hours of homework¹⁶. 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 use separate indices for self-perceived orderliness, grit, school-work impulsivity, work discipline, homework management, and hours of homework.

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. 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 use 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

¹⁵Not included in the Grade 6 index as this sub-index was not collected then.

¹⁶Not included in the Grade 6 index as this sub-index was not collected then.

scales.

3.3.5 School-reported Behavior

If the intervention can change perceived return to effort and character, we hypothesize that behavior at school will improve. We use a summary index of school-reported behavior aggregating information from school administrative registers: the number of absences (counted in half-days, meaning that any hour missed counts as a half-day), the number of times the student was late, the number of sanctions (e.g., expulsion from class or detention), and the 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, while disciplinary actions are taken by a disciplinary board for more severe misconduct. The school-reported behavior summary index is constructed using the same methodology as described above.

3.3.6 Educational and Professional Aspirations

In France, there are two main types of high school: technical high school (including a two-year track and a three-year track)¹⁷ and academic high school¹⁸. 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.

Regarding career aspirations, the question on their preferred job was open-ended. We code the answer using the National Institute for Statistics and Economic Studies (INSEE) classification of jobs.¹⁹ 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 construct 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 storekeepers²⁰, 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 have been assigned to the low-

¹⁷Here we group “lycée professionnel”, “CAP” and “apprentissage”.

¹⁸In French: “lycée général et technologique”.

¹⁹<https://www.insee.fr/fr/metadonnees/pcs2003/categorieSocioprofessionnelleAgreguee/1?champRecherche=true>

²⁰Except students who aspire to become entrepreneurs, whom we assigned to the high-skilled job category.

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 is 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 use the GPA—average grades from all major academic courses—collected each year for all students. We aggregate and standardize the GPA score using the same methodology as described above.

4 Validation of the Experimental Protocol

In this section, we verify that the experimental protocol was 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. 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 to sample fluctuations. 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 represent a small proportion of the tested differences in means, in line with pure sampling fluctuations. In the remainder of the paper, we thus favor the average differences between the treatment and control groups without covariates.

4.2 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 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 is balanced across the treatment and control groups so it does not affect the internal validity of our results. Moreover, the comparison between Appendix Table D1 and Appendix Table D2 shows 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 are 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 are 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.

4.3 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. Students who registered in the school after grade 6 did not receive the full intervention but they are included in the impact estimates. The proportion of students who were enrolled from Grade 6 on is 92% in Grade 7, 82% in Grade 8, and 75% in Grade 9. Therefore, 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,²¹ 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 median²² (hereafter, “high-achievers” *versus* “low achievers”), and whether her school behavior at baseline was above the median²³ (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. 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 the stacked sample as a robustness check and a synthetic view of impact heterogeneity.

²¹See AEA RCT Registry Number AEARCTR-0000376.

²²Academic performance at baseline is GPA during the first three months in Grade 6.

²³School 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.

5 Impact on Academic Achievement

Table 2 shows the impact of the program on academic achievement in each grade. The 4-year intervention increased GPA by 0.067 SD in Grade 9, which represents 6% of the Priority Education *versus* national average achievement gap.²⁴ Because only 75% of students attended the same school from Grade 6 to Grade 9, 25% did not receive the full program. Therefore, a treatment-on-the-treated analysis would produce up to 33% larger effect sizes. The impact on GPA in previous grades is smaller, but still significant: 0.025 SD in Grade 6, 0.036 SD in Grade 7, and 0.032 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, 0.3 SD and 0.8 SD in Grades 6-9 respectively). Even more impressive, 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 (+0.051 SD), aid-recipients (+0.042 SD), low-achievers (+0.05 SD) and to a lesser extent poorly-behaved students (non significant +0.031 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 0.051 SD between girls and boys, 0.028 SD between aid recipients and non-recipients students, and 0.068 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.

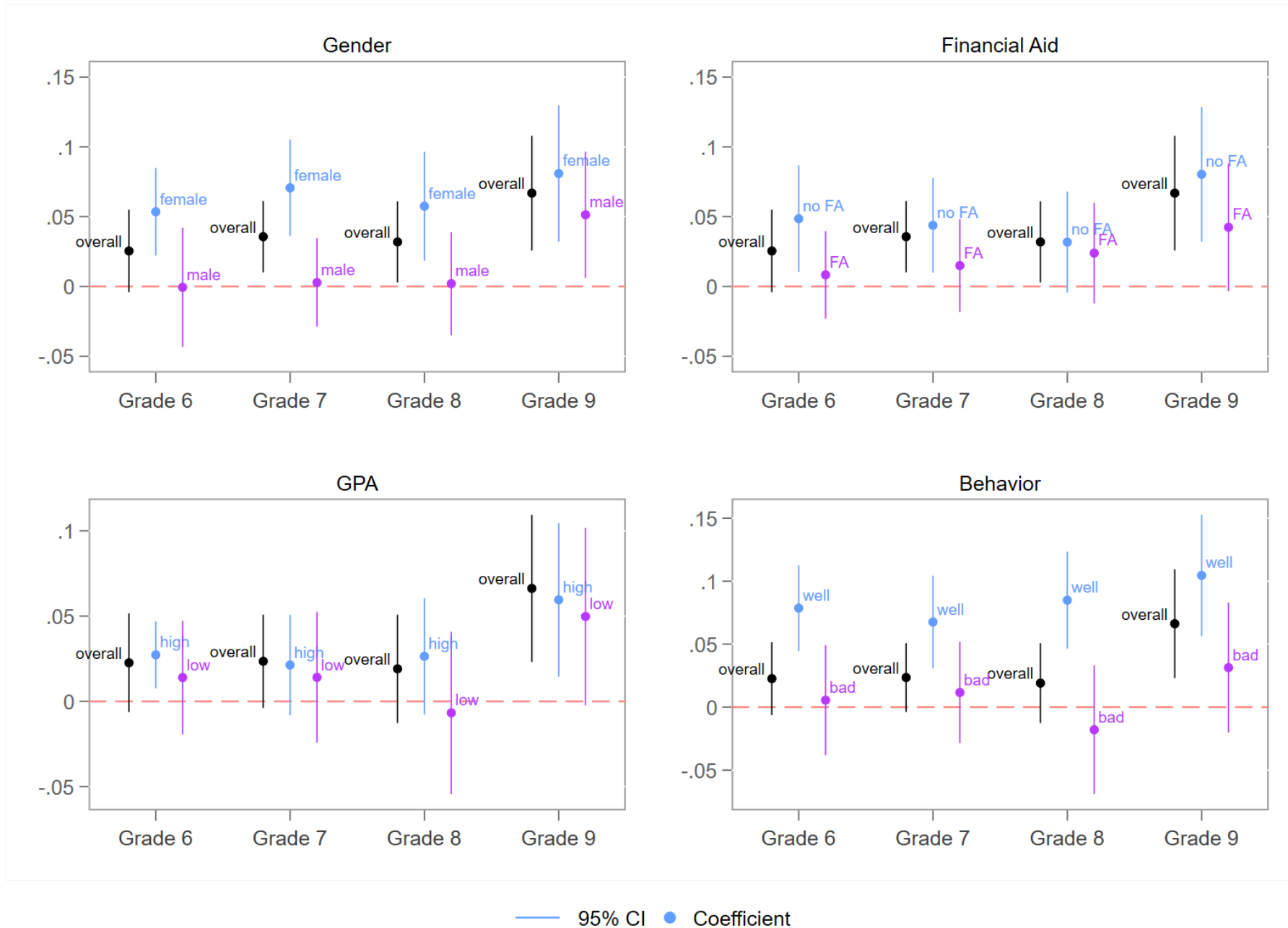
Are the magnitudes of these impacts substantively meaningful? As mentioned in the introduction, large-scale education interventions in high-income countries often fail or have fairly small effects. Cheung and R. E. Slavin, 2016 find average effect sizes on academic achievement of 0.16 SD among 197 Randomized 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 costs. Kraft, 2020 provides effect-size benchmarks

²⁴According to Table B1, the achievement gap between Priority Education schools and the national average in Grade 9 is equal to 1.2 SD.

with a corresponding set of per-pupil cost benchmarks from 242 studies evaluating educational programs. The results show that the effect size of *Energie Jeunes* on the whole population (0.07 SD) is at the 50th percentile of the distribution of effect sizes, but the program 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. Moreover, 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. The fact that this experiment was conducted on a large number of students ($\approx 24,000$) and on schools fairly representative of the population of priority education schools in France confers more importance to our results.

To conclude, given the small cost and large scale of the program, effect sizes such as 0.067 SD overall—0.08-0.10 SD for girls, non-recipients, and well-behaved students, who account for half of the population by construction—are impressive.

Figure 1: Impact on GPA



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 but cannot prove that the following results *explain* the impact on GPA, as these are just associated impacts.

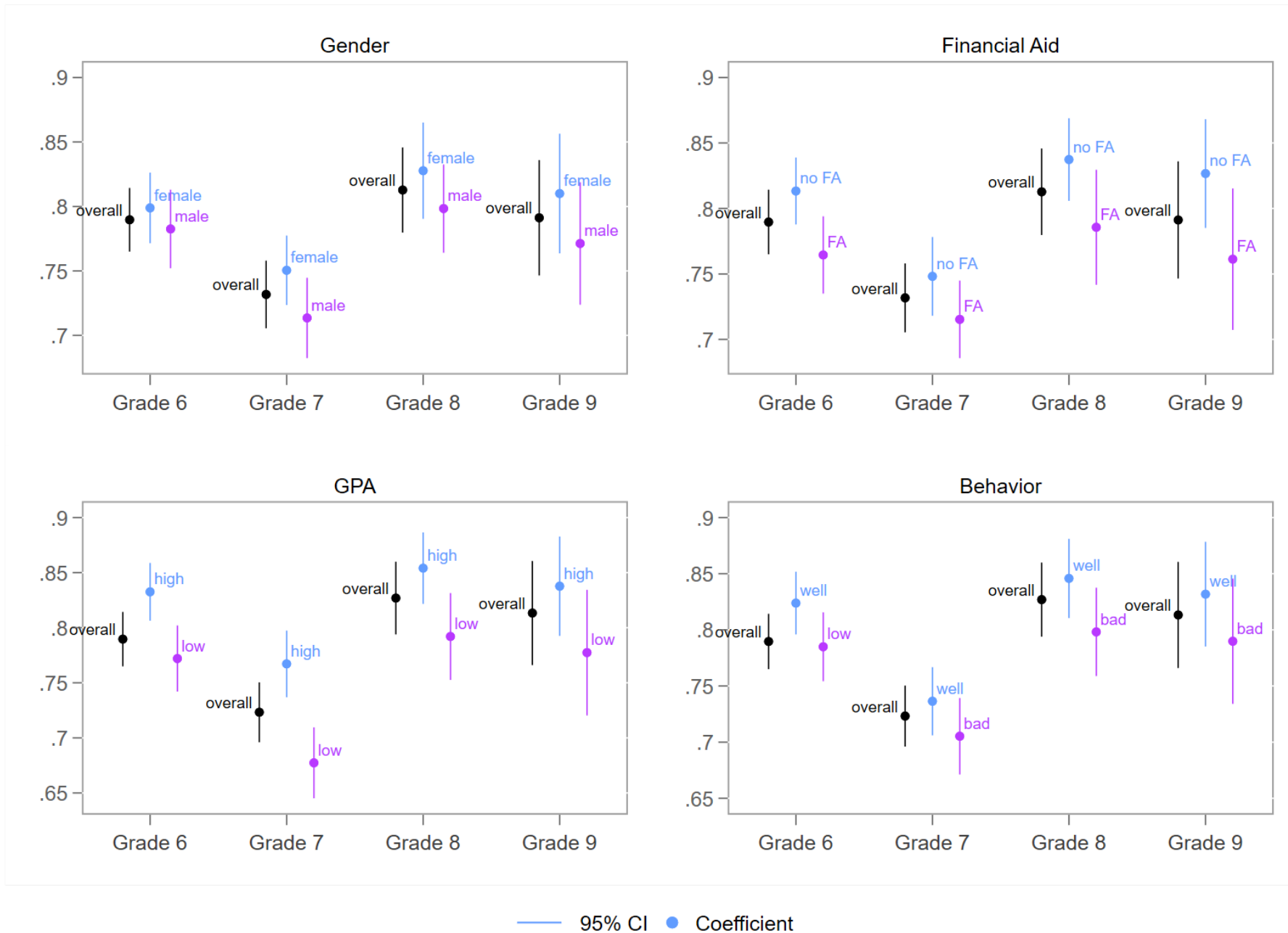
6.1 Take-Up

Table 3 shows that there are large differences in take-up between the treatment and the control groups. Although students in the control group were not supposed to participate in any of the sessions, there may have been mistakes in implementing the program. Based on survey data, participation among control students was between 3.6% and 18.7%. These survey responses also may be due to students misinterpreting the question. In the treatment group, participation in at least one session increases by 73-81 percentage points, which means that 83-95% of students in the treatment group declared that they participated in at least one session. Students in the treatment group report having attended about 2.5 more sessions than the control group, where 3 was expected under perfect compliance with the protocol. Overall, compliance is not perfect but still high.

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 play the game (but do not always succeed at it), and that it is harder to involve older students than younger ones.

Do heterogeneous impacts on GPA come from differential participation in the program? Figure 2 shows participation by grade and by subgroup. We see that girls, non-recipients, high-achievers, and well-behaved students consistently participate more in every grade. 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. These results provide clear evidence that take-up is positively correlated with predictors of better school outcomes, probably because boys, aid recipients, low-achievers, and poorly-behaved students are more likely to be absent and miss parts of the intervention. Differences in participation contribute to downstream differences in impacts on GPA.

Figure 2: Impact on Participation



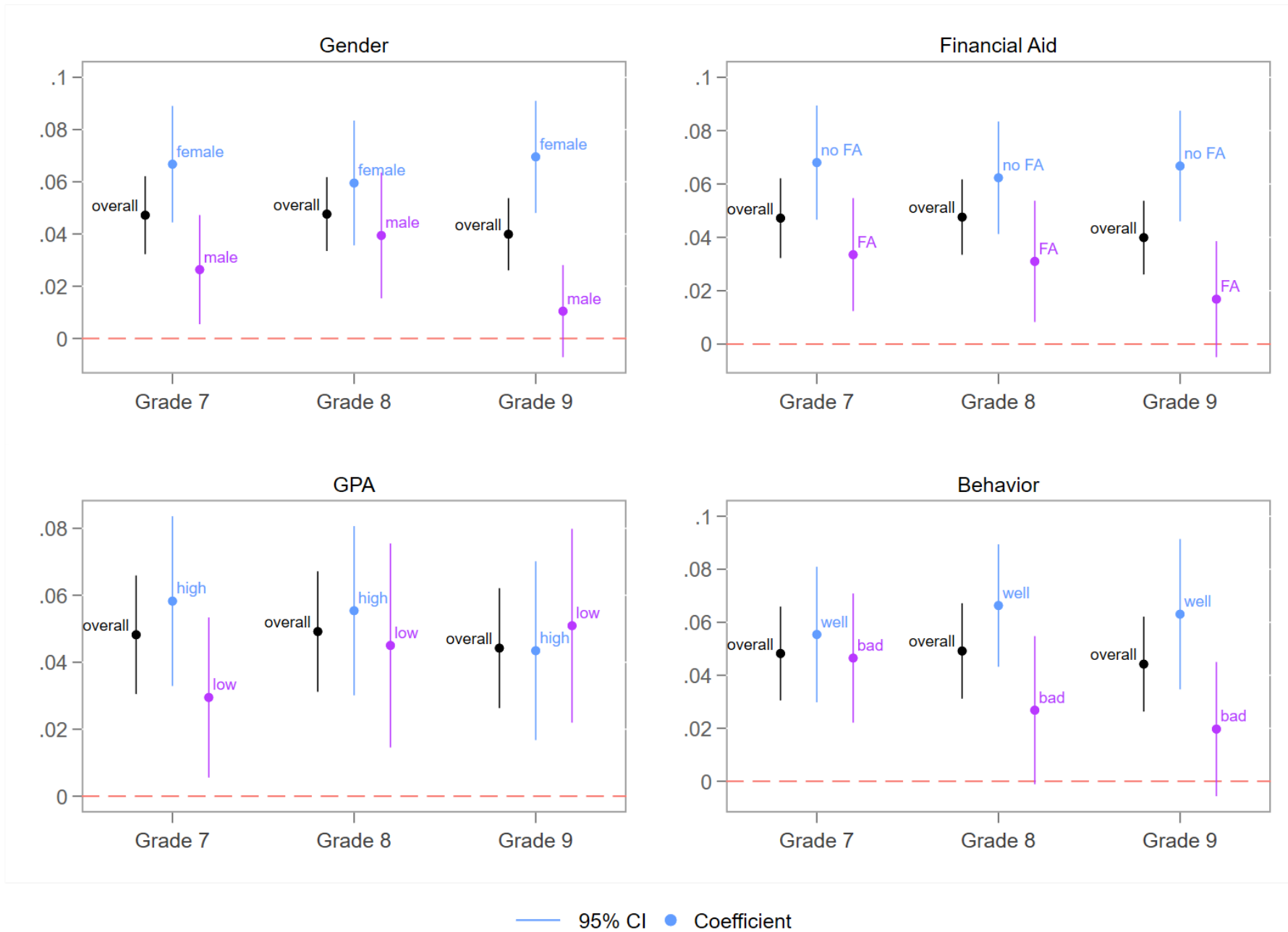
6.2 Perceived return to effort

Table 2 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 0.048 SD. Students are generally prone to update their perceptions and beliefs regarding their chances of success at school. However, the malleability of perceptions varies according to gender, socioeconomic background, and baseline behavior. Figure 3 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 0.024 SD for boys and 0.057 SD for girls (the difference between boys and girls is significant at the 1% level). Similarly, Figure 3 shows that 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, non-recipients students 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 non-recipients and well-behaved, although the estimates on the interaction term are less precise and not significant. This confirms that growth mindset is an important pathway of our impact on GPA.

Figure 3: Impact on Growth Mindset



6.3 Self-reported diligence

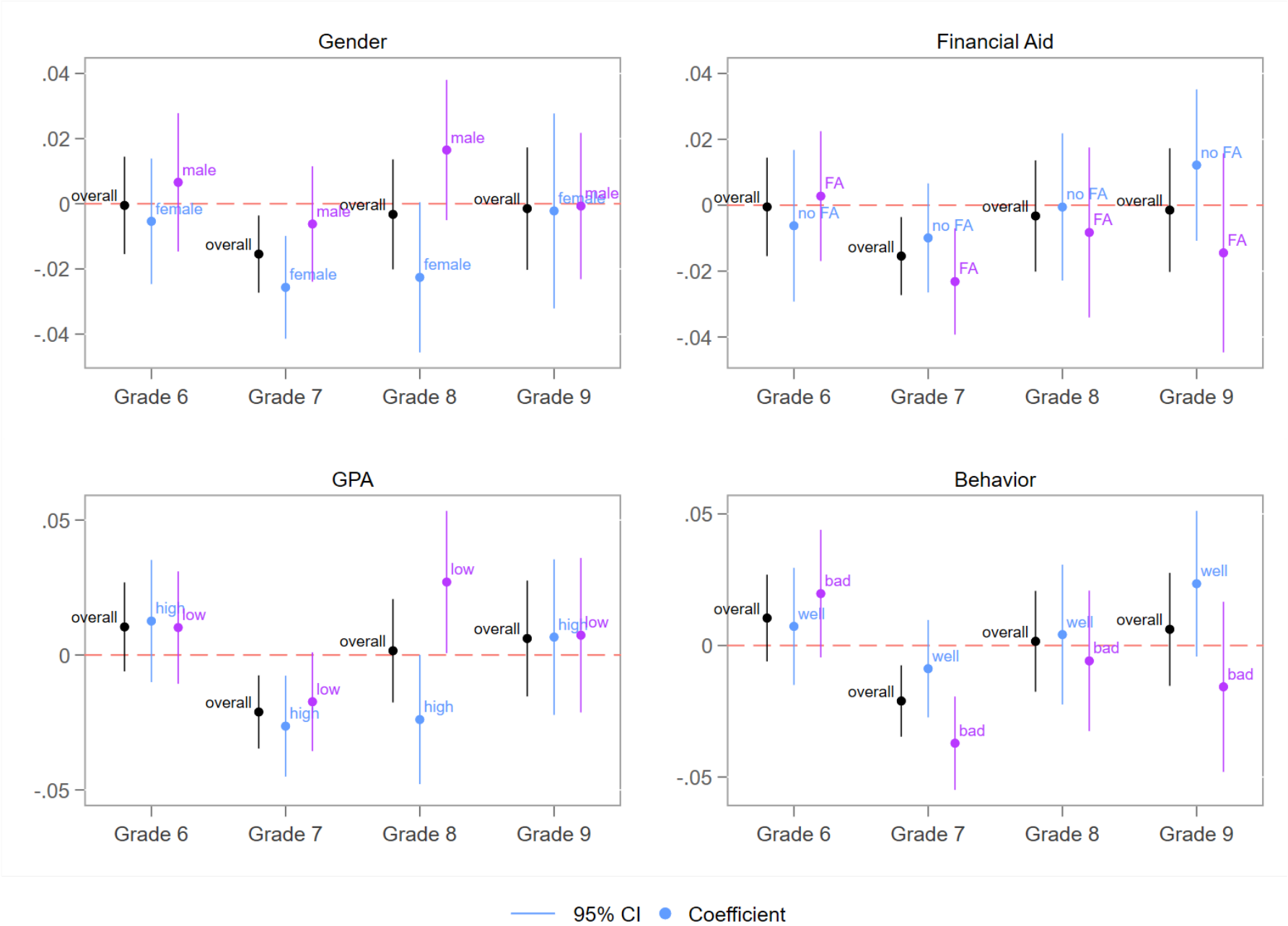
Table 2 shows null impacts of the intervention on self-reported diligence, and even a small negative impact in Grade 7. The subgroup analysis in Tables 4 and 5 shows that the negative impact in Grade 7 concerns mostly girls and poorly-behaved students (see Figure 4). The differential response according 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 -0.023 to -0.052 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.

Whether these updates in self-assessment are correct or not is a difficult question. One explanation may be true deterioration in students' behavior due to possible reactance to *Energie Jeunes* messages. 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.

Figure 4: Impact on Self-reported Diligence



6.4 Teacher-reported Character

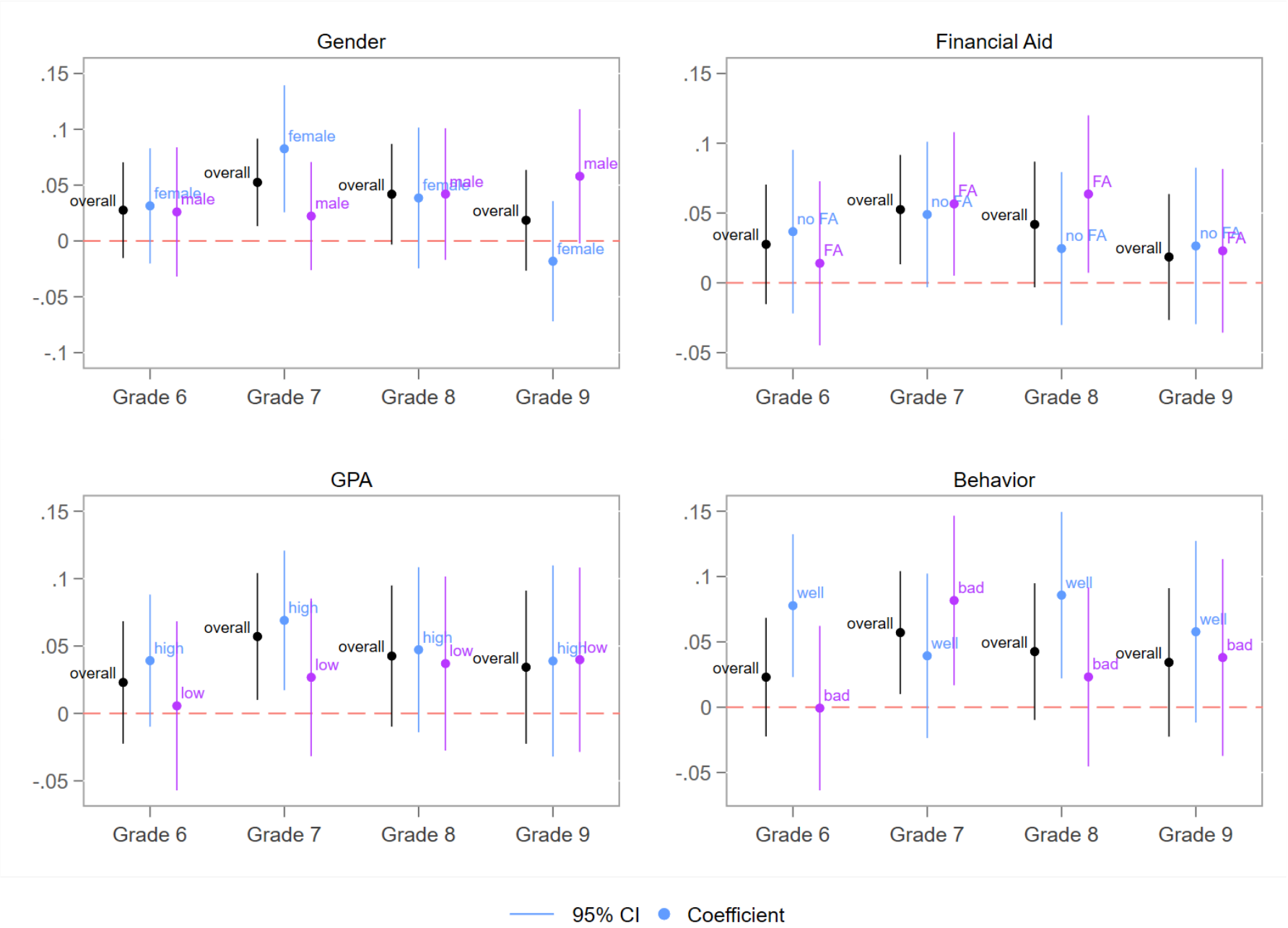
Table 2 shows that the intervention had a positive impact on student character as perceived by their teachers. The effect is 0.028 SD (non-significant) in Grade 6, 0.052 SD in Grade 7, 0.042 SD in Grade 8, and 0.019 SD (non-significant) in Grade 9. When we stack all grades together to avoid multiple hypothesis testing, we find an overall significant effect of 0.039 SD (Appendix Table E1).

Figure 5 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 0.047 SD and 0.082 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.

Figure 5: Impact on Teacher-reported Character



6.5 School-reported Behavior

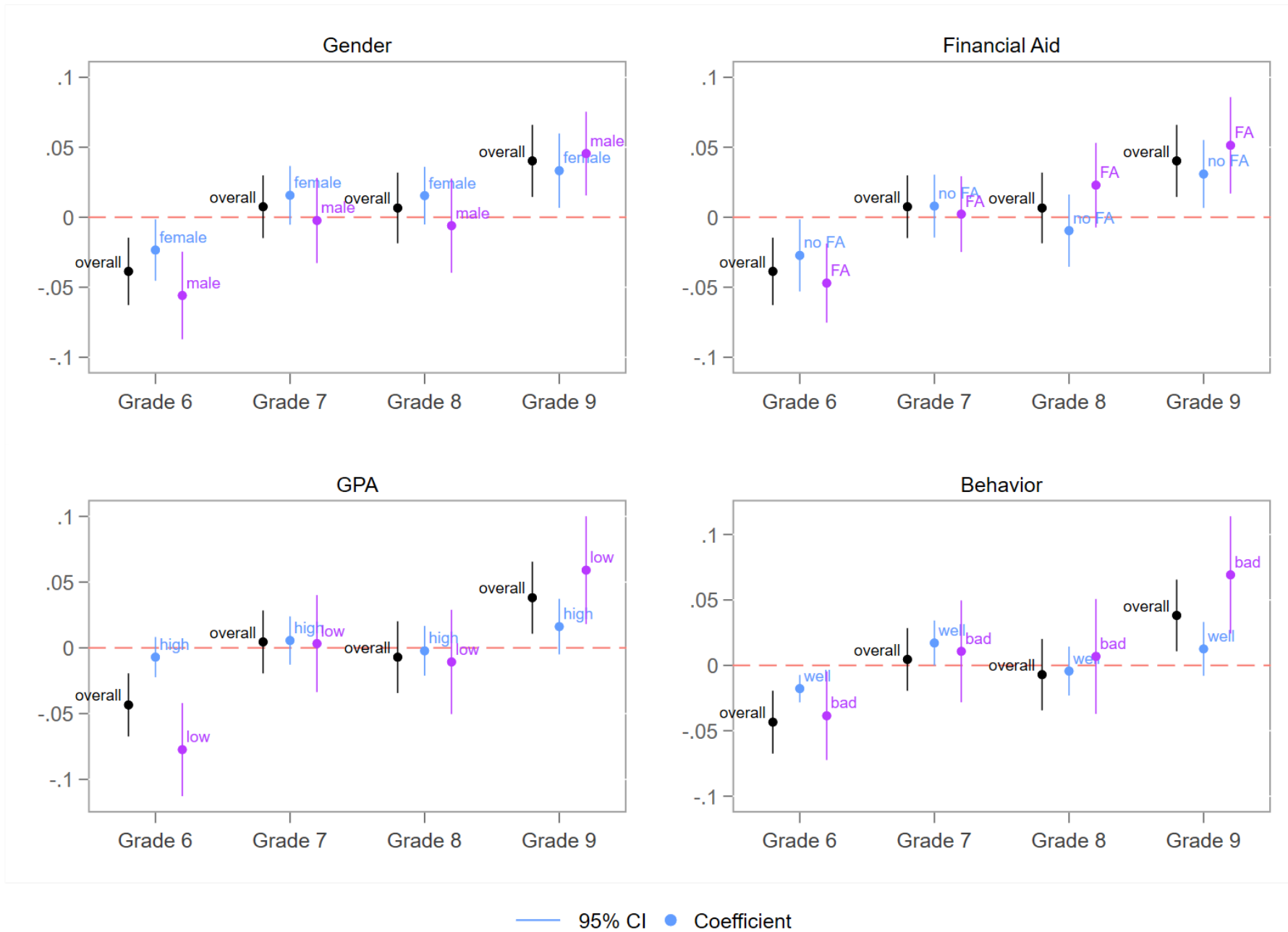
Table 2 shows that the intervention had a positive impact on student behavior reported in school registers, i.e., absences, lateness and discipline, once the whole intervention was delivered. Surprisingly, the initial response of the students was negative in Grade 6, then null in Grades 7 and 8, and 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 improve behavior at school.

Figure 6 (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 (+0.059 SD and +0.069 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' behavior, 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 (+0.045 SD, +0.051 SD, +0.059 SD and +0.069 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.

In line with teachers' views of students' character, Appendix Figure G5 shows that school behavior worsens over the course of middle school for all subgroups. The intervention worked again as a moderator of this natural deterioration, in particular among students whose behavior is more problematic to start with. Overall, impacts on behavior are not fully aligned with impacts on GPA, as they seem stronger in different subgroups, and not always concomitant. 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, behavioral changes are a prerequisite for academic achievement. Our paper only shows associations between impacts on different outcomes, so the evidence is only suggestive.

Figure 6: Impact on School-reported Behavior



6.6 Aspirations

We finally find a positive impact of *Energie Jeunes* on students' aspirations, both educational and professional. As shown in Table 2, the aspiration score is 0.035 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. 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 (+0.059 SD), low-achievers (+0.11 SD), and well-behaved students (+0.078 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 to²⁵.

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 encourage more effort and therefore higher school outcomes later on.

²⁵Note that in the aspiration score, the “I do not know which job I aspire to” category is loaded negatively.

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 and an internal locus of control in adolescents, associated with improvements in school-related behavior, academic and professional aspirations, and GPA. 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 exceptionally 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 (Befy and Davezies, 2013; Bénabou et al., 2009; Bressoux et al., 2016; Goux, Gurgand, 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. The next question is whether mindset interventions such as the one studied here could produce larger effect sizes if the intensity and cost were increased, for instance through more intensive and individualized sessions with students, or through integrating mindset interventions into teacher training programs or curriculum.

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.
- Avvisati, Francesco, Marc Gurgand, Nina Guyon, and Eric Maurin (Sept. 2013). “Getting Parents Involved: A Field Experiment in Deprived Schools”. In: *The Review of Economic Studies* 81.1, pp. 57–83.
- 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.
- Black, Sandra E., Paul J. Devereux, and Salvanes Kjell G. (2005). “Why the Apple Doesn’t Fall Far: Understanding Intergenerational Transmission of Human Capital”. In: *The American Economic Review* 95.1, pp. 437–449.
- 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] <https://www.ipp.eu/wp-content/uploads/2016/03/evaluation-pre-prog-reussites-educ-rapport-IPP-mars2016.pdf> (page consultée le 17 mai 2018)*.
- 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.
- Chetty, Raj, John N Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny Yagan (2011). “How does your kindergarten classroom

- affect your earnings? Evidence from Project STAR". In: *The Quarterly Journal of Economics* 126.4, pp. 1593–1660.
- Chetty, Raj, Nathaniel Hendren, and Lawrence Katz (2016). "The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Project". In: *American Economic Review* 106.4.
- 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.
- 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.

- Goux, Dominique and Eric Maurin (2007). "Close neighbours matter: Neighbourhood effects on early performance at school". In: *The Economic Journal* 117.523, pp. 1193–1215.
- 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, Sergio Urzua, and Edward Vytlačil (2006). "Understanding instrumental variables in models with essential heterogeneity". In: *The Review of Economics and Statistics* 88.3, pp. 389–432.
- 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. 510–51.
- 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.

Tables

Table 1: Samples and Attrition

	Administrative data				Student-reported Survey Data				Teacher-reported Survey Data			
	G6	G7	G8	G9	G6	G7	G8	G9	G6	G7	G8	G9
Actual												
<i>Full</i>	24,142	23,095	23,751	23,588	5,836	6,573	6,003	5,774	4,699	4,832	4,602	4,503
<i>Control</i>	11,914	11,330	11,817	11,645	2,868	3,215	2,973	2,805	2,868	3,215	2,973	2,805
<i>Treatment</i>	12,228	11,765	11,934	11,943	2,968	3,358	3,030	2,969	2,968	3,358	3,030	2,969
Expected												
<i>Full</i>	24,142	23,095	24,349	24,708	5,130	6,965	7,070	7,231	5,130	6,965	7,070	7,231
<i>Control</i>	11,914	11,330	12,070	12,079	2,525	3,437	3,514	3,570	2,525	3,437	3,514	3,570
<i>Treatment</i>	12,228	11,765	12,279	12,629	2,605	3,528	3,556	3,661	2,605	3,528	3,556	3,661
Attrition												
<i>Full</i>	0.000	0.000	0.025	0.045	0.050	0.059	0.153	0.208	0.232	0.309	0.351	0.382
<i>Control</i>	0.000	0.000	0.021	0.036	0.050	0.065	0.156	0.223	0.227	0.318	0.353	0.417
<i>Treatment</i>	0.000	0.000	0.028	0.054	0.049	0.052	0.151	0.193	0.238	0.299	0.349	0.349
<i>Differential</i>	0.000	0.000	0.007	0.018	-0.001	-0.013	-0.005	-0.030	0.011	-0.019	-0.004	-0.069
	(0.000)	(0.000)	(0.025)	(0.032)	(0.025)	(0.025)	(0.046)	(0.050)	(0.045)	(0.050)	(0.056)	(0.055)

The table gives the sample size and attrition rate for the administrative data, the student-reported and the teacher-reported measures. In rows, we first provide the number of observations in the full sample, control group and treatment group (*Actual*), then the number of observations that we expected if no students attrited (*Expected*), and last the comparison between the actual and the expected sample (*Attrition*). The last row (*Differential*) gives the result of the regression of the attrition variable on the treatment variable, controlled for school and cohort fixed effect. The corresponding robust and clustered at school*cohort level standard errors are given below in parenthesis.

* 10%, ** 5%, *** 1%

Table 2: EJ Impacts on Summary Indices

	Grade 6		Grade 7		Grade 8		Grade 9	
	Obs.	Impact	Obs.	Impact	Obs.	Impact	Obs.	Impact
Return to effort	.	.	6,027	0.047*** (0.008)	5,496	0.048*** (0.007)	5,485	0.040*** (0.007)
Student-rep. dilig.	5,506	-0.001 (0.008)	6,458	-0.015** (0.006)	5,706	-0.003 (0.009)	5,497	-0.001 (0.007)
Teacher-rep. char.	4,494	0.028 (0.022)	4,826	0.052*** (0.020)	4,596	0.042* (0.023)	4,503	0.019 (0.023)
School-rep. behav.	22,074	-0.039*** (0.012)	22,449	0.008 (0.011)	22,445	0.007 (0.013)	22,305	0.040*** (0.013)
GPA	20,783	0.025* (0.015)	21,443	0.036*** (0.013)	19,713	0.032** (0.015)	19,330	0.067*** (0.021)
Aspiration	5,497	0.035* (0.021)
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 variable. 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 3: 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	0.156	0.798*** (0.012)	5,998	0.187	0.731*** (0.013)	5,485	0.097	0.813*** (0.017)	5,490	0.036	0.791*** (0.023)
# of sessions attended	5,244	0.208	2.459*** (0.035)	5,698	0.209	2.417*** (0.034)	5,334	0.112	2.434*** (0.044)	5,490	0.066	2.261*** (0.065)
Made commitment	5,446	0.094	0.660*** (0.011)	5,998	0.111	0.595*** (0.012)	5,485	0.044	0.587*** (0.014)	5,490	0.020	0.519*** (0.019)
Honored commitment	5,447	0.080	0.461*** (0.010)	5,998	0.087	0.443*** (0.011)	5,485	0.035	0.396*** (0.011)	5,490	0.016	0.337*** (0.014)
Student sample size	24,142	11,914		23,095	11,330		23,751	11,817		23,588	11,645	
Number of clusters	194	97		194	97		190	95		186	94	

The table provides several 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 average in the control group and *Impact* the coefficient from the regression of the outcomes on the treatment variable. Regressions are controlled 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 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												
Return to effort	.	.	.	0.026**	0.036**	0.040***	0.039***	0.047***	0.020	0.010	-0.006	0.059***
	.	.	.	(0.011)	(0.015)	(0.015)	(0.012)	(0.016)	(0.020)	(0.009)	(0.014)	(0.013)
Stud-rep. dilig.	0.007	0.042***	-0.012	-0.006	0.003	-0.019*	0.016	-0.035***	-0.039***	-0.001	-0.038**	-0.001
	(0.011)	(0.012)	(0.014)	(0.009)	(0.011)	(0.012)	(0.011)	(0.013)	(0.014)	(0.011)	(0.017)	(0.019)
Teacher-rep. char.	0.026	0.155***	0.005	0.022	0.086***	0.060*	0.042	0.066*	-0.004	0.058*	0.158***	-0.076**
	(0.029)	(0.032)	(0.034)	(0.025)	(0.031)	(0.033)	(0.030)	(0.035)	(0.040)	(0.030)	(0.034)	(0.034)
School-rep. beha.	-0.056***	0.184***	0.032***	-0.002	0.156***	0.018	-0.006	0.145***	0.021*	0.045***	0.128***	-0.012
	(0.016)	(0.013)	(0.012)	(0.015)	(0.016)	(0.013)	(0.017)	(0.015)	(0.013)	(0.015)	(0.013)	(0.011)
GPA	-0.001	0.299***	0.054**	0.003	0.326***	0.068***	0.002	0.354***	0.055**	0.051**	0.318***	0.030
	(0.022)	(0.023)	(0.024)	(0.016)	(0.023)	(0.022)	(0.019)	(0.022)	(0.025)	(0.023)	(0.024)	(0.024)
	EJ	nFA	EJ*nFA	EJ	nFA	EJ*nFA	EJ	nFA	EJ*nFA	EJ	nFA	EJ*nFA
Panel B : FA												
Return to effort	.	.	.	0.034***	-0.028*	0.035**	0.031***	-0.011	0.031*	0.017	-0.014	0.050***
	.	.	.	(0.011)	(0.015)	(0.015)	(0.012)	(0.017)	(0.016)	(0.011)	(0.014)	(0.016)
Stud-rep. dilig.	0.003	0.023	-0.009	-0.023***	-0.009	0.013	-0.008	-0.015	0.008	-0.014	-0.005	0.027
	(0.010)	(0.015)	(0.015)	(0.008)	(0.011)	(0.011)	(0.013)	(0.012)	(0.017)	(0.015)	(0.016)	(0.019)
Teacher-rep. char.	0.014	0.125***	0.023	0.057**	0.142***	-0.008	0.064**	0.134***	-0.039	0.023	0.108***	0.003
	(0.030)	(0.035)	(0.040)	(0.026)	(0.033)	(0.033)	(0.029)	(0.031)	(0.033)	(0.030)	(0.034)	(0.033)
School-rep. beha.	-0.047***	0.110***	0.020	0.002	0.105***	0.006	0.023	0.119***	-0.032***	0.051***	0.096***	-0.020
	(0.014)	(0.011)	(0.012)	(0.014)	(0.012)	(0.010)	(0.015)	(0.013)	(0.012)	(0.017)	(0.014)	(0.014)
GPA	0.008	0.314***	0.040*	0.015	0.279***	0.029	0.024	0.241***	0.008	0.042*	0.059*	0.038
	(0.016)	(0.028)	(0.020)	(0.017)	(0.030)	(0.023)	(0.018)	(0.032)	(0.022)	(0.023)	(0.031)	(0.024)
Observations	22,662			22,905			23,266			22,688		
Clusters	188			194			190			186		

The table presents the heterogeneous impacts (in SD) on the summary indices 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 and EJ*F their interaction. In Panel B, EJ gives the impacts for financial aid students, nFA the differences between non-financial aid and financial aid students and EJ*nFA their interaction. Regressions are controlled for school and cohort fixed effects. Standard errors are robust to heteroscedasticity and clustered at the school*cohort level. For the interaction terms and the EJ impacts, we also interact the school and cohort fixed effects with the variable of heterogeneity (F in panel A and nFA in panel B). The sample is composed of 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	$\mathbb{1}_{g>p50}$	$\text{EJ}^* \mathbb{1}_{g>p50}$	EJ	$\mathbb{1}_{g>p50}$	$\text{EJ}^* \mathbb{1}_{g>p50}$	EJ	$\mathbb{1}_{g>p50}$	$\text{EJ}^* \mathbb{1}_{g>p50}$	EJ	$\mathbb{1}_{g>p50}$	$\text{EJ}^* \mathbb{1}_{g>p50}$
Panel A: GPA												
Return to effort	.	.	.	0.029**	0.018	0.029*	0.045***	0.026	0.010	0.051***	0.088***	-0.007
	.	.	.	(0.012)	(0.018)	(0.017)	(0.015)	(0.018)	(0.020)	(0.015)	(0.016)	(0.021)
Stud-rep. dilig.	0.010	0.093***	0.002	-0.017*	0.065***	-0.009	0.027**	0.070***	-0.051***	0.007	0.044***	-0.001
	(0.011)	(0.014)	(0.014)	(0.009)	(0.012)	(0.012)	(0.013)	(0.015)	(0.016)	(0.014)	(0.015)	(0.018)
Teacher-rep. char.	0.006	0.608***	0.034	0.027	0.504***	0.042	0.037	0.450***	0.010	0.040	0.401***	-0.001
	(0.032)	(0.030)	(0.034)	(0.030)	(0.034)	(0.034)	(0.033)	(0.034)	(0.039)	(0.034)	(0.032)	(0.039)
School-rep. beha.	-0.077***	0.331***	0.070***	0.003	0.317***	0.002	-0.011	0.291***	0.008	0.059***	0.296***	-0.043***
	(0.018)	(0.017)	(0.015)	(0.019)	(0.016)	(0.016)	(0.020)	(0.017)	(0.015)	(0.021)	(0.022)	(0.016)
GPA	0.014	1.504***	0.013	0.014	1.345***	0.007	-0.007	1.269***	0.033	0.050*	1.037***	0.010
	(0.017)	(0.018)	(0.017)	(0.019)	(0.022)	(0.021)	(0.024)	(0.023)	(0.026)	(0.026)	(0.023)	(0.021)
	EJ	$\mathbb{1}_{b>p50}$	$\text{EJ}^* \mathbb{1}_{b>p50}$	EJ	$\mathbb{1}_{b>p50}$	$\text{EJ}^* \mathbb{1}_{b>p50}$	EJ	$\mathbb{1}_{b>p50}$	$\text{EJ}^* \mathbb{1}_{b>p50}$	EJ	$\mathbb{1}_{b>p50}$	$\text{EJ}^* \mathbb{1}_{b>p50}$
Panel B : Behavior												
Return to effort	.	.	.	0.046***	0.014	0.009	0.027*	-0.026	0.040**	0.020	0.023	0.043**
	.	.	.	(0.012)	(0.016)	(0.017)	(0.014)	(0.017)	(0.017)	(0.013)	(0.018)	(0.020)
Stud-rep. dilig.	0.020	0.087***	-0.012	-0.037***	0.044***	0.028**	-0.006	0.043***	0.010	-0.016	0.038**	0.039*
	(0.012)	(0.015)	(0.015)	(0.009)	(0.012)	(0.012)	(0.014)	(0.016)	(0.017)	(0.016)	(0.018)	(0.020)
Teacher-rep. char.	-0.001	0.395***	0.078**	0.082**	0.383***	-0.042	0.023	0.278***	0.063	0.038	0.270***	0.020
	(0.032)	(0.031)	(0.037)	(0.033)	(0.039)	(0.041)	(0.035)	(0.029)	(0.041)	(0.038)	(0.037)	(0.043)
School-rep. beha.	-0.039**	0.551***	0.021	0.011	0.371***	0.006	0.007	0.339***	-0.011	0.069***	0.328***	-0.057***
	(0.017)	(0.018)	(0.016)	(0.020)	(0.016)	(0.017)	(0.022)	(0.018)	(0.019)	(0.023)	(0.022)	(0.019)
GPA	0.006	0.703***	0.073***	0.012	0.676***	0.056**	-0.018	0.622***	0.103***	0.031	0.517***	0.073***
	(0.022)	(0.024)	(0.023)	(0.020)	(0.026)	(0.024)	(0.026)	(0.028)	(0.030)	(0.026)	(0.025)	(0.027)
Observation	21,114			18,825			17,465			15,772		
Clusters	183			185			184			180		

The table presents the heterogeneous impacts (in SD) on summary indices by baseline GPA and Behavior score. GPA heterogeneity is computed using $\mathbb{1}_{g>p50}$ 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 $\mathbb{1}_{b>p50}$ which takes 1 when the student scored in the top half of the behavior score at baseline (i.e. Grade 6 first quarter). Regressions are controlled for school and cohort fixed effects and standard errors are robust to heteroscedasticity and clustered at the school*cohort level. For EJ and interaction columns, we interact the school and cohort fixed effect with the heterogeneity variable ($\mathbb{1}_{b>p50}$ or $\mathbb{1}_{g>p50}$). The sample is composed of students with baseline GPA and behavior scores.

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

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	0.284	-0.004 (0.010)	-0.016 (0.014)	-0.036** (0.018)	0.026 (0.021)	-0.004 (0.013)	0.019 (0.016)	-0.000 (0.017)
Medium skill	5,372	0.201	0.021*** (0.007)	0.011 (0.010)	0.091*** (0.017)	0.019 (0.017)	0.044*** (0.012)	0.037** (0.015)	-0.045** (0.017)
Low skill	5,372	0.337	-0.022* (0.012)	-0.004 (0.013)	-0.049** (0.019)	-0.035* (0.020)	-0.050*** (0.018)	-0.054** (0.021)	0.049** (0.021)
No aspiration	5,372	0.178	0.006 (0.009)	0.009 (0.011)	-0.006 (0.015)	-0.011 (0.016)	0.010 (0.012)	-0.002 (0.015)	-0.004 (0.016)
Educational Aspirations									
Academic High School	5,497	0.685	0.018 (0.011)	-0.000 (0.014)	0.084*** (0.019)	0.040* (0.021)	0.021 (0.017)	0.064*** (0.023)	0.001 (0.020)
Technical High School	5,497	0.310	-0.020* (0.011)	0.002 (0.013)	-0.080*** (0.018)	-0.046** (0.021)	-0.028* (0.017)	-0.070*** (0.024)	0.010 (0.020)
Repeat G9	5,497	0.005	0.002 (0.002)	-0.001 (0.003)	-0.005 (0.003)	0.006* (0.003)	0.008*** (0.002)	0.005* (0.003)	-0.011*** (0.003)
Index									
Aspiration Score (SD)	5,497	-0.001	0.035* (0.021)	-0.001 (0.026)	0.154*** (0.037)	0.080* (0.041)	0.059* (0.031)	0.119*** (0.040)	-0.040 (0.036)
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. Under *Full Sample*, Column *Obs* gives the number of observation, *Control* the average in the control group and *Impact* the result of the regression of the outcomes on the treatment variable. Under *Gender Heterogeneity*, column EJ gives the treatment effect for male students, F the difference between female and male students and EJ their interaction. Under FA heterogeneity, EJ is the for financial aid students, nFA the difference between non-financial aid and financial aid students and EJ*nFA their interaction. All regressions are controlled for cohort and school fixed effect. For the heterogeneity analysis, EJ and the interaction columns also control for the interaction between cohort and school fixed effect and the heterogeneity variable (nFA or F). Standard errors are robust to heteroscedasticity and clustered at the school*cohort level. The sample is restricted to the students with gender and financial status information.

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

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

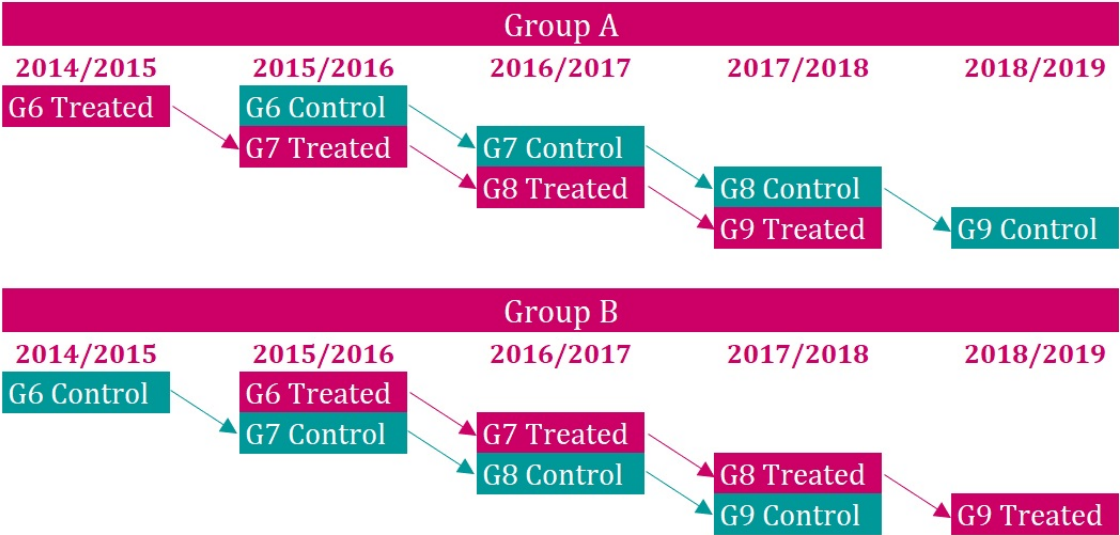
	Full Sample			GPA heterogeneity			Behavior heterogeneity		
	Obs.	C	Impact	EJ	$\mathbb{1}_{g>p_{50}}$	$\text{EJ}^* \mathbb{1}_{g>p_{50}}$	EJ	$\mathbb{1}_{b>p_{50}}$	$\text{EJ}^* \mathbb{1}_{b>p_{50}}$
Professional Aspirations									
High skill	3,932	0.292	-0.005 (0.012)	0.020 (0.019)	0.049 (0.077)	-0.038 (0.025)	0.007 (0.020)	0.059*** (0.022)	-0.022 (0.026)
Medium skill	3,932	0.205	0.029*** (0.008)	0.024* (0.013)	-0.096*** (0.017)	0.010 (0.020)	0.034*** (0.013)	0.034** (0.017)	0.002 (0.019)
Low skill	3,932	0.315	-0.029** (0.014)	-0.043* (0.023)	-0.141*** (0.031)	0.026 (0.028)	-0.019 (0.020)	-0.075*** (0.022)	-0.037 (0.027)
No aspiration	3,932	0.188	0.005 (0.011)	-0.000 (0.014)	0.188*** (0.046)	0.002 (0.020)	-0.022 (0.016)	-0.018 (0.021)	0.058** (0.024)
Educational Aspirations									
Academic High School	3,995	0.722	0.018 (0.013)	0.057*** (0.021)	0.283*** (0.060)	-0.060** (0.027)	-0.012 (0.020)	0.103*** (0.023)	0.068*** (0.025)
Technical High School	3,995	0.275	-0.020 (0.013)	-0.065*** (0.021)	-0.280*** (0.065)	0.069*** (0.026)	0.006 (0.020)	-0.103*** (0.023)	-0.060** (0.025)
Repeat G9	3,995	0.004	0.002 (0.002)	0.008*** (0.003)	-0.003 (0.006)	-0.010*** (0.004)	0.005* (0.003)	-0.000 (0.003)	-0.007** (0.003)
Index									
Aspiration Score (SD)	3,995	0.053	0.042* (0.024)	0.107*** (0.040)	0.261*** (0.035)	-0.097* (0.052)	0.027 (0.036)	0.205*** (0.038)	0.051 (0.047)
Observations	4,190	2,047							
Clusters	158	79							

Same as above Table 6 but with heterogeneous impacts by baseline GPA and baseline behavior score, using the same definitions as in Table E2. The sample is restricted to students with baseline GPA and baseline behavior score.

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

Appendix A: Evaluation Design

Figure A1: Evaluation design



Appendix B: External Validity

Table B1: Comparison between EJ sample and other French Schools - 1/2

	EJ sample	EJ sample - PE schools	EJ sample - Public schools	EJ sample - French schools
Panel A: Schools' characteristics				
Public school (%)	1.000	0.000 (0.000)	0.000 (0.000)	0.243*** (0.044)
Private school (%)	0.000	0.000 (0.000)	0.000 (0.000)	-0.243*** (0.044)
Priority education (%)	0.794	-0.206*** (0.013)	0.598*** (0.041)	0.646*** (0.036)
Number of students/school	528.052	18.751 (22.294)	32.531* (19.670)	49.309** (21.958)
Number of teachers/school	46.082	1.893 (1.486)	7.159*** (1.315)	7.916*** (1.536)
Number of students/class	23.189	0.638*** (0.194)	-1.743*** (0.254)	-2.022*** (0.312)
Remoteness index	-2.072	-1.057*** (0.409)	-2.243*** (0.305)	-2.102*** (0.290)
Panel B: Teachers' characteristics				
Years of service	5.893	-0.062 (0.226)	-1.382*** (0.221)	-1.254*** (0.218)
< than 2 years in school (%)	0.349	0.006 (0.013)	0.049*** (0.011)	0.050*** (0.012)
Part-time (%)	0.057	-0.004 (0.005)	-0.031*** (0.006)	-0.025*** (0.006)
Female (%)	0.596	0.001 (0.009)	-0.045*** (0.009)	-0.053*** (0.009)
Permanent (%)	0.910	0.007 (0.012)	-0.031*** (0.007)	0.012 (0.013)
<i>Agrégés</i> (%)	0.043	-0.000 (0.004)	-0.009* (0.005)	0.001 (0.005)
Turnover (%)	0.182	-0.007 (0.012)	0.027** (0.012)	0.027** (0.012)

The table tests the differences between our experimental sample of middle schools and the population of French middle schools on school aggregate outcomes. Column *EJ sample* provides the average for the experimental sample. Column *EJ sample - PE schools*, *EJ sample - public schools* and *EJ sample - French schools* compares the schools in our sample with respectively schools in Priority Education Zone (PE schools), Public Schools and all french middle schools. The remoteness index is computed using the proportion of students living in a rural commune, the range of courses offered around the school, the distance of the school from sports and cultural facilities (the higher the index value, the more remote a school is). *Agrégés* are qualified teachers who have passed the high-level competitive *Aggregation* exam in France. Standard errors of the regressions are reported in parenthesis.

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

Table B2: Comparison between EJ sample and other French Schools - 2/2

	EJ sample	EJ sample - PE schools	EJ sample - public schools	EJ sample - French schools
Panel C: Students' characteristics				
G6 held back at least once (%)	0.117	-0.001 (0.006)	0.041*** (0.005)	0.046*** (0.005)
G6 National Test	229.571	0.046 (0.091)	-1.018*** (0.095)	-1.183*** (0.103)
G9 National Test	7.044	-0.085 (0.080)	-1.090*** (0.090)	-1.305*** (0.102)
Female (%)	0.488	-0.003 (0.003)	-0.003 (0.003)	-0.001 (0.005)
With financial aid (%)	0.494	-0.017 (0.015)	0.202*** (0.016)	0.239*** (0.016)
Living in a deprived area (%)	0.421	0.054* (0.029)	0.320*** (0.020)	0.333*** (0.018)
Social position index	0.828	0.024** (0.011)	-0.175*** (0.016)	-0.207*** (0.017)
High SES (%)	0.168	0.026*** (0.008)	-0.127*** (0.015)	-0.168*** (0.018)
Low SES (%)	0.782	-0.015 (0.011)	0.118*** (0.015)	0.156*** (0.017)
<i>Post-G9 enrollment:</i>				
Academic high school (%)	0.593	0.021* (0.011)	-0.047*** (0.011)	-0.065*** (0.013)
Technical high school (%)	0.350	-0.009 (0.010)	0.071*** (0.010)	0.096*** (0.011)
G9 held back (%)	0.027	0.002 (0.003)	0.006*** (0.002)	0.005* (0.002)

Same note as Table B1. For the (non-standardized) outcomes *GPA G6* and *GPA G9* which are the average scores obtained in Mathematics and French at standardized national tests administered at the beginning of Grade 6 and at the end of Grade 9, respectively, coefficients are standardized so that they are expressed in terms of standard deviations of the mean measured for all French public middle schools. The social position index is created from data with a detailed description of the family environment: parental education, income, cultural practices, housing conditions, and so on (the higher the index value, the more the students' family background is favourable to learning).

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

Appendix C: Content of the Indices and Sub-Indices

Table C1: Sources and Items included in the Indices and Sub-Indices - 1/2

Indice & sub-indices	Items	Source
GPA	Average grade French, Maths, Physics, Biology, History, Geography, Sports, Foreign languages, Technical skills	Administrative data
Perceived Return to Effort		
<i>Locus of control</i>	Prob. success if from poor neighborhood	Guyon and Huillery (2020)
	Gap poor/wealthy neighborhood	<i>idem</i>
	Prob. success if parents without degree	<i>idem</i>
	Gap with and without college degree	<i>idem</i>
<i>Growth Mindset</i>	Intelligence is something that can't be changed	Claro et al. (2016)
	You can learn new things, not your intelligence	<i>idem</i>
	I prefer problems that I'll learn a lot from...	Li and Bates (2017)
	Prob. success if gifted but do not study hard	Guyon and Huillery (2020)
	Prob. success if study regularly	<i>idem</i>
	Prob. success if who is under-performing	<i>idem</i>
Teacher-reported character		
<i>Achievement</i>	I finish whatever I begin	Park et al. (2017)
	I come to class prepared	<i>idem</i>
	I work independently with focus	<i>idem</i>
	I get to work right away rather than procrastinating	<i>idem</i>
	I remembers and follow directions	<i>idem</i>
	I believe that effort will improve my future	<i>idem</i>
	I try very hard even after experiencing failure	<i>idem</i>
	I actively listen to others	<i>idem</i>
<i>Intelligence</i>	I show enthusiasm	<i>idem</i>
	I invigorate others	<i>idem</i>
	I actively participate	<i>idem</i>
	I ask and answer questions to deepen understanding	<i>idem</i>
	I am eager to explore new things	<i>idem</i>
<i>Social</i>	I keep my temper in check	<i>idem</i>
	I remain calm even when criticized or otherwise provoked	<i>idem</i>
	I get over frustrations and setbacks quickly	<i>idem</i>
	I am polite to adults and peers	<i>idem</i>
	I demonstrate respect for the feelings of others	<i>idem</i>
	I allow others to speak without interruption	<i>idem</i>
	I am able to find solutions during conflicts with others	<i>idem</i>
	I recognize and show appreciation for other	<i>idem</i>
	I recognize and show appreciation for my opportunities	<i>idem</i>
School-reported behavior		
	1/2 day absenteeism	Administrative data
	number of recorded lateness	<i>idem</i>
	disciplinary actions	<i>idem</i>
	sanctions	<i>idem</i>

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

Table C2: Sources and Items included in the Indices and Sub-Indices - 2/2

Indice & sub-indices	Items	Source
Self-reported diligence		
<i>Orderliness</i>	Is organized and neat	Goldberg (1990)
	Works carefully, takes one's time to get things right	<i>idem</i>
	Plans things ahead, thinks before acting	<i>idem</i>
	Is reliable, shows up on time	<i>idem</i>
<i>Grit</i>	New ideas and projects sometimes distracting	Angela Lee Duckworth and Quinn
	Setbacks are not discouraging, doesn't give up	<i>idem</i>
	Is obsessed with projects but quickly loses interest	<i>idem</i>
	Is obsessed with projects but quickly loses interest	<i>idem</i>
	Is a hard worker	<i>idem</i>
	Sets a goal then changes goal	<i>idem</i>
	Has problems focusing on long-term projects	<i>idem</i>
	Finishes whatever is started	<i>idem</i>
	Is diligent, never gives up	<i>idem</i>
<i>School-work impul.</i>	Forgets things needed for class	Tsukayama et al. (2013)
	Interrupts other students	<i>idem</i>
	Says rude things	<i>idem</i>
	Loses things because messy (eg, desk, bedroom)	<i>idem</i>
	Loses temper at home or at school	<i>idem</i>
	Forgets what teacher instructed	<i>idem</i>
	Lets mind wander instead of listening	<i>idem</i>
	Talks back to teacher or parent when upset	<i>idem</i>
<i>Work discipline</i>	Is always prepared	Goldberg et al. (2006)
	Gets chores done right away	<i>idem</i>
	Start tasks right away	<i>idem</i>
	Gets to work at once	<i>idem</i>
	Carries out plans	<i>idem</i>
	Wastes my time	<i>idem</i>
	Finds it difficult to get down to work	<i>idem</i>
	Needs a push to get started	<i>idem</i>
	Has difficulty starting tasks	<i>idem</i>
<i>Homework Manag.</i>	Postpones decisions	<i>idem</i>
	Locates the materials needed for homework	Xu and Wu (2013)
	Finds a quiet area	<i>idem</i>
	Removes things from table	<i>idem</i>
	Makes enough space to work	<i>idem</i>
	Turns off the TV	<i>idem</i>
	Sets priority and plans ahead	<i>idem</i>
	Keeps track of what remains to be done	<i>idem</i>
	Reminds oneself of the remaining time	<i>idem</i>
	Tells oneself to work more quickly when late	<i>idem</i>
	Finds ways to make homework more interesting	<i>idem</i>
	Praises oneself for good work/effort	<i>idem</i>
	Reassures oneself about one's abilities	<i>idem</i>
	Tries not to be bothered with previous mistakes	<i>idem</i>
	Tries to pay attention to what needs to be done	<i>idem</i>
	Tries to calm down	<i>idem</i>
	Tells oneself that one can do it	<i>idem</i>
	Daydreams during homework	<i>idem</i>
	Starts conversations unrelated to work	<i>idem</i>
	Plays around during homework	<i>idem</i>
	Stops homework to eat or drink	<i>idem</i>
<i>hours of homework</i>	Time spent on homework yesterday?	Authors
	Time spent on homework day before yesterday?	<i>idem</i>

Appendix D: Balance Checks

Table D1: Baseline Balancing 1/3 - Administrative Data Sample

	G6 sample		G7 sample		G8 sample		G9 sample	
	C	Impact	C	Impact	C	Impact	C	Impact
Panel A: Full Sample								
Date of Birth	2003.30	-0.000	2003.29	0.004	2003.27	0.006	2003.26	-0.004
	[0.664]	(0.005)	[0.667]	(0.005)	[0.690]	(0.005)	[0.715]	(0.006)
Female	0.489	0.004	0.492	0.005	0.489	0.008	0.488	0.008
	[0.500]	(0.005)	[0.500]	(0.006)	[0.500]	(0.005)	[0.500]	(0.006)
Was held back	0.185	-0.001	0.173	-0.004	0.168	-0.005	0.156	-0.006
	[0.389]	(0.004)	[0.378]	(0.004)	[0.374]	(0.004)	[0.363]	(0.004)
Financial aid	0.504	0.013	0.520	-0.028***	0.476	0.012	0.444	-0.002
	[0.500]	(0.008)	[0.500]	(0.008)	[0.499]	(0.008)	[0.497]	(0.005)
Single parent family	0.178	-0.007	0.173	-0.007	0.182	-0.008**	0.196	-0.011
	[0.382]	(0.005)	[0.378]	(0.004)	[0.386]	(0.004)	[0.397]	(0.008)
Blue collar family	0.766	-0.013***	0.761	-0.009**	0.758	-0.012***	0.744	-0.013***
	[0.424]	(0.005)	[0.426]	(0.004)	[0.429]	(0.004)	[0.436]	(0.004)
Foreigner	0.125	0.003	0.134	-0.002	0.150	-0.003	0.140	0.003
	[0.331]	(0.004)	[0.341]	(0.003)	[0.357]	(0.003)	[0.347]	(0.005)
Baseline GPA	13.177	0.019	13.310	0.024	13.346	0.028	13.475	0.047
	[2.674]	(0.039)	[2.589]	(0.040)	[2.571]	(0.038)	[2.527]	(0.039)
Baseline behavior (-)	-0.011	0.022*	-0.009	0.021	-0.008	0.010	0.001	0.002
	[0.637]	(0.012)	[0.614]	(0.013)	[0.622]	(0.013)	[0.618]	(0.013)

The table gives the differences between the treatment and the control group on outcomes not supposed to be affected by the EJ program. Under column *C*, we provide the average in the control group, under *Impact* the coefficient of the regression of the outcome on the treatment variable, controlled for school and cohort fixed effects. (-) indicates that the value was inverted so that a higher value means a better outcome. Standard errors of the regressions, given under the impact coefficients, are robust to heteroscedasticity and clustered at the school*cohort level.

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

Table D2: Baseline Balancing 2/3 - Student Survey Sample

	G6 sample		G7 sample		G8 sample		G9 sample	
	C	Impact	C	Impact	C	Impact	C	Impact
Panel B: Student reported survey								
Date of Birth	2003.44	-0.016*	2003.31	0.011*	2003.26	0.011	2003.31	0.015
	[0.631]	(0.008)	[0.655]	(0.006)	[0.649]	(0.009)	[0.694]	(0.010)
Female	0.497	0.005	0.489	0.018*	0.496	0.000	0.508	-0.006
	[0.500]	(0.010)	[0.500]	(0.011)	[0.500]	(0.012)	[0.500]	(0.011)
Was held back	0.166	0.014*	0.158	-0.008	0.160	-0.017**	0.148	-0.011
	[0.372]	(0.007)	[0.365]	(0.006)	[0.367]	(0.008)	[0.356]	(0.008)
Financial aid	0.522	0.013	0.518	-0.039***	0.484	0.018*	0.474	0.006
	[0.500]	(0.011)	[0.500]	(0.011)	[0.500]	(0.010)	[0.499]	(0.011)
Single parent family	0.178	-0.002	0.163	-0.012*	0.186	-0.021**	0.186	-0.023**
	[0.383]	(0.008)	[0.370]	(0.006)	[0.389]	(0.008)	[0.389]	(0.011)
Blue collar family	0.762	0.007	0.756	-0.006	0.756	-0.014*	0.730	-0.003
	[0.426]	(0.007)	[0.430]	(0.007)	[0.430]	(0.008)	[0.444]	(0.008)
Foreigner	0.123	0.002	0.128	0.000	0.136	0.001	0.121	0.020***
	[0.328]	(0.007)	[0.334]	(0.006)	[0.342]	(0.006)	[0.327]	(0.007)
Baseline GPA	13.348	-0.006	13.473	0.030	13.503	0.026	13.720	0.058
	[2.508]	(0.060)	[2.537]	(0.059)	[2.521]	(0.064)	[2.420]	(0.069)
Baseline behavior (-)	-0.050	0.037**	-0.043	0.009	-0.051	0.015	-0.044	0.000
	[0.524]	(0.017)	[0.570]	(0.015)	[0.532]	(0.018)	[0.542]	(0.017)

Same as D1 for the sub-sample composed of students who completed the Student Survey Sample (i.e. the survey that corresponds to the Return to effort or the self-reported diligence score.

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

Table D3: Baseline Balancing 3/3 - Teacher Survey Sample

	G6 sample		G7 sample		G8 sample		G9 sample	
	C	Impact	C	Impact	C	Impact	C	Impact
Panel C: Teacher reported survey								
Date of Birth	2003.42	-0.015	2003.31	0.029***	2003.24	0.022*	2003.33	-0.003
	[0.631]	(0.010)	[0.649]	(0.007)	[0.643]	(0.012)	[0.694]	(0.012)
Female	0.496	0.020	0.489	0.025**	0.492	0.017	0.515	0.007
	[0.500]	(0.012)	[0.500]	(0.011)	[0.500]	(0.015)	[0.500]	(0.012)
Was held back	0.168	0.015*	0.163	-0.020***	0.166	-0.029***	0.147	-0.006
	[0.374]	(0.009)	[0.369]	(0.007)	[0.372]	(0.011)	[0.354]	(0.009)
Financial aid	0.523	-0.001	0.524	-0.037***	0.482	0.038***	0.479	0.004
	[0.500]	(0.013)	[0.500]	(0.011)	[0.500]	(0.011)	[0.500]	(0.013)
Single parent family	0.178	-0.002	0.154	0.000	0.184	-0.019*	0.189	-0.024**
	[0.383]	(0.009)	[0.361]	(0.008)	[0.387]	(0.011)	[0.391]	(0.009)
Blue collar family	0.761	0.004	0.756	0.008	0.756	-0.019*	0.739	-0.001
	[0.426]	(0.009)	[0.430]	(0.008)	[0.429]	(0.010)	[0.439]	(0.010)
Foreigner	0.128	0.002	0.129	-0.004	0.141	0.002	0.117	0.025***
	[0.334]	(0.008)	[0.336]	(0.007)	[0.349]	(0.008)	[0.322]	(0.009)
Baseline GPA	13.326	0.003	13.462	0.055	13.473	0.016	13.699	0.091
	[2.502]	(0.071)	[2.516]	(0.066)	[2.540]	(0.075)	[2.421]	(0.091)
Baseline behavior (-)	-0.039	0.028	-0.030	0.005	-0.037	0.013	-0.038	-0.000
	[0.539]	(0.021)	[0.597]	(0.015)	[0.560]	(0.022)	[0.561]	(0.020)

Same as D1 for the sub-sample composed of students whose teachers have completed the teacher survey. Teacher survey corresponds to the Teacher-reported character index.

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

Appendix E: Stacked dataset Impacts

Table E1: Heterogeneous Effects by Gender and Financial Aid Status - Stacked dataset

	Full Sample			Gender			Financial Aid		
	Obs	Control	Impact	EJ	F	EJ*F	EJ	nFA	EJ*nFA
Participation	22,147	0.121	0.765*** (0.012)	0.744*** (0.013)	-0.015** (0.006)	0.042*** (0.007)	0.741*** (0.013)	-0.031*** (0.007)	0.045*** (0.009)
Return to effort	17,008	-0.000	0.041*** (0.005)	0.024*** (0.007)	0.027*** (0.009)	0.033*** (0.010)	0.030*** (0.008)	-0.018* (0.010)	0.025** (0.011)
Student-rep. dilig.	23,167	0.000	-0.006 (0.005)	0.002 (0.006)	-0.008 (0.008)	-0.016* (0.008)	-0.010 (0.007)	-0.001 (0.007)	0.006 (0.009)
Teacher-rep. char.	18,419	0.001	0.039*** (0.012)	0.040*** (0.015)	0.117*** (0.017)	0.003 (0.018)	0.046*** (0.014)	0.128*** (0.018)	-0.009 (0.017)
School-rep. behav.	89,273	0.002	0.003 (0.008)	-0.005 (0.010)	0.154*** (0.009)	0.016* (0.009)	0.006 (0.009)	0.108*** (0.008)	-0.006 (0.007)
GPA	81,269	0.005	0.037*** (0.012)	0.011 (0.014)	0.323*** (0.017)	0.051*** (0.018)	0.020 (0.013)	0.228*** (0.026)	0.028* (0.016)
Observations	91,521								
Clusters	194								

The table presents the impacts of the treatment on summary indices using the stacked dataset. Indices are presented in rows. Under *Full Sample*, we present the overall results on the full sample. Under *Gender*, we look at the heterogeneity by gender. *EJ* is the impact of the program on male students, column *female* is the difference between female and male students, and *female*EJ* is the interaction term capturing the additional impact for female students. In the second set of columns, aid recipients are defined as students who receive need-based financial aid (FA). Column *EJ* is the impact of the program for non-recipients, column *FA* the difference between non-recipients and aid recipients students and *EJ*FA* is the interaction term capturing the additional impact for aid recipients. Columns *Obs* gives the number of observations when the four years are stacked together. Row Observations and clusters is the total number of observations and clusters in the stacked dataset. All regressions are controlled for cohort and school fixed effect. For the heterogeneity analysis, EJ and the interaction columns also control for the interactions between cohort and school fixed effects and the heterogeneity variable (nFA or F).

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

Table E2: Heterogeneous Effects by Baseline GPA and Baseline Behavior - Stacked dataset

	Full Sample			GPA			Behavior		
	Obs	Control	Impact	EJ	$\mathbb{1}_{g>p50}$	$\text{EJ}^*\mathbb{1}_{g>p50}$	EJ	$\mathbb{1}_{b>p50}$	$\text{EJ}^*\mathbb{1}_{b>p50}$
Participation	18,033	0.123	0.771*** (0.013)	0.731*** (0.014)	-0.030*** (0.007)	0.073*** (0.008)	0.745*** (0.013)	-0.021** (0.009)	0.045*** (0.010)
Return to effort	13,254	-0.002	0.043*** (0.006)	0.038*** (0.009)	0.041*** (0.010)	0.008 (0.013)	0.031*** (0.009)	0.006 (0.010)	0.022* (0.011)
Student-rep. dilig.	18,812	-0.004	-0.001 (0.005)	0.007 (0.007)	0.068*** (0.008)	-0.014* (0.008)	-0.011 (0.007)	0.053*** (0.008)	0.018** (0.009)
Teacher-rep. char.	14,955	0.014	0.040*** (0.014)	0.025 (0.017)	0.499*** (0.018)	0.028 (0.019)	0.037* (0.019)	0.332*** (0.020)	0.026 (0.022)
School-rep. behav.	71,950	0.025	-0.006 (0.008)	-0.014 (0.013)	0.310*** (0.013)	0.016 (0.011)	0.006 (0.013)	0.406*** (0.015)	-0.005 (0.012)
GPA	68,316	0.070	0.026** (0.012)	0.013 (0.015)	1.313*** (0.017)	0.014 (0.015)	0.008 (0.018)	0.641*** (0.021)	0.068*** (0.021)
Observations	73,177	35,810							
Clusters	189	119							

Same as E1 for baseline GPA and behavior score and for a sample restricted to students with a baseline GPA and behavior score. For the columns under *GPA*, high-achievers are defined as students in the top half of the GPA distribution in Grade 6 first quarter i.e. at baseline. EJ is the impact of the program on low-achievers, $\mathbb{1}_{g>p50}$ the difference between high-achievers and low-achievers, and $\mathbb{1}_{g>p50}^*\text{EJ}$ is the interaction term capturing the additional impact for high-achievers. For columns under *Behavior*, well-behaved students are defined as students in the top half of the behavior score distribution during Grade 6 first quarter i.e. at baseline. EJ is the impact of the program on poorly-behaved, $\mathbb{1}_{b>p50}$ the difference between well-behaved and poorly-behaved, and $\mathbb{1}_{b>p50}^*\text{EJ}$ is the interaction term capturing the additional impact for well-behaved students.

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

Appendix F: Impacts on Sub-Indices

Table F1: Sub-Indices Impacts by Gender - Stacked dataset

	Obs.	EJ	F	EJ*F
Return to effort	17,008	0.024*** (0.008)	0.027*** (0.009)	0.033*** (0.010)
... <i>Locus of Control</i>	17,008	0.046*** (0.015)	0.083*** (0.015)	0.011 (0.018)
... <i>Growth Mindset</i>	16,984	0.022** (0.009)	0.004 (0.010)	0.038*** (0.011)
Student-rep. dilig.	23,167	0.003 (0.007)	-0.008 (0.008)	-0.016* (0.008)
... <i>Orderliness</i>	22,903	0.015 (0.022)	0.146*** (0.021)	-0.027 (0.022)
... <i>Grit</i>	23,159	-0.023* (0.013)	-0.017 (0.016)	-0.029* (0.018)
... <i>Work impulsivity (-)</i>	23,040	0.028** (0.012)	0.050*** (0.013)	-0.026* (0.014)
... <i>Work discipline</i>	22,446	-0.003 (0.011)	-0.028** (0.011)	-0.013 (0.012)
... <i>Homework manag.</i>	17,356	-0.017 (0.014)	-0.061*** (0.014)	-0.012 (0.016)
... <i>hours of homework</i>	22,858	0.024 (0.022)	0.011 (0.023)	-0.037 (0.025)
Teacher-rep. char.	18,419	0.035** (0.017)	0.117*** (0.017)	0.003 (0.018)
... <i>Intelligence</i>	18,419	0.031 (0.021)	0.088*** (0.023)	0.024 (0.024)
... <i>Achievement</i>	18,416	0.037* (0.020)	0.304*** (0.021)	0.003 (0.022)
... <i>Social</i>	18,417	0.051*** (0.019)	0.273*** (0.018)	-0.021 (0.020)
School-rep. behav.	89,273	-0.005 (0.012)	0.154*** (0.009)	0.016* (0.009)
... <i>1/2 day absenteeism (-)</i>	88,916	-0.031* (0.017)	0.074*** (0.013)	0.024* (0.014)
... <i>lateness (-)</i>	88,146	-0.042 (0.026)	0.133*** (0.019)	0.040** (0.020)
... <i>disciplinary actions (-)</i>	87,922	-0.009 (0.025)	0.257*** (0.024)	0.022 (0.015)
... <i>sanctions (-)</i>	87,995	-0.006 (0.030)	0.340*** (0.024)	0.019 (0.018)
Observations	91,521			
Clusters	194			

The table presents the heterogeneous treatment impacts by gender on summary indices and sub-indices using the stacked dataset. (-) indicates that the value was inverted so that a higher value means a better outcome. All regressions include cohort and school fixed effect. EJ and EJ*F also include the interactions between the cohort and school fixed effects and the gender variable.

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

Table F2: Sub-Indices Impacts by FA Status - Stacked dataset

	Obs.	EJ	nFA	nFA*EJ
Return to effort	17,008	0.029*** (0.009)	-0.018* (0.010)	0.025** (0.011)
... <i>Locus of Control</i>	17,008	0.041** (0.017)	-0.088*** (0.019)	0.023 (0.019)
... <i>Growth Mindset</i>	16,984	0.030*** (0.010)	0.010 (0.010)	0.022* (0.013)
Student-rep. dilig.	23,167	-0.010 (0.008)	-0.001 (0.007)	0.006 (0.009)
... <i>Orderliness</i>	22,903	-0.004 (0.021)	0.016 (0.018)	0.009 (0.021)
... <i>Grit</i>	23,159	-0.030** (0.014)	0.003 (0.014)	-0.019 (0.016)
... <i>Work impulsivity (-)</i>	23,040	0.006 (0.013)	-0.007 (0.013)	0.018 (0.016)
... <i>Work discipline</i>	22,446	-0.016* (0.009)	0.002 (0.009)	0.009 (0.012)
... <i>Homework manag.</i>	17,356	-0.031** (0.014)	-0.008 (0.014)	0.018 (0.017)
... <i>hours of homework</i>	22,858	0.001 (0.019)	0.023 (0.018)	0.008 (0.022)
Teacher-rep. char.	18,419	0.043*** (0.016)	0.128*** (0.018)	-0.009 (0.017)
... <i>Intelligence</i>	18,419	0.049** (0.020)	0.150*** (0.023)	-0.013 (0.022)
... <i>Achievement</i>	18,416	0.040* (0.020)	0.171*** (0.023)	0.004 (0.022)
... <i>Social</i>	18,417	0.043** (0.019)	0.144*** (0.021)	0.006 (0.020)
School-rep. behav.	89,273	0.007 (0.010)	0.108*** (0.008)	-0.006 (0.007)
... <i>1/2 day absenteeism (-)</i>	88,916	-0.022 (0.016)	0.098*** (0.016)	0.006 (0.011)
... <i>lateness (-)</i>	88,146	-0.017 (0.022)	0.184*** (0.014)	-0.006 (0.014)
... <i>disciplinary actions (-)</i>	87,922	0.004 (0.019)	0.110*** (0.014)	-0.003 (0.012)
... <i>sanctions (-)</i>	87,995	0.013 (0.022)	0.165*** (0.017)	-0.013 (0.016)
Observations	91,521			
Clusters	194			

The table presents the heterogeneous impacts on summary indices and sub-indices by socioeconomic background using the stacked dataset. Aid recipients are defined as in Table E1. (-) indicates that the values was inverted so that a higher value means a better outcome. The regressions include school and cohort fixed effect. The standard errors, given below in parenthesis, are clustered at the school*cohort level and are robust to heteroscedasticity.

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

Table F3: Impacts on Sub-Indices by Baseline GPA

	Obs.	EJ	$\mathbb{1}_{g>p50}$	$EJ^* \mathbb{1}_{g>p50}$
Return to effort	13,254	0.040*** (0.009)	0.041*** (0.010)	0.008 (0.013)
... <i>Locus of Control</i>	13,254	0.044** (0.019)	-0.061*** (0.018)	0.018 (0.022)
... <i>Growth Mindset</i>	13,235	0.040*** (0.011)	0.080*** (0.012)	0.009 (0.014)
Student-rep. dilig.	18,812	0.005 (0.008)	0.068*** (0.008)	-0.014* (0.008)
... <i>Orderliness</i>	18,630	0.001 (0.023)	0.307*** (0.021)	-0.020 (0.023)
... <i>Grit</i>	18,807	-0.037** (0.015)	0.213*** (0.016)	-0.002 (0.017)
... <i>Work impulsivity (-)</i>	18,732	0.023 (0.015)	0.042*** (0.014)	-0.004 (0.014)
... <i>Work discipline</i>	18,271	0.001 (0.012)	0.133*** (0.013)	-0.030** (0.013)
... <i>Homework manag.</i>	13,506	-0.006 (0.016)	0.102*** (0.016)	-0.032* (0.019)
... <i>hours of homework</i>	18,595	0.011 (0.023)	0.064** (0.025)	0.002 (0.024)
Teacher-rep. char.	14,955	0.025 (0.020)	0.499*** (0.018)	0.028 (0.019)
... <i>Intelligence</i>	14,955	0.029 (0.024)	0.718*** (0.023)	0.024 (0.025)
... <i>Achievement</i>	14,953	0.014 (0.025)	0.797*** (0.023)	0.030 (0.026)
... <i>Social</i>	14,954	0.032 (0.022)	0.557*** (0.020)	0.016 (0.022)
School-rep. behav.	71,950	-0.017 (0.016)	0.310*** (0.013)	0.016 (0.011)
... <i>1/2 day absenteeism (-)</i>	71,776	-0.053** (0.022)	0.365*** (0.019)	0.023 (0.015)
... <i>lateness (-)</i>	70,855	-0.078** (0.037)	0.422*** (0.025)	0.066** (0.027)
... <i>disciplinary actions (-)</i>	71,045	-0.015 (0.035)	0.317*** (0.035)	0.020 (0.020)
... <i>sanctions (-)</i>	70,665	-0.027 (0.039)	0.482*** (0.029)	0.039 (0.028)
Observations	73,177			
Clusters	189			

The table presents the heterogeneous impacts on summary indices and sub-summary indices by baseline GPA using the stacked dataset. High-achievers are defined as in Table E2. (-) indicates that the value was inverted so that a higher value means a better outcome. The regressions also include school and cohort fixed effect. The standard errors, given below in parenthesis, are clustered at the school*cohort level and are robust to heteroscedasticity.

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

Table F4: Impacts on Sub-Indices by Baseline Behavior

	Obs.	EJ	$\mathbb{1}_{b>p50}$	$EJ^*\mathbb{1}_{b>p50}$
Return to effort	13,254	0.034*** (0.010)	0.006 (0.010)	0.022* (0.011)
... <i>Locus of Control</i>	13,254	0.037** (0.018)	-0.051** (0.020)	0.032 (0.020)
... <i>Growth Mindset</i>	13,235	0.035*** (0.012)	0.021* (0.012)	0.023* (0.013)
Student-rep. dilig.	18,812	-0.009 (0.008)	0.053*** (0.008)	0.018** (0.009)
... <i>Orderliness</i>	18,630	0.011 (0.024)	0.268*** (0.020)	-0.013 (0.023)
... <i>Grit</i>	18,807	-0.044*** (0.014)	0.124*** (0.015)	0.030* (0.017)
... <i>Work impulsivity (-)</i>	18,732	0.014 (0.015)	0.068*** (0.013)	0.022 (0.015)
... <i>Work discipline</i>	18,271	-0.015 (0.011)	0.088*** (0.011)	0.013 (0.013)
... <i>Homework manag.</i>	13,506	-0.024 (0.015)	0.076*** (0.015)	0.009 (0.018)
... <i>hours of homework</i>	18,595	-0.013 (0.024)	0.049** (0.024)	0.042 (0.027)
Teacher-rep. char.	14,955	0.037* (0.021)	0.332*** (0.020)	0.026 (0.022)
... <i>Intelligence</i>	14,955	0.022 (0.026)	0.307*** (0.025)	0.060** (0.027)
... <i>Achievement</i>	14,953	0.028 (0.024)	0.496*** (0.024)	0.035 (0.026)
... <i>Social</i>	14,954	0.048** (0.022)	0.453*** (0.022)	0.007 (0.025)
School-rep. behav.	71,950	0.006 (0.016)	0.406*** (0.015)	-0.005 (0.012)
... <i>1/2 day absenteeism (-)</i>	71,776	-0.037 (0.023)	0.511*** (0.020)	0.023 (0.017)
... <i>lateness (-)</i>	70,855	-0.039 (0.040)	0.563*** (0.033)	0.026 (0.028)
... <i>disciplinary actions (-)</i>	71,045	0.015 (0.034)	0.362*** (0.036)	-0.015 (0.023)
... <i>sanctions (-)</i>	70,665	0.004 (0.043)	0.537*** (0.031)	0.014 (0.036)
Observations	73,177			
Clusters	189			

The table presents the heterogeneous impacts on aggregated and sub-aggregated outcomes baseline behavior using the stacked dataset. Well-behaved students are defined as in Table E2. (-) indicates that the value was inverted so that a higher value means a better outcome. The regressions also include school and cohort fixed effect. The standard errors, given below in parenthesis, are clustered at the school*cohort level and are robust to heteroscedasticity.

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

Appendix G: Main Outcomes Dynamics throughout Middle School - Control Group

Figure G1: Evolution of GPA (standardized on Grade 6)

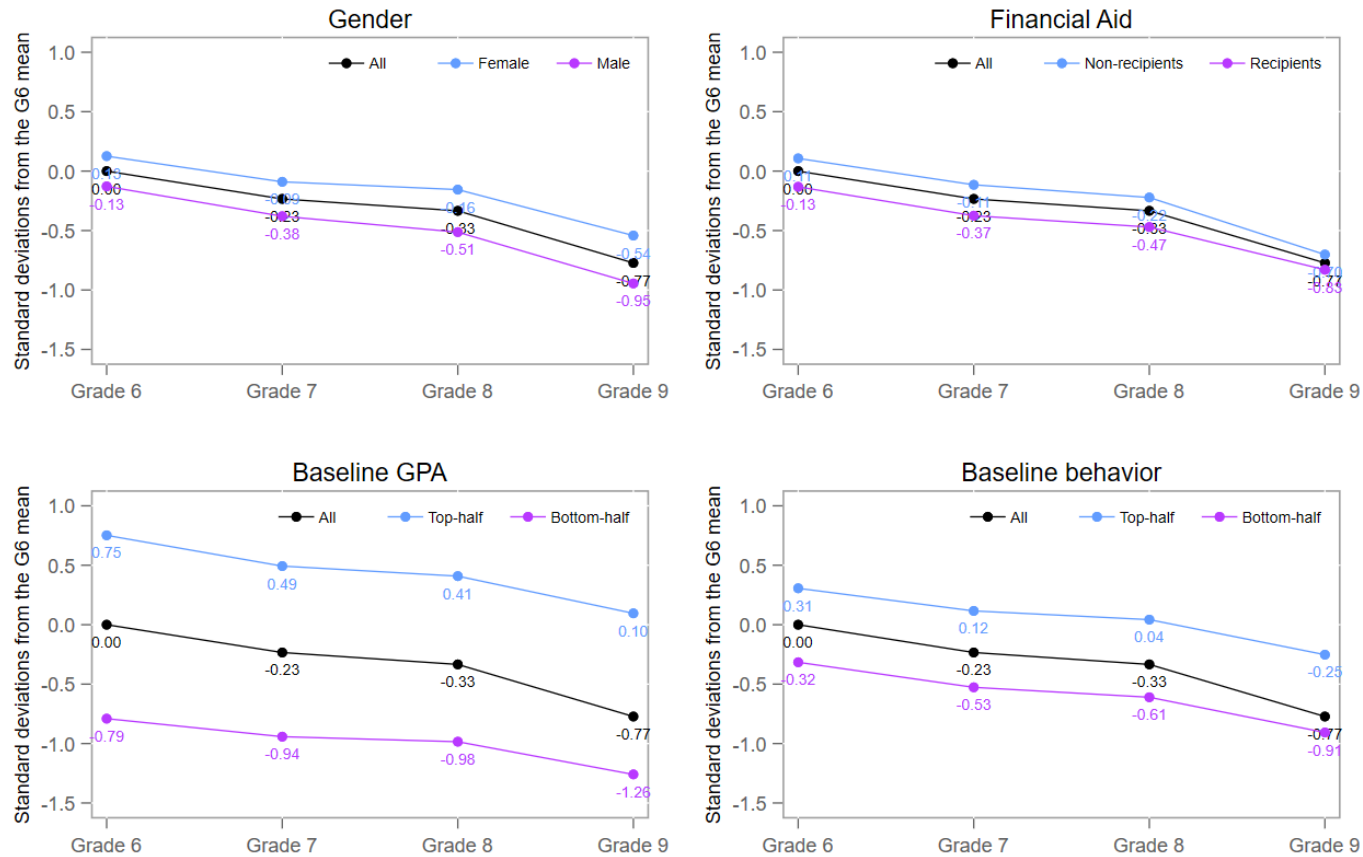


Figure G2: Evolution of return to effort (standardized on Grade 7)

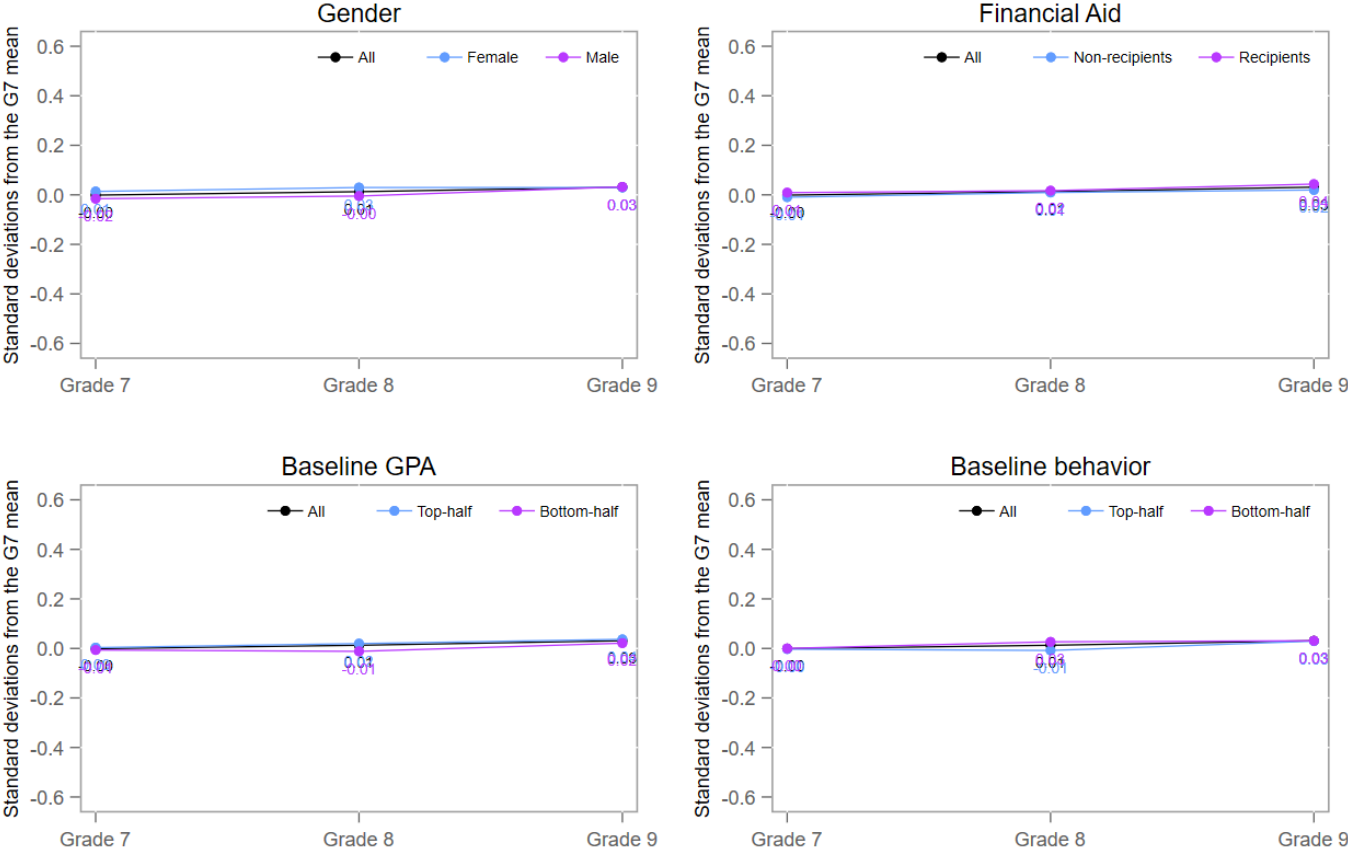


Figure G3: Evolution of student-reported diligence (standardized on Grade 6)

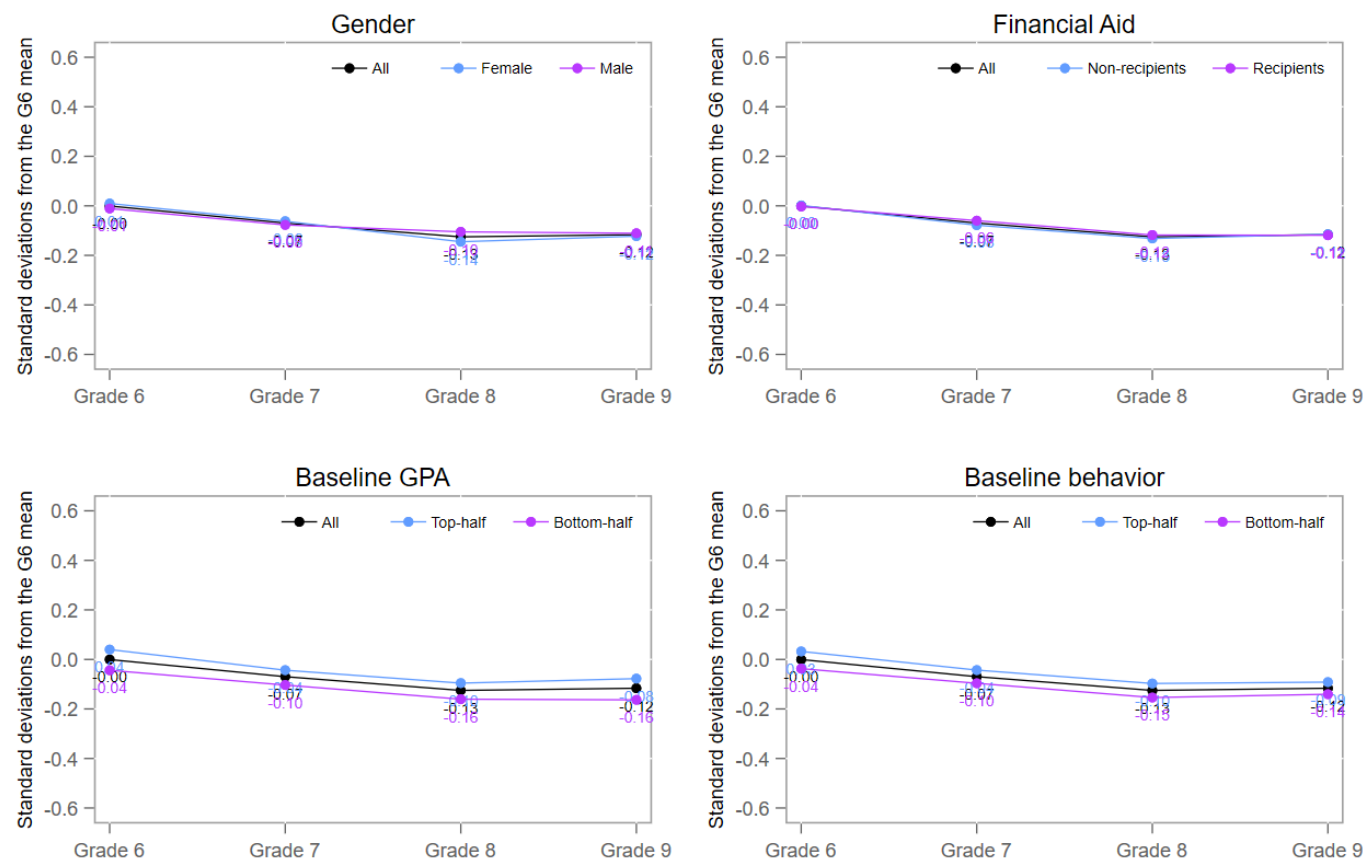


Figure G4: Evolution of teacher-reported character (standardized on Grade 6)

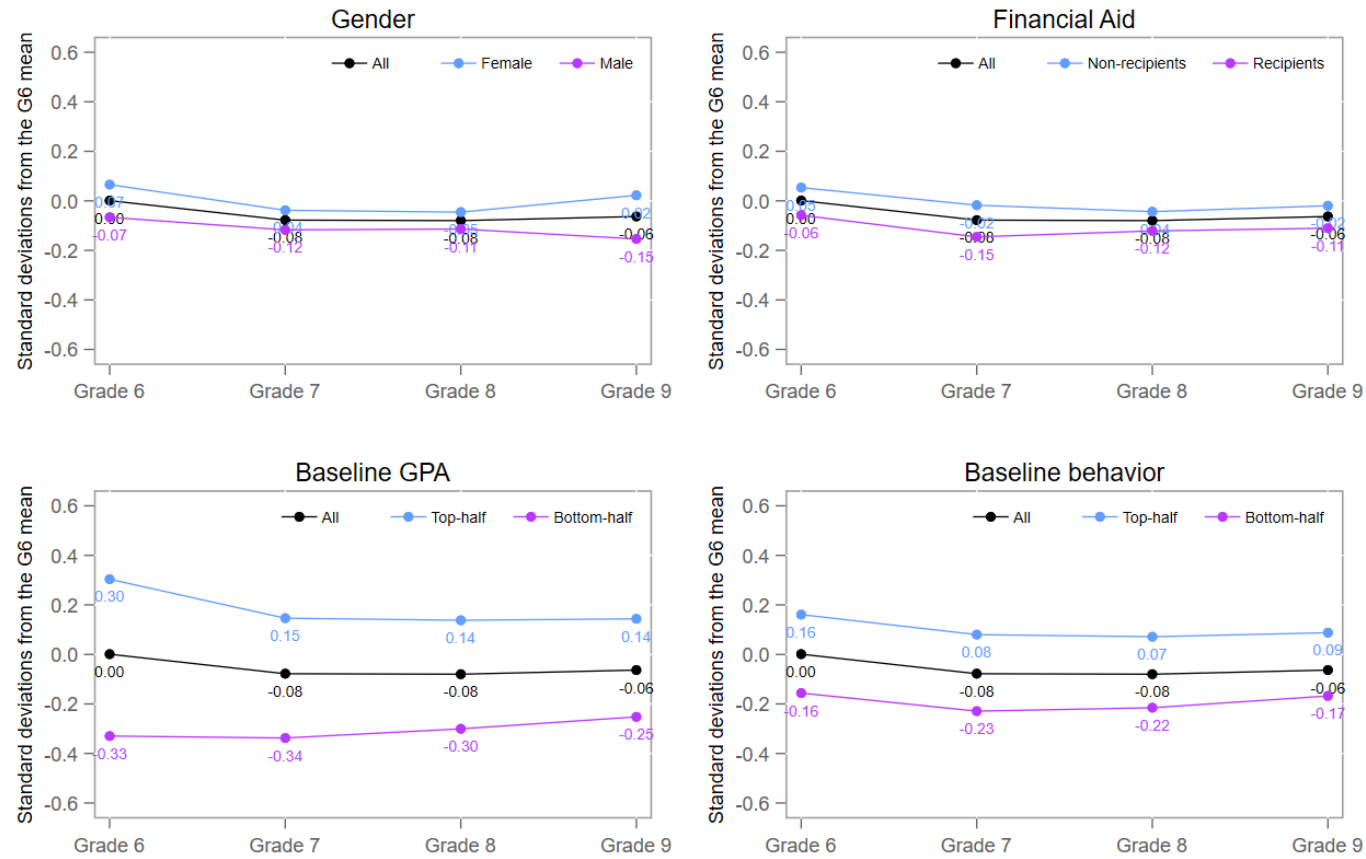


Figure G5: Evolution of school-reported behavior (standardized on Grade 6)

