Getting Parents Involved:
A Field Experiment in Deprived Schools.[1]

Francesco Avvisati, Marc Gurgand, Nina Guyon, Eric Maurin[2]

Paris School of Economics

[1] This research was supported by a grant from the French Experimental Fund for the Youth. We are very grateful for the support of the schools and administrative teams from the *rectorat* de Créteil, and particularly to Bénédicte Robert. We thank the many J-Pal Europe research assistants that worked on this project. We also thank seminar participants at LSE (London), TSE (Toulouse), PSE (Paris) GREQAM (Marseille) and the University of Mannheim for useful comments, as well as discussants and participants at SOLE/EALE conference (London), ESPE conference (Essen) and EEA congress (Glasgow).
[2] Corresponding author. PSE, 48, Boulevard Jourdan, 75014 Paris (France). Eric.maurin@ens.fr.

Abstract:

This paper presents a randomized field experiment conducted in French middle schools located in a deprived educational district. Parents in test groups were invited to participate in a program of parent-school meetings on how to improve their involvement in their children's education. The experiment follows a "partial population" design, whereby volunteer parents are listed at the beginning of the year and treatment classes are then randomized. As only volunteer parents attend meetings, this provides for robust identification of both the overall effect of the program on the beneficiaries and spillover effect on the untreated.

At the end of the school-year, we find that treated families effectively increased their school- and home-based involvement activities. Children of families who were directly targeted by the program developed more positive behavior and attitudes in school, and had less literacy problems. Importantly, for all behavioral outcomes we find large spillover effects of the program on classmates of treated families. This experiment proves that schools are able to increase parents' awareness and that parental inputs have strong effects on pupil behavior. Our results on spillovers demonstrate that similar initiatives can be effective even in case of low parental take-up of the program.

Keywords: Parental Involvement, Cluster Randomized Trial, Classroom Peer-Effects, Child support

JEL Classification: I21, J13, J18

I. Introduction

Middle schools in modern societies face the challenge of providing basic skills to very heterogeneous populations. The problems of truancy, violence, and pupil indiscipline are epidemic, especially in deprived urban areas. After spending three or four years in middle schools, many pupils are still far from reaching the basic requirements of curricula (OECD, 2010). In this context, the view that better informed and more involved parents could contribute to overcome many difficulties enjoys a large consensus. Local initiatives abound, and plans to foster parental involvement are already eligible to federal funding in the US ("No Child left Behind" Act) and part of the national education policy in the UK ("Every Child Matters" Green Paper).

Yet, there is still very little evidence on whether such policies make a difference. It is not clear whether involvement policies conducted by schools can effectively increase parents' participation in education-related activities, especially among the most disadvantaged. It is not even clear whether improved parental involvement has any positive effect on pupils' behavior. The most involved parents differ from the less involved across many observed and unobserved dimensions and it is far from obvious that the empirical correlation between parental involvement and pupils' outcomes represents any causal effect at all.

This paper proposes entirely novel evidence on these issues, using a field experiment undertaken in Paris area middle schools. We show that a simple program of parent-school meetings increased the level and quality of school-related parental care; that this improved involvement of parents translated into a significant reduction of truancy and misbehavior in test classes; and that the program ultimately resulted in less literacy problems for children whose parents were invited to the program. Impacts are extremely strong: their order of

magnitude is that of average differences between white-collar and blue-collar families in the control group (where white-collar families represent the top 20% of the population in term of socio-economic status). Moreover, while all actions on parents were limited to those who did participate in the program, we find that the behavior of all pupils was affected, including pupils whose parents did not participate. Robust evidence on spillover effects comes from the "partial population" (Moffitt, 2001) design that was implemented: volunteer parents were listed in all classes prior to class randomization. Non-volunteer parents never attended the program, but their children were randomly exposed to treated classmates.

The experiment started in September 2008, at the beginning of the 2008-2009 academic year, in 37 middle schools of a relatively deprived educational district. In 200 classes, some 1000 parents (22%) of sixth graders agreed to enroll in a program of three debates with the school staff on how to successfully manage the transition from primary school to middle school. The program offered information on the functioning of schools and advice on how to help children with homework. The transition between primary and secondary school represents one of the most critical stages of an educational career and this is why we chose to focus on 6th grade, i.e., the first year of secondary education.

During the enrolment period, schools made it clear to parents that agreement to enroll would not necessarily result in participation and that only a random selection of enrolled parents would be effectively invited to participate in the program. By early October 2008, the enrolment period was closed. Of the 200 classes, 102 were randomly chosen to effectively run the program in November and December 2008. In each school and each class we are therefore able to clearly identify volunteer and non-volunteer families prior to the random decision of running the program. By comparing volunteer families in test classes and volunteer families in control classes, we capture the direct effect on the treated. By comparing their children's classmates, we capture the indirect effect of having treated families within the class. Finally,

by comparing all pupils and families in test classes with all pupils and families in control classes, this randomization design is able to capture in a simple way the equilibrium impact of this program, under the assumption that there are no spillovers across classes.

The program had a positive impact on school-related involvement activities of enrolled families. For instance, the proportion of enrolled parents which actively participate in parents' organization at their school is 35% in test classes, whereas it is only of 24% in control classes. On aggregate, these differences are of the same magnitude (0.27 standard deviations) as the differences between white-collar families and blue-collar families observed in the control sample (0.35 standard deviations). Our results therefore secure that schools have a critical ability to influence parental attitudes and behaviors.

As a consequence, the behavior of pupils was affected along many dimensions. We find that the program is associated, by the end of the school year, with a decline by about 0.09 standard deviations in truancy in test classes, and a decline of the same magnitude in the probability of being sanctioned. At the same time, indicators of positive attitudes and behavior improved, thus signaling that the impact is not limited to the lower end of the distribution. Everything seems to indicate the existence of a causal link between parents' involvement and their pupils' behavior. Furthermore, improvements in pupils' behavior are not limited to children of enrolled families: the effects spread out with almost the same average magnitude to their classmates, and especially to the most exposed to disciplinary problems. These facts contradict the view that involvement policies are bound to benefit to a small fraction of volunteer families only.

Finally, we find that pupils in test classes are more able to master the easier reading exercises at the end of the year: over the school-year, children of enrolled families in test classes gained 0.21 standard deviations, and their classmates 0.08 standard deviations, over their counterparts in control classes.

Our paper lies within the scope of several different strands of the literature. First of all, we contribute to the economic literature on the importance of parental inputs for children's education. The few existing studies in this field within the economics discipline all adopt a structural approach, using survey data.[3] Most studies make use of NLSY panel data: Todd and Wolpin (2007) emphasize the preeminent role of "home inputs" relative to "school inputs"; Cunha and Heckman (2008) extend the analysis to distinguish the effect of parents' involvement on cognitive and non-cognitive skills. With a narrower focus on parents' supervision after school, Aizer (2004) or Welsch and Zimmer (2008) quantify its impact with different fixed-effect strategies. The identification of causal links and impacts in these papers has to rely on model assumptions about the form of the education production function. To the best of our knowledge, our paper is one of the first to provide large scale experimental evidence on the potential benefits of parental involvement for children's success up to late childhood.

We also contribute to the debate over the policy levers that can be used to improve pupils' behavior and performance. Parental attitude and involvement at school are widely perceived as key inputs, but little is known on whether such inputs can effectively be manipulated through simple policy initiatives.[4] As it turns out, they are rooted in parents' own past and belong to the private sphere. Our study constitutes one of the very few social experiments demonstrating that such inputs can be significantly upgraded through simple participation programs and that such initiatives have a strong potential for reducing indiscipline in young teenagers.

[3] For a recent survey, see Avvisati, Besbas and Guyon (2010). To the best of our knowledge, the only related experimental evaluation has been conducted by psychiatrists in four London elementary schools (Scott, O'Connor and Futh., 2006). Its focus is on a parenting program for five- and six-years olds.

[4] Desforges and Abouchaar (2003, p.5) note that "evaluations of interventions [in the area of parental involvement] are so technically weak that it is impossible on the basis of publicly available evidence to describe the scale of the impact on pupils' achievement".

Our paper is also related to the large and still growing literature on social interactions and spillover effects in education (see e.g., Hoxby 2000; Angrist and Lang, 2004). Specifically we provide new insights on how the school context can influence the behavior of pupils. We show that an early intervention at the parents' level at the beginning of the school year translates into a progressive improvement of their own pupils' behavior and performance, with a maximum improvement at the end of the year. This result suggests that pupils have indirectly benefited from the parental treatment all over the year (through repeated interactions with their parents) and that this increasingly large "dose" of family interactions has contributed to a progressive modification of their behavior at school. Also, our specific clustered randomized design makes it possible to separately identify spillover effects on non-volunteer parents (and on non-volunteer pupils) without any parametric assumptions, such as linear-in-means models. Most interestingly, we find no spillover effects on non-volunteer parents, but strong spillover effects on their children, especially on behavioral outcomes and, again, with peak improvement by the end of the school year. Overall, our results are consistent with the assumption that the initial treatment has first influenced the attitude of children of volunteer parents through repeated family interactions, which has in turn progressively influenced the attitude of children of non-volunteer parents through repeated classroom interactions all over the school year. To the best of our knowledge, this social experiment is the first to provide such a decomposition of the working of social interactions across parents and pupils in the same school. It improves our understanding of how educational policies can exploit spillover effects and social interactions to enlarge the number of beneficiaries beyond the small fraction that volunteers to participate.

Finally, the paper contributes to the ongoing debate on community and user empowerment policies in western societies. As it happens, many developed countries are faced with the problem of an increasingly fragmented urban landscape with increasing

disparities between poor and rich neighborhoods. Within this context, it is often argued that enclaves of social exclusion deserve special policies, relying on much greater involvement of local communities and inhabitants. Our paper provides new experimental evidence on the outcome of increasing local residents' involvement in one key public service (education), in the context of a poor urban district of a western country.[5]

The remainder of this paper proceeds as follows. Section 2 provides background information on the context in which the program took place and describes the interventions and its objectives. Section 3 introduces outcome measures, experimental design and estimation strategy, performs balancing tests on baseline data and discusses take-up and attrition issues. Section 4 presents the main results of the study. Section 5 discusses the results. Section 6 concludes with implications for policy and future research.

II. The Program

A. Institutional context

The French state-run educational system is highly centralized with schools having limited autonomy. All schools are required to complete the same national curriculum and teachers are civil servants, selected through national examinations, who all hold the same qualifications. After 5 years of elementary school, children enter middle schools at the age of 11. There is no streaming by ability across schools, and French parents are not free to choose the state school that their children will attend. In middle schools each subject is taught by a different teacher. For sixth graders, a typical week consists of 29 school hours, distributed across 9 different subjects, and, hence, different teachers.

[5] In developing countries, there is a similar debate on whether community empowerment policies would also result in reducing corruption and better public services, with mixed empirical evidence (Banerjee et al. 2008, Bjorkman & Svensson, 2009).

Pupils stay in the same class throughout the school year, and in every subject. The class is therefore a very distinct and closed entity; interactions with children of other classes are very limited. Classes are groups of 20 to 30 pupils. Each class has one specific *professeur principal* (reference teacher), two parental delegates and two elected representatives of pupils. Within each school, a *conseil de classe* (class council) is formed for each class, composed by all teachers, and representatives of the school administration. Parental delegates and pupils' representatives are also allowed to attend the meetings of the *conseil de classe* at the end of each term, with an informative role only.

The year is divided into three terms. At the end of each term, the *conseil de classe* meets to discuss each student's work, achievement, and behavior. The *conseil* bestows honors and disciplinary warnings that are transmitted to families, together with teacher grades, on the report card; at the end of the year, the *conseil* decides about grade repetition, and about the optional courses that each pupil will be allowed to take in the future. Indeed, only the best students are allowed to take the optional courses which are considered prestigious (Latin, Greek, additional hours in Chinese, German or English, etc…). Through these decisions, teacher assessments have a lasting influence on later tracking decisions (general versus occupational tracks) which are taken at the end of middle school (9^{th} grade).

B. Participants and information campaign

The experiment took place in the educational district of *Créteil*, which includes all suburbs located to the east of Paris. The district covers an area of approximately 6400 km² and 4 millions inhabitants. This mostly urban and suburban area has the highest density of immigrant populations in France (according to the 1999 census, 20.9% of the population were first-generation immigrants, born outside Metropolitan France) and includes some of the most deprived areas of the Paris region.

The academic year begins in September. Over the summer before the start of academic year 2008-2009, the heads of 37 state-run middle schools from the district volunteered to participate in the experimental study. Out of the 37 middle schools which entered the study, 21 are located in an "educational priority zone" – a label that distinguishes historically deprived areas.[6] Experimental schools also have lower than average pass-rates at the national examination that takes place at the end of middle school ("*brevet des collèges*"): in 2008, the pass rate is 72% against a national average of 83%. Many families attending these schools are relatively poor.

Just after the start of academic year, during September 2008, experimental schools advertised the program to the families of their 6th-graders. The universe to which the program was offered, and baseline and follow-up data were collected, consists of 37 schools, 215 classes, and the families of some 5000 pupils enrolled at these schools.

School heads mainly used a standardized leaflet to inform families of the program. They were also encouraged to contact the families who are the less familiar with the school system, through directed phone calls, or by taking advantage of their demands for allowances. The program was presented as an outreach effort, distinct from usual parent-teacher meetings. The school would organize a series of three evening meetings/debates with parents of 6th graders to help them understand the role of each member of the educational community, the schools' organization, and to help them develop positive involvement attitudes towards their children's school education. It was always explicit that actual eligibility to the program would occur only conditional on a random draw which would select eligible classes. The leaflet explicitly stated that the experimental nature of the program implied a limitation on the number of classes which could benefit from the program. By mid-October each school listed all families who signed up, and closed the registration phase. This list defines the population

[6] This label distinguishes 874 middle schools nationwide, out of a total of over 5000 state-run schools, with a national rate (17%) three times smaller than the rate observed in our experimental set of schools.

of what we call "volunteer families", and has not been amended thereafter. Volunteer families constitute approximately one fifth of the total population (1056 out of 5017).

Overall, the initial information campaign defined very clearly two distinct populations within each school and each class: volunteers and non-volunteers. In substance, volunteers are the fraction of parents who are the most receptive to the policy under consideration; in the absence of any evaluation study, we would expect a high take-up for the program among volunteers and no take-up by non-volunteers. The ability to evaluate separately the effect of the program on volunteers and non-volunteers is one of the very attractive features of our experimental design. Within each school, classes with at least one volunteer were eligible to random assignment to the treatment and control arms (200 out of 215 classes). The draw took place immediately after the end of the registration phase at the school level, and the school direction informed volunteer families in the randomly selected classes about the exact calendar of the program.

C. Interventions

The experimental program consists mainly of a sequence of three meetings/debates which take place every two to three weeks, between November and December (early January in some cases).[7] Sessions start at 6pm at the school. To introduce each session, the school head can draw on precise guidelines, designed by the districts' educational experts. School heads are also invited to project excerpts from a specially conceived DVD introducing the main actors in middle schools, and what is at stake in this stage of education. Both local and district-level documents are distributed at these meetings, explaining the functioning and opportunities of the school attended by their child.

[7] The program was named "*La Mallette des Parents*" (the parents' schoolbag). An official description of the program can be found at http://www.ac-creteil.fr/jahia/Jahia/site/rectoratCreteil/lang/fr/mallette-des-parents (in French; accessed in January 2010).

The two initial sessions of the program focus on how parents can help their children and involve at school and at home with their education. The last session takes place after the first *conseil de classe* (class council) and end-of-term report card. It offers parents advice on how to adapt to the results of the first term. Parents are encouraged to ask questions, explain their problems and share their own experience.

The district-level guidelines insist that the facilitator of the debate should develop the following arguments in discussions. (a) All parents can help their child, no matter what their own school record was and how familiar they are with the institution: what matters most is that children feel that their parents are interested in their school experience, and feel encouraged to talk often about it. (b) To succeed, work in the classroom is not sufficient; homework and regular exercise are extremely important. (c) Parents should regularly scrutinize homework diaries and notebooks, and stay close to children while they repeat their lessons or do exercise. (d) To develop the best attitudes, children must feel that their parents have a good perception and knowledge of the school and that they adhere to the demands of teachers and administration.

At the end of the third session, the principal asks participants whether they would like to participate in additional sessions (a) on parenting issues (in continuity with the first three meetings/debates) or (b) on the use of (school-related) internet or (c) in sessions specifically designed for those who are not fluent in French. These additional sessions include more training elements, and are lead by qualified adult trainers or experts in children development.

Generally speaking, the program and its materials were developed by educational experts at the district level in accordance with state-of-the-art psychological literature on parental involvement (see e.g., Hoover-Dempsey and Sandler 1995, 1997). According to this literature, parental involvement depends on three basic ingredients: (1) Parents become involved in schools if they hold the belief that they should be involved, (2) if they believe that

their involvement can exert a positive influence on children's educational outcomes, and (3) if they perceive that the child and the school want them to be involved. As it happens, the program explicitly increased the level of invitations, and simultaneously raised the opportunities offered by the school to parents in the test group. Also, the topics developed at the meeting insisted on arguments, drawn from role-model and efficacy theories in psychology, about the ways in which involved parents can exert a positive influence on children's achievement. According to these theories parents can increase the quality of the effort exerted by children by giving them interest, attention, praise and rewards related to behaviors that lead to school success.

III. Outcome Measures, Randomization, and Statistical Methods

A. Randomization and Methods

By mid-October, the initial information campaign was closed and a random allocation of classes across test and control groups was implemented. This randomization procedure was carried out separately within each school, in the presence of the school head.[8] The implementation of the random assignment rules resulted in the selection of 102 classes in the test group and 98 classes in the control group (see Figure 1). Volunteer families belonging to test classes were informed by the administration, in the days after randomization was performed, of the exact dates at which the three meetings would take place.

The randomization procedure defines four basic groups of families within each school: volunteers in test classes; non-volunteers in test classes; volunteers in control classes; non-volunteers in control classes. Of these four groups, only volunteers in test classes are effectively invited to the program. The design thus corresponds to a "partial population

[8]A detailed description of the randomization procedure is provided in the working paper version of this paper, see Avvisati, Gurgand, Guyon et Maurin (2010).

experiment" (Moffitt, 2001). Within this framework, any difference in outcomes between volunteers in test and control groups will capture the causal effect of eligibility for the program on the population of volunteers. In addition, any difference in outcomes between non-volunteers in test and control groups will capture the causal effect of having eligible peers on the population of non-volunteers. This will be interpreted as a treatment spillover. Finally, we will also provide estimates of the average equilibrium effect of the program by comparing all pupils in test classes to all pupils in control classes.

To estimate these effects, we use the following statistical models for each outcome Y,

(1) *volunteers:* $Y_{ics} = \alpha^{V} T_c + u_s + v_{ics}$

(2) *non-volunteers:* $Y_{ics} = \alpha^{NV} T_c + n_s + \eta_{ics}$

(3) *all (equal effects):* $Y_{ics} = \alpha T_c + n_s + u_s V_i + \varepsilon_{ics}$

where, for each individual i in class c and school s, the variable V_i is a dummy indicating whether the family of i is a volunteer, and T_c is a dummy indicating whether class c is a test class. Parameters u_s and n_s represent two potentially distinct sets of school fixed effects while variables v_{ics}, η_{ics} and ε_{ics} represents unobserved individual random effects. The parameters of interest are α^{V}, α^{NV} and α. The identifying assumption is that the unobserved random factors and the treatment variable T_c are uncorrelated conditional on the enrollment status V_i. The credibility of this assumption is a direct consequence of the experimental nature of the treatment assignment variable T_c. As discussed below, we do not find any significant correlation between T_c and pupils' observed characteristics for both the group of volunteer and the group of non-volunteer pupils, which is consistent with the identifying assumption. Also, for each regression analysis, we have checked that results are not affected when we include observed baseline characteristics as additional control variables (analysis available on request). Within this framework, parameter α^{NV} is identified as the difference in average outcomes between non-volunteer pupils in treated and untreated classes whereas α^{V} is

identified as the variation across volunteers in treated and untreated classes. Both parameters can be estimated through standard fixed effect OLS procedures. Standard errors are clustered at the class level, and clustered standard errors are used to assess significance of the coefficient on the "test" dummies.

B. Outcome measures: Parents

To assess the impact of the program on parental involvement attitudes, we distributed a short questionnaire to all families at the end of the school year. The questionnaire was distributed in all schools on 11 May 2009 to each family via their children; parents were asked to send it back within a week. The parent questionnaire is a self-administered short questionnaire, with 12 questions on school-based involvement, home-based involvement as well as on parents' perception of the school. Specifically, the questionnaire consisted of 3 questions on school-based involvement (participation in parents' organization – a necessary condition for being a representative – participation in parents/teachers general meetings, individual appointments with teachers), 4 questions on home-based involvement and parental control (help with homework, knowledge of grades, control over time spent watching TV, control over time spent on videogames) and 4 questions on understanding and general perception (knowledge of available optional courses, plans about child's future, satisfaction with school, anxiety about violence). Finally, one question asks whether parents have been summoned to school to discuss their child's behavior. Never being summoned to school may be interpreted both as a symptom of the child's good discipline and a consequence of a proactive partnership with schools.

The answers to these questions define our basic measures for parental involvement. We have also constructed four synthetic scores – a global parenting score, a school-based involvement score, a home-based involvement score and an understanding and perception

score – by applying correspondence analysis to the indicator matrix of all responses: we then computed the position of each parent on the first axis derived from correspondence analysis. Scores are standardized to have mean 0 and standard deviation 1. The global parenting score applies this scoring technique to the 12 questions in the questionnaire; the three other scores apply the same technique to subsections of the questionnaire.

C. Outcome measures: Pupils

Pupils' individual outcomes are mostly measured based on administrative registry data. First, we collected data on "honors" awarded from the *conseil de classe* (about 30% of pupils get honors). In most schools, we have also been able to collect the official "conduct mark" given by the *conseil de classe* to each pupil at the end of each term. This conduct mark has usually a very skewed distribution, with about 30% pupils having either the maximum or next-to-maximum mark (i.e., no behavior problem). With respect to behavior, we also collected data on whether pupils were given an official "disciplinary warning" or were temporarily excluded during each term. Temporary exclusions signal violent behaviors or repeated transgression of the rules. They are sentenced by the school head.

Overall, we defined three dummy variables: a "honors" dummy, taking value 1 if the child had honors, a "good conduct" dummy taking value 1 if the child earned the maximum, or next-to-maximum conduct mark, and a "sanctions" dummy, taking value 1 if the child was punished with an official warning or temporarily excluded during the term. The "honors" dummy identifies high-performance whereas the "good conduct" dummy identifies the absence of problematic behavior during school hours. By contrast, the "sanctions" dummy identifies the most problematic behaviors.

Finally, in 28 schools we could access information on absenteeism. We define our measure of absenteeism as the number of half-days where the child is not at school without a

valid justification from its parents (an occasional hour skipped counts as a half-day if no justification is given). Note that this information is completely independent from teachers' assessment or *conseil de classe* deliberation.

The above-defined variables constitute a rich set of measures about pupils' behavior, coming from independent data sources and reflecting both subjective and objective outcomes. Reassuringly, we have checked that whenever a pupil has a sanction, then the "good conduct" and "honors" dummies are equal to 0. By contrast, for children who are not sanctioned, we found that the "good conduct" (about 25% of pupils) and "honors" dummies (another 25%) were reasonably independent of each other[9]. Overall, the "sanction" variable separates the small proportion of students with heavy conduct problems, and can therefore measure an improvement occurring at the bottom of the distribution of behaviors. By contrast, "good conduct" and "honors" make possible to capture improvements in performance and/or behavior among the better students. Together, these three measures shed light on changes taking place at both ends of the distribution of behaviors. With respect to absenteeism, we find that it decreases sharply from an average of nine half-days per term to one half-day per term when binary indicators of behavior vary from (sanction, conduct, honors)=(1,0,0), to (0,0,0), to (0,0,1) or (0,1,0) and finally to (0,1,1). Absenteeism can thus be considered as providing an independent continuous measure of the quality of pupil behavior.

We were also interested in measuring the impact on children's achievement. We use two sources of information. First, we collected the teacher-given marks reported on end-of-term sheets and transmitted to families after each *conseil de classe*. For each subject and each term, they represent the average mark given by the corresponding teacher. In addition to these teacher marks, we ran two pre- and post-treatment tests (in Math and French) which are identical across schools and classes, and were externally graded.

[9] A detailed statistical analysis of how pupils' outcomes relate to each other is available in the working paper version of this work (Avvisati, Gurgand, Guyon et Maurin, 2010).

D. Baseline data

Baseline data originate from two sources: the administrative database on pupils and families, and a pre-test in French and Mathematics, which, as part of a national evaluation process, took place in September 2008. The administrative database contains information from a registration form collected in July 2008, when parents registered their children for the next school-year (gender, date of birth, family situation, socioeconomic status…). While administrative data are available on all pupils, we were not able to access the pre-test results for one out of 37 schools. In addition, some students were absent on the day the test was taken, which resulted in missing observations on the pre-test.

This data can be used to check that observable characteristics are balanced across treatment and control groups. Comfortingly, Table 1 shows that differences are weak and we can never reject the null that the differences occur by chance at standard levels of significance, including within subpopulations of volunteers and non-volunteers.

Table 1 also shows the basic pre-treatment differences between our samples of volunteer and non-volunteer families (see "Mean" column). Volunteers have slightly more often white-collar occupation than non-volunteer (21% vs 18%) and belong more often to two-parent families (75% vs 72%). But there are no significant differences in pre-treatment test scores between volunteers and non-volunteers. Overall, Table 1 reveals no strong observable pre-treatment differences between the two populations. This fact is maybe a consequence of the principals having tried to inform and attract all categories of parents, even those whose involvement is usually very weak.

E. Take-up

At the beginning of each session of the initial sequence of meetings/debates, we asked participants to sign in and collected the attendance lists. Table 2 shows the effective take-up

rates across the four basic categories of families: volunteers in test classes, non-volunteers in test classes, volunteers in control classes, non-volunteers in control classes. Comfortingly take-up is large and significant for volunteers in test class only (Test/Vol. column), even though it remains far from 100%. Specifically, about 58% of families in this group participated in at least one session, and about 16% attended all the three basic debates. As a result of imperfect compliance, any significant difference between the test and control groups will be driven by a relatively small proportion of actually treated families.

As discussed above, families attending the last meetings in the initial program could determine whether, and in what form, to continue with additional sessions. "Parenting sessions" ended-up being organized in only 17 schools out of 37; 15 schools offered sessions for parents on the use of the internet, and only a handful ran additional sessions for non-French speakers. Overall, the number of families which participated in at least one additional session, beyond the initial three meetings, makes up about 15% of eligible parents (test volunteers): 80 parents participated in about 3 additional debates on parenting issues; 57 participated in additional sessions about the internet (4 sessions on average); and 19 participated in sessions specifically designed for non-French speaking (5 sessions on average). It must be understood, therefore, that additional sessions have got a marginal turnout only. For the vast majority of eligible parents, the program has consisted in participating to initial debates only.

F. Response Rate

The number of observations included in each analysis is only limited by non-response, or more generally unavailability of the information. Response rates for our main outcome measures are presented in Appendix Table A1 and A2.

The parent questionnaire was returned back in due time by approximately two thirds of volunteer and non-volunteer families (Table A1). This proportion is very similar across treatment arms. On the subset of volunteer families, we made a special effort to improve the response rate: all volunteer families which did not return the questionnaire after a week were called, between June 3 and June 10, to answer the questions during a short phone interview. This has increased the response rate for volunteers to about 80%. Non-response acts as a filter on the information flow, and could seriously bias the test-control comparison. In our case, however, non-response is balanced across treatment arms. Also, by performing baseline comparisons again on the sample with observed response, we have checked that the initial balancing properties are still valid even after attrition.

For outcome measures related to pupils' behavior, availability of information varies between 61% (good conduct) and 90% (honors) of the initial sample (Table A2). Attrition here does not stem from intentional behavior, but rather from varying school or, sometimes, class-level practices. Indeed, for all outcomes most attrition is at the school-level (with entire schools missing from our data), or at the class level, and the residual individual-level attrition is at the same level (about 6%) for all four outcomes. School-level attrition does not have the potential to introduce biases in estimation, as randomization was stratified by school. In principle, class-level attrition, and individual-level attrition, might cause more trouble, but we have checked for each outcome that resulting samples remained balanced with respect to baseline characteristics. The residual individual-level attrition can with high probability be attributed to school-migration during the school year, or, in some cases, over the summer preceding the school-year.

Finally, teacher marks in French and Math are available from all 37 schools, and the response rate is about 90% for these outcomes. Post-tests could be conducted at 35 schools; in

the case of post-tests, individual-level attrition due to absenteeism is significant, but the overall response rate is still above 80%.

IV. Parental Involvement, Pupils' Behaviors and Spillovers

In this section, we analyze in turn the effect of the program on parental involvement, pupils' attitude and pupils' performance. For each type of outcome, we provide a separate analysis of the direct effect on volunteers and the indirect effect on non-volunteers.

A. Increases in Parental Involvement

The experimental evidence suggests that the program was successful at significantly improving parental attitudes. Table 3 reveals higher levels of parental involvement by parents in test classes, as well as a better perception and understanding of the school. Families in test classes also declare having less often been summoned to the school for disciplinary reasons.

Table 3 also enables us to draw finer conclusions: the improvement in parental attitudes and better perception of the school seems entirely attributable to volunteer families, those who effectively could attend the program. Among volunteer parents, the difference across treatment arms in levels of institutional involvement equals more than 30% of a standard deviation of our school-based parental involvement score. The standardized effect size for home-based involvement practices is about 10%, and the program increased the parents' perception and understanding of school by almost 20% of a standard deviation of our score. Overall, the program has a very significant effect on school-based involvement and this effect extends, although to a lower extent, to home-based involvement.

A metric in which the magnitude of the results can be assessed is given by the differences across these same dimensions between white-collar families and non white-collar

families in control classes.[10] Table 4 shows that these are of the same magnitude as those created by the program. In other terms, the invitation of the school head created, among volunteer families in different experimental arms, more or less the same difference in levels of parental involvement as those which pre-existed between the 20% of families with higher socio-economic status and the rest of families.

By contrast, having eligible parents in the same class does not affect the involvement of non-eligible families. When we restrict the analysis to non-volunteers, we find positive differences in involvement and perceptions between test and control groups, but they are small and not statistically significant at standard levels.

Appendix Table A3 proposes a more in-depth description of the observed differences in parents' behavior between test and control volunteers, across the 12 original dimensions measured by the questionnaire (from which the synthetic scores are computed). This table confirms, for instance, that volunteers who received invitations to attend the debates asked for more individual appointments with teachers, report to attend more often traditional parent/teacher evenings and meetings organized by parents' associations. Fewer volunteers from test classes allow their children to watch television after 9 pm on weekdays, and so on.

B. Improvement in Pupils' Behavior

Turning to children, the data from the third and last school-term unanimously point to a better quality in children's relation to school in test than in control classes, across the complete range of available measures on behavior and attitudes (Table 5): children in test classes skip less classes (absenteeism is lower by 0.64 half-days), are less likely to be punished for disciplinary reasons (8.5% against 10.9% in control classes), are more likely to get honors (38.5% against 34.2%) and are more likely to earn the top marks for their conduct

[10] Roughly speaking, white-collar families (managers, professors, engineers…) represent the top 20% of the population in term of social status.

(37.4% against 32.6%). In terms of standardized effect size, both for absenteeism and for the global behavior score that resumes information in the three dummy variables, the advantage of the test group over the control group is about 10% of a standard deviation. All of these differences are above the significance thresholds.

The observed difference in absenteeism or in the likelihood of having disciplinary sanctions is of the same magnitude as the difference between children of white-collar families and other children whereas the difference in the likelihood of receiving honors or in the likelihood of receiving the top conduct-mark is about one-third of the difference in this likelihood between white-collar children and children of other socio-economic background. Overall, estimated improvements are large and present at all levels of the distribution of behavior: very bad behavior is less frequent and very good one more frequent. Truancy, that was shown above to form a continuous, independent and objective measure is also strongly affected.

C. Direct Effects and Peer Effects

So far, our estimates capture average differences between all pupils in treated and non-treated classes, i.e. a mix between direct and equilibrium effects of the program. Estimates in the bottom panels of Table 5 compare test volunteers with control volunteers, and test non-volunteers with control non-volunteers.[11] It reveals that the advantage of pupils in test classes is observed among both volunteers and non-volunteers. Absenteeism, for instance, is reduced by almost the same amount among non-volunteers than among volunteers. The same result holds true for the behavioral score. In other words, inviting parents to the meetings has produced a net improvement over the year not just in the behavior of children whose families were effectively invited, but also on all other children belonging to the same classes. The

[11] Remember that only volunteer parents where offered the program in test classes.

existence of such large spillovers provides a major argument in favor of developing that kind of policy, since it extends its benefits to all children, despite directly involving only a fraction of all parents. It provides a tool to increase average outcomes without increasing inequalities between the small fraction of children whose parents volunteer to attend evening meetings at school and all the other children.

D. Improvement in Basic Language Achievement

Through its influence on the perceptions and attitudes of families, or through its effects on the behavior and motivation of pupils, the program could extend its benefits to academic achievement measures. We have two sets of measures on achievement to test this claim: teacher marks, collected at the end of each term, and standardized test scores. Teacher marks are essential in shaping pupils' opportunities; they influence grade retention decisions, future high-school plans, and, in the mid-term, the choice of optional subjects. One issue with this outcome, however, is that teachers can adjust their grading practice to the average level of their pupils. In such a case, the comparison between marks given in treated and control classes provides an estimate of the effect of the program which may be downward biased. For this reason, we also conducted externally set and marked tests, in French and Mathematics, which were taken at the end of the school year; these tests supposedly deliver a more objective measure of academic abilities. Their limit, however, resides precisely in their very objective nature: pupils do not have any true incentive to succeed at these tests, as they do not have any consequence for their future (this is especially true for end-of-the-year tests). Teachers' assessments provide plausibly a better measure of the effect of the program on pupils' extrinsic motivations (i.e., motivations coming from external reward such as good grades and academic success) than external tests.

Both for teacher marks and tests, we have two measures – one at the beginning of the year, one at the end of the year. For teacher marks, our "baseline" measure is the mark given at the end of the first term (December), and the end-line measure is the last term's mark (June). For tests, the baseline measure was given by a national evaluation protocol; the end-line measure is an ad-hoc test, built in strict resemblance to the national evaluation.

Building on this information, we will measure the impact of the program on gains over the school year rather than on end-of-year levels.[12] This is improves precision of the estimates, as persistence is very high for achievement measures.

Table 6 displays the differences between test and control arms in the progress in Mathematics and French as measured by teacher marks and test scores. End-of-year measures are standardized prior to the analysis, so that the differences can be interpreted as standardized effects and estimates can be compared across outcomes. Using teacher marks, we do find a significantly larger progress in French for test pupils, relative to control pupils; the magnitude of the differential is 6.5% of a standard deviation. However, we are not able to measure significant differences in the progresses in Mathematics. Turning to test scores, we do not find evidence of significant differences across treatment arms. When we distinguish, within the French test, the easier tasks (those with the highest success rates) from the remaining tasks, there is some evidence that children in test classes have significantly higher success rates at those items. These exercises, labeled "observation" items, tend to measure the ability to find and exploit explicit indications given in short texts, and do not require writing skills.

In contrast to pupils' behavior, the impacts on achievement gains are different on volunteers and non-volunteers. The bulk of the effects is indeed attributable to large impacts on the subpopulation of volunteers (gains for test volunteers over control volunteers are as

[12] Specifically, the dependent variable is computed as the difference between the standardized end-of-year score or mark and the standardized start-of-year score or mark. In a few cases where the start-of-year score or mark is unobserved, its value is set to 0; two dummy variables (one dummy for cases where the pupils' observation is missing, one dummy for cases where the corresponding measure is missing for the whole school) are added to estimation.

large as +15.1% of a standard deviation for French marks), with small and insignificant spillovers on their classmates. Intuitively, it is easier to adapt behaviors to the conduct of other children in the same class than to catch up academically. This is possibly why spillover effects on behaviors are much more perceptible than on test scores.

The improved success rate on the easiest tasks may suggest that the intervention bore some benefits on the progress made by the weakest children – those who were not able to complete, by the beginning of the year, the easier reading comprehension tasks**.** It could also signal higher effort (rather than ability) levels by these same children: subjective attitudes towards learning tend to form part of teacher assessment as well. These hypotheses are intriguing, but remain conjectural.

E. Robustness and Subgroup Analysis

Table 7 further expands the analysis to allow for different impacts across subgroups. Treatment dummies in this table are fully interacted with group indicators for SES status (white-collar vs. others), pupils' sex (girls vs. boys), or pre-test achievement levels (top, medium and bottom third of the distribution of test scores within each school).

Although for most of the coefficients, we cannot reject homogenous impacts across subgroups, we do find evidence of some significant contrasts. Effects on parental attitudes seem to be more significant for families with higher SES; effects on children's behavior are stronger for boys than for girls; and effects on achievement gains are more important for pupils in the lowest initial ability group.

With respect to parental attitudes and behavior, the impact on the "school perception" score is large and significant for white-collar families only (+.312 of a standard deviation); this same outcome is also positively affected among non-volunteer white-collar families (+.199 of a standard deviation). This outcome reflects primarily the parents' satisfaction with

their child's school. Parents with higher SES status seem to appreciate the fact that the school makes special efforts to involve families, even when these efforts are not specifically targeted to them; in addition, these parents may also be the most aware of the improved classroom ambiance. This is an important result in the context of relatively deprived neighborhoods, where public schools are exposed to competition from private schools, and struggle to retain families with higher SES status.

Improvements in pupil behavior, and particularly the reduction in absenteeism, are mainly driven by boys. For volunteer and non-volunteer boys, the reduction in truancy observed in test classes corresponds to a net decrease of total days absent, whereas for girls – which on average have lower levels of absenteeism – we cannot reject that absenteeism is unaffected. Finally, the evidence on achievement gains points to the fact that volunteer pupils belonging to the bottom third of the ability distribution are responsible for most of the observed improvements in literacy, as measured by French marks or by scores on the easiest test items in French. Overall, the benefits of this parental involvement program are evenly distributed across most subgroups; when this is not the case, benefits seem to be mildly targeted to those groups of children and families which are of greatest policy concern.

V. Program Spillovers and Class-Level Effects

A. Spillover Effects: Placebo or Classroom Interactions?

The evidence presented in the previous section unanimously points to very significant spillover effects of the program across all measured dimensions of pupil behaviour. Also, these spillover effects on pupils do not seem to derive from spillover effects on parents since we do not find any significant difference in parental involvement across non-volunteer parents

in treated and untreated classes. Non-volunteer parents do not seem to have been influenced by the program (nor by treated parents).

Given these facts, one possible interpretation for the improvement in the behaviour of non-volunteer pupils is classroom interactions and peer effects. The influence of peers can trickle along many channels, including direct influence of peers' behaviour on own behaviour or more indirect influence through progressive modification of the context of teaching and learning within the classroom.

An alternative interpretation would link the observed improvements to a change in teachers' attitude towards selected classes that is independent from the programme itself and that affects volunteer and non-volunteer students indifferently. Indeed, telling the staff that some classes were selected and other control could have had, as such, an impact on their behavior towards the selected classes. For example, if teachers want the programme to be a success (because it provides additional resources), they may tend to better assess selected classes, regardless of pupils' true outcomes.

To test for such placebo effects, we have first compared teachers' subjective marks during term 1 with initial test scores. We find that the difference between teachers' marks and externally marked test scores is not significantly different in test and control classes.[13] It suggests that the random selection process has, as such, no effect on teachers' assessment.

In fact, we have also checked that test and control classes do not differ either in honors and sanctions granted at the end of the first trimester by the teachers of the class during the first class council (Table 8). The end of term 1 is in the middle of the sequence of debates, and approximately one month after the assignment lottery took place. This is the moment where the experimental context of the program is most salient to teachers and school staff. There is nonetheless no detectable impact of the program on outcomes at this point of time.

[13] Regression (3), with the difference between Term 1 Marks and initial test scores as the dependent variable, yields non-significant coefficients of -.011 (.124) for French and .007 (.113) for Math.

Effects become perceptible and significant at the end of the school year only, after the non-volunteers in selected classes were exposed to repeated interaction with their treated peers. This constitutes the most direct set of evidence that teacher assessments are not influenced by class assignment status.

There is a second key testable difference between actual peer effects and placebo. Specifically, peer effects are likely to increase with increased level of interactions between volunteer and non-volunteer pupils, whereas placebo effects are, by construction, independent from the "dose" of social interactions received by non-volunteer pupils. Any dose-response relationship between the magnitude of spillover effects and the level of classroom interactions is compatible with peer effects only.

As it turns out, the prediction that more interactions with treated peers lead to larger impacts on non-treated peers is borne out by our data: on the subsample of non-volunteers, the impact is indeed larger, for most outcomes, when there are many volunteers in a treated class (more than one sixth of the class volunteered) than in classes with few volunteers (Table 9). Since we did not randomize treatment intensity across classes, we cannot exclude that these differences in spillovers across classes with high and low numbers of volunteers reflect (at least to some extent) the heterogeneity of spillover effects across classes which are *ex-ante* different. The exercise which is presented in Table 9 is, however, clearly suggestive of a dose-response relationship between the quantities of interactions with treated peers to which non-volunteers were exposed over the school year and the quality of their behavior at the end of the year.

To sum up, we do find support for a positive dose-response relationship – more interactions leading to larger spillovers – across both the intensive (time) margin and the extensive margin.

B. Class Level Analysis

Overall, our findings suggest that the involvement of parents modifies the functioning of their children's class both directly and indirectly, i.e., through the parents' direct influence on their own children and through the influence of their own children on other children in the class. In this section, we provide an evaluation of the effect of this combination of direct and indirect influences on the average outcomes of a class. As discussed below, this class-level analysis provides a picture of the causal effect of parental involvement on classes' outcomes which is somewhat different from the picture that would emerge from a non-experimental approach.

To start with, panel A of Table 10 shows the results of regressing the average outcomes of a class on a dummy indicating whether the class has been randomly selected for the program or not. This confirms that selecting a class into the program increases significantly the average involvement of parents[14] (as measured by our global score) and improves significantly the average behavior of pupils, with a decrease of about 0.6 half day in average absenteeism during the last term, a decrease of about 3 percentage points in the proportion of sanctions, and an increase of about 5 percentage points in the proportion of honors. Also, these reduced-form regressions confirm that the selection into the program generates improvements in average performance in Math and French that are positive, although not statistically significant at standard levels. We find a marginally significant effect on the easiest part of the French test only. Panel B of the same table provides the corresponding Instrumental Variable (IV) evaluation, i.e., the IV regression estimate of the

[14] For the sake of clarity, these regressions use a normalized version of the class-level parental involvement score. After computing the class-level average of individual scores based on available responses, the distance between the first (P25) and third (P75) quartile of the distribution of class scores is set to 1. Given this normalization, the estimated effect on parental involvement (.296) means that selecting a class into the program has an effect on average parental involvement which is equivalent to 29.6% of the difference observed between the 25% of classes where involvement is maximum (high-involvement classes) and the 25% classes where it is minimum (low-involvement classes).

effect of a unit increase in parents' average involvement on the average outcomes of a class, where the dummy indicating class selection is used as instrumental variable. They confirm that parental involvement has a large and significant effect on all observed aspects of pupils' behavior as well as on their ability to pass the easiest part of the French test. Given the normalization used, results suggest that the difference in parental involvement between the 25% of classes where involvement is highest and the 25% of classes where it is lowest generates, by itself, a reduction of about 11.2 percentage points in the proportion of pupils getting sanctions for their misbehavior, an increase of 19.1 percentage points in the proportion of pupils earning honors, and an incremental gain of about 28.6% of a standard deviation in the average score obtained at the easiest part of the French test. These effects are large and suggest that differences in parental involvement represent, as such, one very important explanation for the high level of heterogeneity in pupils' behavior across classes.

For the sake of comparison, the last panel of Table 10 shows the results of the corresponding OLS regression. This OLS regression of average outcomes on average parental involvement provides a naive estimate of the effect of parental involvement on class-level outcomes, which only controls for selection on observables. OLS estimates are most likely biased, as their identification does not allow for the possibility that parents adjust their involvement level after observing their child's classroom ambiance and in reaction to their child's behavior. As it turns out, the estimated OLS effects of parental involvement on pupils' behavior (and on French test scores) are significant and positive, but systematically less large than the IV estimates (even though the differences between the two set of estimates are marginally significant only). In other words, a non-experimental approach would have provided a downward biased picture of the potential of parental involvement for improving the functioning of classes. One possible interpretation is that parental involvement is boosted, ceteris paribus, whenever children have problems or are assigned to peers with problems. In

such a case, the rough correlation between parents' involvement and the classes' outcomes does not reflect the true causal effect of involvement on outcomes, but a mix between this causal effect and a negative selection effect. Another possible interpretation for smaller OLS effect is that the impact of parental involvement is heterogeneous and more significant for the group of volunteer families than for the average ones.

VI. Conclusion

Governments and schools are increasingly enthusiastic about improving child outcomes through parenting programs. Parents are sometimes seen as a reserve of underutilized inputs, waiting to be called upon to contribute, at low cost, to the process leading to better school outcomes. This drive has delivered an abundance of policy initiatives, but remarkably little rigorous evidence on whether, and how, interventions fostering parental involvement in education are successful.

This paper provides experimental evidence that middle-school classes in poor neighborhoods are less exposed to truancy and to episodes of misbehavior when parents receive invitations and support to become effectively involved in their children's school education. Teaching and learning activities take place in a more cooperative environment in those classes and pupils' command of the most basic reading skills improves.

Generally speaking, our results show that low levels of parental involvement are not a fatality in poor neighborhoods. Schools have the critical ability to trigger higher levels of involvement among some parents, and this can be enough to improve the outcomes for all children. On a population of pre-adolescents, our findings show that these interventions first and mostly deliver improvements in their behavior at school, which might be instrumental for gains in achievement. Furthermore, the results of this study not only stress the influence that parental behavior has on pupils, but also the role of peer-pressure in shaping pupils' behavior.

Taken together, these two influences redefine high level of parental involvement as a club good at the class level, rather than a private investment: all pupils in a same class benefit from higher monitoring and involvement efforts by some parents.

A central debate in education is whether remedial programs should be targeted at the individual level. Our results on spillovers demonstrate some benefits of universal provision of parenting programs over the alternative of targeting at-risk families only. Providing support to entire communities has the advantage of minimizing the stigma associated with individual targeting. In the context of parenting programs, this does not come necessarily at the cost of smaller benefits for individual pupils, given the large spillovers at play.

Despite universal provision, the evaluated program had low take-up rates among potential beneficiaries. The net benefits from this intervention could probably be increased if more parents took up the program. Even the most rational among parents do not internalize the large positive externalities of their efforts on classmates. The provision of small and targeted incentives to compensate some parents for their effort to attend school-based meetings could be a way to improve take-up and cost-effectiveness.

References

Aizer, Anna (2004), "Home alone: supervision after school and child behavior", *Journal of Public Economics*, vol. 88, pp. 1835-1848.

Angrist, Joshua and Kevin Lang (2004), "Does School Integration Generate Peer Effects? Evidence from Boston's Metco Program", *American Economic Review*, vol. 94(5), pp. 1613-1634.

Avvisati, Francesco, Bruno Besbas and Nina Guyon (2010), "Parental Involvement in Schools: A Literature Review", *Revue d'Economie Politique*, vol. 120, no. 5, pp. 761-778, September-October 2010.

Avvisati, Francesco, Marc Gurgand, Nina Guyon and Eric Maurin (2010), "Getting Parents Involved: a Field Experiment in Deprived Schools", *CEPR Discussion Paper* No. 8020

Banerjee, Abhijit, Rukmini Banerji, Esther Duflo, Rachel Glennerster and Stuti Khemani (2008), "Pitfalls of Participatory Programs: Evidence from a Randomized Evaluation in Education in India", *NBER Working Paper* No. 14311.

Björkman, Martina and Jakob Svensson (2009), "Power to the People: Evidence from a Randomized Field Experiment on Community-Based Monitoring in Uganda", *Quarterly Journal of Economics*, vol. 124(2), pp. 735-769.

Cunha, Flavio and James J. Heckman (2008), "Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation", *Journal of Human Resources*, vol. XLIII(4), pp. 739-782.

Desforges, Charles and Alberto Abouchaar (2003), "The Impact of Parental Involvement, Parental Support and Family Education on Pupil Achievements and Adjustment: a Literature Review", *Department for Education and Skills*, Research Report No. 433.

Hoover-Dempsey, Kathleen V. and Howard M. Sandler (1995), "Parental Involvement in Children's Education: Why Does it Make a Difference?", *Teachers College Record*, vol. 97, pp. 310-331.

Hoover-Dempsey, Kathleen V. and Howard M. Sandler (1997), "Why Do Parents Become Involved in Their Children's Education?", *Review of Educational Research*, vol. 67, pp. 3-42.

Hoxby, Caroline (2000), "Peer Effects in the Classroom: Learning from Gender and Race Variation", *NBER Working Paper* No. 7867.

Moffitt, Robert A. (2001), "Policy Interventions, Low-Level Equilibria, and Social Interactions", pp. 45-82 of Durlauf, Steven N. and H. Peyton Young, (Eds.), Social Dynamics. *The MIT Press*, for Brookings Institution.

OECD (2010), "PISA 2009 Results: What Students Know and Can Do - Student Performance In Reading, Mathematics and Science (Vol. I)", *Organisation for Economic Co-Operation and Development*.

Scott, Stephen, Thomas O'Connor and Annabel Futh (2006), "What makes parenting programmes work in disadvantaged areas?", *Joseph Rowntree Foundation*.

Todd, Petra E. and Kenneth E. Wolpin (2007), "The Production of Cognitive Achievement in Children: Home, School, and Racial Test Score Gaps", *Journal of Human Capital*, vol.1, no.1, pp. 91-136.

Welsch, David M. and David M. Zimmer (2008), "After-School Supervision and Children's Cognitive Achievement", *The B.E. Journal of Economic Analysis & Policy*, vol. 8, art. 49.

Table 1: Difference in Pre-Treatment Characteristics across Treatment Arms

	Mean	Std	T - C	(se)	N.Obs.
		All			
Parents					
Employment status	0.85	0.36	-0.003	(0.011)	4660
Intact family	0.73	0.44	-0.001	(0.013)	4729
White collar	0.19	0.39	0.001	(0.011)	4529
Children					
Girl	0.48	0.50	0.006	(0.011)	4729
6th grade repetition	0.06	0.23	0.001	(0.005)	4729
Age (sept 2008)	11.45	0.57	0.002	(0.019)	4729
Tests					
French test (sept 2008)	-0.01	1.00	-0.047	(0.045)	4163
Maths test (sept 2008)	-0.01	1.00	0.003	(0.037)	4165
		Volunteers Only			
Parents					
Employment status	0.84	0.36	-0.015	(0.021)	1056
Intact family	0.75	0.43	-0.020	(0.023)	1056
White collar	0.21	0.41	-0.014	(0.022)	1037
Children					
Girl	0.46	0.50	-0.016	(0.029)	1056
6th grade repetition	0.05	0.21	0.007	(0.011)	1056
Age (sept 2008)	11.41	0.57	0.019	(0.034)	1056
Tests					
French test (sept 2008)	-0.07	1.00	-0.019	(0.073)	985
Maths test (sept 2008)	-0.04	1.01	-0.053	(0.069)	994
		Non Volunteers Only			
Parents					
Employment status	0.85	0.35	0.001	(0.012)	3604
Intact family	0.72	0.45	0.005	(0.014)	3673
White collar	0.18	0.39	0.006	(0.012)	3492
Children					
Girl	0.48	0.50	0.013	(0.012)	3673
6th grade repetition	0.06	0.24	-0.001	(0.006)	3673
Age (sept 2008)	11.46	0.57	-0.002	(0.021)	3673
Tests					
French test (sept 2008)	0.01	1.00	-0.056	(0.047)	3178
Maths test (sept 2008)	0.00	1.00	0.021	(0.040)	3171

The first two columns are the mean and standard-error of each row variable in the control group. Column "T - C" displays the coefficient from the regression of the row variable on a test dummy and school fixed effects. Each line corresponds to a separate regression. When volunteers and non-volunteers are pooled, the regression also includes school*volunteer fixed effects (school dummies interacted with volunteer status dummies). Robust standard errors allowing for correlated residuals within classes are shown in parentheses.
*: Significant at the 10% level. **: significant at the 5% level.

Table 2: Take-up Rates for the 4 Populations (Volunteers and Non-Volunteers in Test and Control Classes)

	Test		Control	
	Vol	N.Vol.	Vol	N.Vol.
Initial Workshops				
At least 1 debate	57.8	1.1	4.1	0.2
At least 2 debates	35.8	0.2	0.6	0.0
All 3 debates	16.9	0.1	0.6	0.0
Additional Workshops				
Parenting	11.7	0.6	0.0	0.0
Internet	7.8	0.5	0.4	0.0
French as foreign language	3.2	0.0	0.0	0.0
Any of the above	16.7	0.9	0.4	0.0

All rates are expressed in percentage terms. **Reading:** 57.8% of test volunteers took part in at least one debate.

Table 3: Impact of the Program on Parental Attitudes and Behavior

Dependent Variable	Mean C	T - C	(se)
All			
Global Parenting Score	-0.072	0.109**	(0.035)
School-Based Involvement Score	-0.052	0.093**	(0.034)
Home-Based Involvement Score	-0.023	0.041	(0.032)
Understanding & Perceptions Score	-0.029	0.072*	(0.037)
Never been summoned to the school	0.79	0.029**	(0.015)
Volunteers			
Global Parenting Score	-0.147	0.266**	(0.071)
School-Based Involvement Score	0.173	0.323**	(0.076)
Home-Based Involvement Score	0.016	0.103*	(0.057)
Understanding & Perceptions Score	-0.182	0.184**	(0.071)
Never been summoned to the school	0.72	0.077**	(0.029)
Non Volunteers			
Global Parenting Score	-0.049	0.044	(0.040)
School-Based Involvement Score	-0.124	-0.001	(0.036)
Home-Based Involvement Score	-0.035	0.016	(0.040)
Understanding & Perceptions Score	0.020	0.027	(0.041)
Never been summoned to the school	0.81	0.009	(0.016)

Score variables are standardized summaries of answers to questions in the corresponding section of the parent questionnaire. "Never been summoned to the school" is a dummy variable. The first column is the mean of each row variable in the control group. Column "T - C" displays the coefficient from the regression of the row variable on a test dummy and school fixed effects. Each line corresponds to a separate regression. When volunteers and non-volunteers are pooled, the regression also includes school*volunteer fixed effects (school dummies interacted with volunteer status dummies). Coefficients in column "T - C" are to be interpreted as standardized effect sizes, except for the dummy indicator "Never been summoned to school", where it corresponds to the predicted change in the probability. Robust standard errors allowing for correlated residuals within classes are shown in parentheses. Sample size corresponds to 3026 for the complete sample, 2192 for non-volunteers and 834 for volunteers.
*: Significant at the 10% level. **: significant at the 5% level.

Table 4: Parental Attitudes and Behavior Scores: Difference by Socio-Economic Status

Dependent Variable	Mean oth.	White-collar	(se)
Global Parenting Score	-0.146	0.346**	(0.068)
School-Based Involvement Score	-0.088	0.215**	(0.062)
Home-Based Involvement Score	-0.090	0.253**	(0.067)
Understanding & Perceptions Score	0.002	-0.131*	(0.067)

Score variables are standardized summaries of answers to questions in the corresponding section of the parent questionnaire. The first column is the mean of each row variable in the non white-collar population. Column "white-collar" displays the coefficient from the regression of each score variable on a white-collar dummy and school fixed effects. Each line corresponds to a separate regression. Only control classes are used in these regressions. Coefficients in column "T - C" are to be interpreted as standardized effect sizes. Robust standard errors allowing for correlated residuals within classes are shown in parentheses. Sample size is 1485.
*: Significant at the 10% level. **: significant at the 5% level.

Table 5: Impact of the Program on Pupils' Behavior (Term 3)

Dependent Variable	Mean C	T - C	(se)	Std	N.Obs.
		All			
Absenteeism	4.319	-0.642**	(0.309)	7.736	3402
Behav. score	-0.012	0.106**	(0.037)	1.024	4468
Discipl. sanctions	*0.109*	*-0.024***	*(0.011)*	*0.296*	*4199*
Good conduct	*0.326*	*0.048***	*(0.024)*	*0.481*	*2969*
Honors	*0.342*	*0.043***	*(0.016)*	*0.481*	*4232*
		Volunteers			
Absenteeism	4.217	-0.771	(0.549)	7.736	786
Behav. score	-0.014	0.124*	(0.065)	1.024	1045
Discipl. sanctions	*0.106*	*-0.036**	*(0.020)*	*0.296*	*975*
Good conduct	*0.289*	*0.044*	*(0.038)*	*0.481*	*676*
Honors	*0.345*	*0.028*	*(0.028)*	*0.481*	*1006*
		Non Volunteers			
Absenteeism	4.345	-0.602*	(0.337)	7.736	2616
Behav. score	-0.012	0.100**	(0.042)	1.024	3423
Discipl. sanctions	*0.110*	*-0.021**	*(0.012)*	*0.296*	*3224*
Good conduct	*0.336*	*0.049**	*(0.026)*	*0.481*	*2293*
Honors	*0.341*	*0.048***	*(0.019)*	*0.481*	*3226*

The first column is the mean of each row variable in the control group. Column "T - C" displays the coefficient from the regression of the row variable on a test dummy and school fixed effects. Each line corresponds to a separate regression. When volunteers and non-volunteers are pooled, the regression also includes school*volunteer fixed effects (school dummies interacted with volunteer status dummies). Robust standard errors allowing for correlated residuals within classes are shown in parentheses.
*: Significant at the 10% level. **: significant at the 5% level.

Table 6: Impact of the Program on Pupils' Achievement Gains

	Impact (std pts)		*Base points*		
Dependent Variable	T - C	(se)	Mean C	Std	N.Obs.
Teacher Marks (Term 3):					
		All			
French	0.065*	(0.036)	10.8	3.8	4271
Maths	0.005	(0.038)	10.9	4.3	4271
		Volunteers			
French	0.151**	(0.048)	10.7	3.8	1009
Maths	0.024	(0.054)	10.9	4.3	1009
		Non Volunteers			
French	0.040	(0.038)	10.9	3.8	3262
Maths	0.003	(0.038)	10.9	4.3	3262
Tests (Term 3):					
		All			
French	0.039	(0.042)	62.6	17.9	3734
Observation	0.109**	(0.045)	78.0	18.3	3734
Maths	-0.013	(0.038)	54.0	19.2	3707
		Volunteers			
French	-0.032	(0.055)	62.0	17.9	881
Observation	0.211**	(0.063)	77.5	18.3	881
Maths	-0.012	(0.055)	53.0	19.2	870
		Non Volunteers			
French	0.060	(0.046)	62.8	17.9	2853
Observation	0.076	(0.050)	78.1	18.3	2853
Maths	-0.014	(0.039)	54.2	19.2	2837

The row variable for each achievement measure is computed as the difference between end-of-year standardized scores and start-of-year standardized scores. Column "T - C" displays the coefficient from the regression of the row variable on a test dummy and school fixed effects. All observations for which end-of-year achievement is available are included in the analysis; dummies for missing Term 1 measures are added to the regression. Each line corresponds to a separate regression. When volunteers and non-volunteers are pooled, the regression also includes school*volunteer fixed effects (school dummies interacted with volunteer status dummies). Robust standard errors allowing for correlated residuals within classes are shown in parentheses. Columns 3 and 4 report descriptive statistics for the control groups' term 3 measures; marks are given on a 0-20 scale, while test scores are in percentage terms.
*: Significant at the 10% level. **: significant at the 5% level.

Table 7: Impact of the Program on Selected Subgroups of Volunteers and Non-Volunteers

Panel A: Impact on Parental Attitudes and Behavior by SES											
		Global Score		School-inv. Score		Home-inv. Score		Und. & Perc. Score		Never Summoned...	
T - C (V):	w-collar	.306**	(.118)	.536**	(.129)	.113	(.096)	.312**	(.121)	.016	(.045)
T - C (V):	others	.253**	(.069)	.332**	(.081)	.100*	(.058)	.129	(.078)	.072**	(.029)
T - C (NV):	w-collar	-.019	(.088)	-.011	(.086)	.011	(.096)	.199*	(.101)	-.016	(.034)
T - C (NV):	others	.042	(.040)	.004	(.045)	.040	(.045)	-.009	(.042)	.003	(.017)
N		3026		3024		3025		3018		3013	
Panel B: Impact on Pupils' Behavior by Gender											
		Absenteeism		Behav. Score		Sanctions		Good C'duct		Honors	
T - C (V):	girls	-.958	(.637)	.171**	(.072)	-.042**	(.021)	.109**	(.050)	.025	(.036)
T - C (V):	boys	-1.247*	(.721)	.181**	(.072)	-.046*	(.026)	.057	(.037)	.060*	(.032)
T - C (NV):	girls	-.126	(.438)	.115**	(.049)	-.034**	(.013)	.069**	(.034)	.057**	(.028)
T - C (NV):	boys	-1.097**	(.446)	.147**	(.054)	-.021	(.019)	.055*	(.028)	.070**	(.021)
N		3401		4467		4198		2971		4234	
Panel C: Impact on Pupils' Achievement Gains by Initial Achievement Group											
		French Mark		Maths Mark		French Test		French Test: Observation		Maths Test	
T - C (V):	top	.166**	(.072)	-.000	(.073)	-.029	(.075)	.122	(.085)	.033	(.058)
T - C (V):	med	.109	(.068)	-.039	(.078)	-.072	(.068)	.093	(.095)	-.113	(.069)
T - C (V):	low	.204**	(.073)	.009	(.074)	.001	(.077)	.396**	(.104)	-.005	(.066)
T - C (NV):	top	.012	(.055)	.014	(.046)	.029	(.048)	.044	(.059)	.009	(.044)
T - C (NV):	med	.096*	(.049)	-.014	(.050)	-.079	(.055)	-.065	(.071)	-.045	(.048)
T - C (NV):	low	.014	(.063)	-.005	(.054)	-.012	(.059)	.070	(.081)	-.015	(.055)
N		4028		4070		3493		3493		3475	

The table reports estimates from an augmented version of equation (3), where the treatment variable is interacted with a volunteer dummy and all the variables of interest are fully interacted with subgroup dummies. All regressions also include school*volunteer fixed effects (school dummies interacted with volunteer status dummies) as well as controls for gender, grade repetition, scholarship, intact family, employment status (3 levels), white-collar occupation; the exact age in days, test scores at baseline tests in French and Maths, plus dummies for missing observations on baseline tests; the average of these individual characteristics over classmates; dummies for low, medium and high proportion of volunteers, fully interacted with own volunteer status. In Panel C pupils are assigned to initial achievement groups based on their point average at baseline tests in French and Maths; consequently, only observations with non-missing baseline tests are used, and dummies for initial achievement group are also added to estimation. Robust standard errors allowing for correlated residuals within classes are shown in parentheses. Each line corresponds to a separate regression. Robust standard errors allowing for correlated residuals within classes are shown in parentheses.
*: Significant at the 10% level. **: significant at the 5% level.

Table 8: Difference in Pupils' Behavior in Term 1

Dependent Variable	Mean C	T - C	(se)	Std	N.Obs.
		All			
Absenteeism	1.125	0.030	(0.113)	2.976	3823
Behav. score	-0.115	0.055	(0.040)	1.159	4603
Discipl. sanctions	*0.086*	*-0.016*	*(0.011)*	*0.265*	*3868*
Good conduct	*0.385*	*0.020*	*(0.028)*	*0.491*	*2902*
Honors	*0.453*	*0.009*	*(0.017)*	*0.499*	*4301*

The first column is the mean of each row variable in the control group. Column "T - C" displays the coefficient from the regression of the row variable on a test dummy and school*volunteer fixed effects (school dummies interacted with volunteer status dummies). Each line corresponds to a separate regression. Robust standard errors allowing for correlated residuals within classes are shown in parentheses.
*: Significant at the 10% level. **: significant at the 5% level.

Table 9: Spillover Effects of the Program as Result of Repeated and Sustained Interaction with Treated Peers

	Pupil Behavior (Term 3)				
	Absent.	Beh. Sc.	Sanct.	Gd C'duct	Honors
T - C (NV): few vol.	-.310	.045	-.005	.041	-.020
	(.682)	(.064)	(.016)	(.048)	(.035)
T - C (NV): many vol.	-1.074**	.168**	-.034**	.063*	.093**
	(.444)	(.055)	(.017)	(.032)	(.024)
N	2615	3422	3223	2295	3228
p-value	.363	.169	.230	.718	.010

The table reports estimates from an augmented version of equation (2) with non-volunteers only, where the variable of interest is fully interacted with dummies for low (<.16%) and medium to high proportion of volunteers in class (⩾16%). One third of classes are in the first category: the threshold is selected to correspond to the first tercile of the distribution. All regressions include school fixed effects and dummies for low, medium and high proportion of volunteers. Each column corresponds to a separate regression. The p-value associated with a test of equality among the coefficients of interest is reported. Robust standard errors allowing for correlated residuals within classes are shown in parentheses.
*: Significant at the 10% level. **: significant at the 5% level.

Table 10: Class-Level Analysis

Panel A: First Stage and Reduced Form Regressions

	1st Stage:	Reduced Form:							
	Parents	Pupils' Behavior				Marks		Tests		
	(Gl. Score)	Absent.	Sanct.	Gd C'duct	Honors	French	Maths	French	Fr.: Obs.	Maths
T - C	.296**	-.577*	-.029**	.057**	.050**	.049	.029	.029	.082	.009
	(.108)	(.308)	(.012)	(.027)	(.017)	(.042)	(.040)	(.042)	(.052)	(.042)
School fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	188	156	187	135	192	187	189	182	182	185

Panel B: Instrumental Variables Regressions

	Pupils' Behavior				Marks		Tests		
	Absent.	Sanct.	Gd C'duct	Honors	French	Maths	French	Fr.: Obs.	Maths
Parents (gl. score)	-1.950*	-.112**	.198*	.191**	.172	.063	.100	.286*	.026
	(1.182)	(.050)	(.113)	(.066)	(.121)	(.126)	(.123)	(.161)	(.129)
School fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	152	175	131	180	176	178	181	181	181

Panel C: OLS Regressions

	Pupils' Behavior				Marks		Tests		
	Absent.	Sanct.	Gd C'duct	Honors	French	Maths	French	Fr.: Obs.	Maths
Parents (gl. score)	-.326	-.032**	.034*	.054**	.032	-.057*	.064**	.109**	.111**
	(.247)	(.011)	(.020)	(.013)	(.028)	(.032)	(.029)	(.037)	(.028)
School fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	152	175	131	180	176	178	181	181	181

All regressions are estimated on class-level averages. Controls include school fixed effects as well as the class composition in terms of gender, grade repetition, scholarship, intact family, employment status, white-collar occupation; the average age in days; average test scores at baseline tests in French and Maths, plus the proportion of missing observations on baseline tests. Each column corresponds to a separate regression.
*: Significant at the 10% level. **: significant at the 5% level.

Figure 1: Flow diagram for the Field Experiment

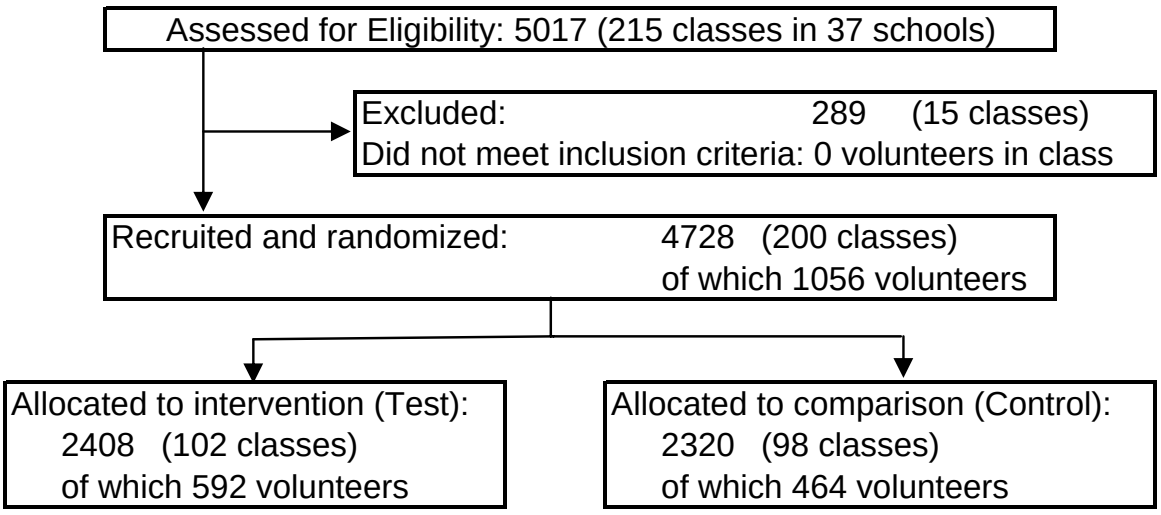

Table Appendix

Table A1: Parent Questionnaire: Response Rate

	Response rate				*Source of attrition*		
Population	Mean C	T - C	(se)	Resp.	Sch.	Cl.	Ind.
Non Volunteers	0.63	-0.039	(0.026)	2192	34	0.06	0.32
Volunteers	0.67	-0.015	(0.035)	698	34	0.10	0.24
Volunteers (incl. call-back)	0.80	-0.011	(0.026)	834	34	0.02	0.16

The first column is the mean of the response rate in the control group. Column "T - C" displays the coefficient from the regression of a response dummy on a test dummy and school fixed effects. Each line corresponds to a separate regression. Robust standard errors allowing for correlated residuals within classes are shown in parentheses. "Source of attrition" indicates: the number of schools contributing observations (out of 37), the share of classes with no information in providing schools and the share of individuals with missing information in providing classes.
*: Significant at the 10% level. **: significant at the 5% level.

Table A2: Pupil Outcomes: Response Rate

		Response rate			*Source of attrition*		
Dependent Variable	Mean C	T - C	(se)	Resp.	Sch.	Cl.	Ind.
Pupil Behavior (Term 3):							
		All					
Absenteeism	0.74	0.000	(0.006)	3402	28	0.05	0.07
Behav. score	0.94	-0.001	(0.006)	4468	37	0.07	0.06
Discipl. sanctions	*0.88*	*-0.004*	*(0.006)*	*4199*	*35*	*0.07*	*0.05*
Good conduct	*0.61*	*0.009*	*(0.021)*	*2969*	*28*	*0.16*	*0.07*
Honors	*0.90*	*-0.021*	*(0.013)*	*4232*	*36*	*0.08*	*0.07*
Teacher Marks & Tests (Term 3):							
		All					
Teacher Marks	0.89	0.021	(0.017)	4269	37	0.02	0.08
French Tests (June 2009)	0.81	-0.024	(0.023)	3736	35	0.05	0.14
Maths Tests (June 2009)	0.80	-0.015	(0.016)	3709	35	0.04	0.16
		Volunteers					
Teacher Marks	0.95	0.014	(0.020)	1009	37	0.02	0.03
French Tests (June 2009)	0.86	-0.040	(0.029)	881	35	0.06	0.08
Maths Tests (June 2009)	0.84	-0.013	(0.025)	870	35	0.04	0.11
		Non Volunteers					
Teacher Marks	0.87	0.023	(0.018)	3260	37	0.02	0.10
French Tests (June 2009)	0.79	-0.019	(0.023)	2855	35	0.05	0.15
Maths Tests (June 2009)	0.79	-0.015	(0.016)	2839	35	0.04	0.17

The first column is the mean of each row variable in the control group. Column "T - C" displays the coefficient from the regression of a response dummy on a test dummy and school fixed effects. Each line corresponds to a separate regression. When volunteers and non-volunteers are pooled, the regression also includes school*volunteer fixed effects (school dummies interacted with volunteer status dummies). Robust standard errors allowing for correlated residuals within classes are shown in parentheses. "Source of attrition" indicates: the number of schools contributing observations (out of 37), the share of classes with no information in providing schools and the share of individuals with missing information in providing classes.
*: Significant at the 10% level. **: significant at the 5% level.

Table A3: Impact of the Program on Parental Attitudes and Behavior: Volunteers Only (Raw Indicators)

Question	Mean C	T - C	(se)
Volunteers			
Global Parenting Score	-0.141	0.266**	(0.071)
School-Based Involvement Score	0.172	0.320**	(0.076)
Several individual appointments with teachers	0.24	0.056*	(0.033)
Has attended parents/teachers meetings	0.80	0.083**	(0.026)
Has participated in parents' organizations	0.24	0.111**	(0.032)
Home-Based Involvement Score	0.015	0.103*	(0.057)
Precise knowledge of child's grades	0.44	0.011	(0.035)
Sometimes helps with homeworks	0.88	0.004	(0.023)
Child does not watch TV daily after 9pm	0.80	0.052**	(0.025)
Child spends less than 1 h/d on other screens	0.88	0.027	(0.019)
Understanding & Perceptions Score	-0.182	0.184**	(0.071)
Knowledge of optional courses offered	0.76	0.093**	(0.028)
Has never been anxious about violence	0.26	0.014	(0.028)
Clear ideas about high-school plans	0.27	0.048	(0.031)
Satisfied with school	0.81	0.048**	(0.021)
Never been summoned to the school	0.72	0.077**	(0.029)

Score variables are standardized summaries of answers to questions in the corresponding section of the parent questionnaire. Dependent variables in italics are dummy variables, constructed from answers to one question. The first column is the mean of each row variable in the control group. Column "T - C" displays the coefficient from the regression of the dependent variable on a test dummy and school fixed effects. Each line corresponds to a separate regression. Coefficients in column "T - C" are to be interpreted as standardized effect sizes, except for the dummy indicators, where it corresponds to the predicted change in the probability. Robust standard errors allowing for correlated residuals within classes are shown in parentheses. Sample size corresponds to 834.
*: Significant at the 10% level. **: significant at the 5% level.