DISCUSSION PAPER SERIES

No. 8020

GETTING PARENTS INVOLVED: A FIELD EXPERIMENT IN DEPRIVED SCHOOLS

Francesco Avvisati, Marc Gurgand, Nina Guyon and Eric Maurin

LABOUR ECONOMICS

Centre for Economic Policy Research

www.cepr.org

Available online at: **www.cepr.org/pubs/dps/DP8020.asp**

ISSN 0265-8003

GETTING PARENTS INVOLVED: A FIELD EXPERIMENT IN DEPRIVED SCHOOLS

Francesco Avvisati, Paris School of Economics
Marc Gurgand, Paris School of Economics
Nina Guyon, Paris School of Economics
Eric Maurin, Paris School of Economics and CEPR

Discussion Paper No. 8020
September 2010

Centre for Economic Policy Research
53–56 Gt Sutton St, London EC1V 0DG, UK
Tel: (44 20) 7183 8801, Fax: (44 20) 7183 8820
Email: cepr@cepr.org, Website: www.cepr.org

CEPR Discussion Paper No. 8020

September 2010

ABSTRACT

Getting Parents Involved: A Field Experiment in Deprived Schools*

This paper presents a randomized field experiment conducted in a set of French middle schools located in a deprived educational district near Paris. Parents in test groups were invited to participate in a simple program of training sessions on how to get better involved in their children's education. 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.

JEL Classification: I21, J13 and J18
Keywords: child support, classroom peer-effects, cluster randomized trial, parental involvement

Francesco Avvisati
Paris School of Economics
48 Boulevard Jourdan
75014 Paris
FRANCE

Email: francesco.avvisati@ens.fr

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=172541

Marc Gurgand
Paris School of Economics
48 Boulevard Jourdan
75014 Paris
FRANCE

Email: gurgand@pse.ens.fr

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=158565

Nina Guyon
Paris School of Economics
48 Boulevard Jourdan
75014 Paris
FRANCE

Email: nina.guyon@gmail.com

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=161529

Eric Maurin
Paris School of Economics
Bureau 105, Bat B
48 Boulevard Jourdan
75014 Paris
FRANCE

Email: eric.maurin@ens.fr

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=1131502

* This research was supported by a grant from the French High Commissioner for Youth. We are very grateful for the support of the school 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) 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).

Submitted 10 September 2010

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. These issues are very high on political agendas[3].

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, 2001) and part of the national education policy in the UK ("Every Child Matters" Green Paper, 2003).

Yet, there is still very little evidence on whether such policies make a difference. In fact, 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 observed correlation between parental involvement and pupils' outcomes represents any causal effect at all.

In this paper, we use a randomized field experiment in middle schools of the Paris area to shed light on these issues. In a relatively deprived educational district near Paris, 37 schools offered a program of debates and training on how to help children succeed at school to families from randomly chosen classes. 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. Children need to adapt to a completely new environment; parents need to develop a new partnership with schools. The program offered information on the functioning of schools and advice on how to help

[3] In 2003, in the combined OECD area, 14% of 15-year-old students are capable of completing only the simplest reading tasks developed for PISA, such as locating a single piece of information, identifying the main theme of a text or making a simple connection with everyday knowledge. 8% of students are even below this level of ability (OECD, 2004).

children with homework. These debates were eventually followed by training sessions on similar issues.

We show that this simple program increased effectively the level and quality of school-related parental care. Also, we find that this improved involvement of parents translated into a significant reduction of truancy and misbehavior in test classes. The program ultimately translated into less literacy problems for children whose parents were invited to the program. Most interestingly, 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 those whose parents did not participate.

The experiment started in September 2008, at the beginning of the 2008-2009 academic year. 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. During this 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 enrolled and non-enrolled families prior to the random decision of running the program. By comparing enrolled families in test classes and enrolled families in control classes (where enrolment did not result in participation), 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.

We find that 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), where white-collar families represent the top 20% of the population in term of socio-economic status. 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 pupils' behaviors.

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 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 education. The few existing studies in this field within the economics discipline all adopt a structural approach, using survey data. 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 actually 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.

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

[4] Desforges and Abouchaar (2003,p.5) noted 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". Pre-school interventions have a longer tradition of rigorous evaluation: the Perry Preschool Project and Head Start Program include parental involvement modules, whose effect is however difficult to sort out from the effect of other modules.

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.

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'

[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).

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 (9th 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.

---

[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.

School heads mainly used a standardized leaflet to inform families of the program, but they also spoke of the program at the usual meetings that take place in the very first weeks of the school year. Schools were particularly 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. A particular effort was made to target the information campaign at families which are usually reluctant to get involved: the wording and design of the leaflet were visual and accessible, and extrinsic motivations mentioning children's success were put forward.

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. It was also stated that a random draw would take place at the end of October to select "one out of two" classes.

By mid-October each school listed all families who signed up, and closed the registration phase. This list defines the population 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 arm (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. The organizer and facilitator is the school head, usually assisted by a second member of the educational community (a teacher or supervisor). To introduce each session, the school head can draw on precise guidelines, designed by the districts' educational experts. Facilitators are 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. To make these meetings accessible to all parents, the availability of translators for parents who were not fluent in French was also announced.

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 first results, discussed at the *conseil de classe*. Parents are encouraged to ask questions, explain their problems and share their own experience.

The district-level guidelines insist that the facilitator 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

[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).

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.

D. Objectives

The program and its materials were developed by educational experts at the district level in accordance with state-of-the-art theories about parental involvement. According to the psychological model proposed by Hoover-Dempsey and Sandler (1995), 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 – reviewed by Hoover-Dempsey and Sandler (1995,1997) in the context of parental involvement – parents model their own child's attitudes by devoting interest and time to activities related to schooling; moreover, 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.

The primary objective of the experiment was to test whether a relatively un-intensive program inspired by these theories could successfully lead to (a) increased levels of parental involvement with education, both at school – the most direct outcome – and at home, and (b) whether increased involvement would result in better non-cognitive and cognitive achievement for children. By exploiting its' cluster design, a secondary objective was (c) to measure the magnitude of classroom spillovers in attitudes, behaviors, and achievement.

III. Outcome Measures, Randomization, and Methods for Statistical Analysis

A. Randomization

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].

As a first step, classes were ranked based on the number of volunteers; classes without volunteers were not eligible to random assignment (15 out of 215 classes). The random assignment protocol distinguished two cases. (a) If the number of eligible classes m was uneven, and the class with the smallest number of volunteers had more than three volunteers, randomization was unrestricted and resulted in the selection of $(m+1)/2$ classes in the test arm. In contrast, if the class with the smallest number of volunteers had three or less volunteers, this class was grouped with the class just above it in the ranking by volunteers, and the two formed a single randomization unit – thereby resulting in an even number of randomization units. (b) If the number of classes with one or more volunteers was even to begin with, or if the mentioned procedure was applied to produce an even number of randomization units, a restricted randomization procedure (blocking) was used: within each pair formed by the ranking by number of volunteers (ranks 1 & 2, ranks 3 & 4, etc), a random sequence selected one out of two randomization units (i.e. classes, or, sometimes, groups of two classes). Generally speaking, this procedure aimed at ensuring a certain balance in the number of treatment and control volunteers within each school (on top of producing balanced treatment and control groups by virtue of random assignment).

The empirical analysis will use weights, which are defined at the class level as the inverse of the ex-ante probability of being assigned to the treatment arm (test or control) to which each class belongs, to ensure that each school has the same weight in the test and control group. In case (a), weights equal $m/(m+1)$ for observations in treatment classes and $m/(m-1)$ for observations in control classes, where m is the total number of eligible classes. In case (b), weights are uniformly equal to 1.

[8]The publicity of the random allocation was intended to ensure trust in the impartiality and transparency of researchers, as was the fact that the “random sequence” was actually based on externally verifiable numbers : the landline number of the school and the school head’s month and day of birth in case. This latter aspect also ensured our ability to replicate the random assignment outside the head’s office.

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.

B. Outcome measures: Parents

A multiplicity of outcome measures was defined, making use of different data sources, to describe the program's impact on parent and child-level outcomes. All outcome measures were collected at the individual level.

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 15 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 appointements 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.

Because the questionnaire was self-administered, we were worried about non-response. On the subset of volunteer families, we made a special effort to minimize this issue: 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.

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' 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. The top panel of Table 1 shows how the discrete outcomes relate to each other, on the subsample of schools for which all measures are available. Reassuringly, whenever a pupil has a sanction, then necessarily the “good conduct” and “honors” dummies are equal to 0. In contrast, for children who are not sanctioned, the “good conduct” (about 25% of pupils) and “honors” dummies (another 25%) seem to be reasonably independent of each other. 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.

The bottom panel in Table 1 exhibits the relation between these three indicators and absenteeism. Interestingly enough, average absenteeism decreases sharply from nine half-days to one half-day 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 Maths and French) which are identical across schools and classes, and were externally graded.

D. Statistical methods

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 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 contrast, 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 externality.

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:* $Y_{ics} = \alpha^{V} T_c + u_s + \nu_{ics}$

(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 ν_{ics}, η_{ics} and ε_{ics} represents unobserved individual random effects. The parameters of interest are $\alpha^{V,}$ 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. 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.

Note that it is possible to perform a joint estimation of the effect of the treatment on volunteers (α^{V}) and on non-volunteers (α^{NV}), and thus to formally test for their difference, by rewriting the estimating equation as

(4) $Y_{ics} = (n_s + \alpha^{NV} T_c)*(1-V_i) + (u_s + \alpha^{V} T_c)*V_i + \varepsilon_{ics}$

In robustness and subgroup analyses, we use available characteristics of pupils, families, and classes at baseline as additional control variables in equation (4)[9]. Adding control variables should only reduce the asymptotic standard errors.

[9]Controls include dummies for girls, 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; along with dummies for low, medium and high proportion of volunteers, fully interacted with own volunteer status.

E. 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. 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.

To start with, these data can be used to compare the populations of volunteers and non-volunteers within each class (see Table 2). Volunteers have slightly more often white-collar occupation (+2.5%) and belong more often to two-parent families (+3.8%); moreover, pupils in volunteer families have been less often held back a grade. But there are no significant differences in gender nor in pre-treatment test scores between volunteers and non-volunteers. Overall, table 2 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.

Baseline data can also be used to check that observable characteristics are balanced across treatment and control groups. Comfortingly, table 3 shows that differences between test and control groups 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.

F. 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. This makes it possible to compute 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 (table 4). Comfortingly take-up is large and significant for volunteers in test class only (Vol-test), 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 (see table 5). Additional "parenting sessions" were finally organized in only 17 schools out of 37; 15 schools offered sessions for parents on the use of 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, makes up about 15% of eligible parents (test volunteers): 80 parents in 17 schools participated in about 3 additional debates on parenting issues. 57 in 15 schools participated in additional sessions about the internet (4 sessions on average) and 19 in 8 schools participated in sessions specifically designed for non-French speaking (5 sessions on average).

It must be understood, therefore, that intensity of treatment is heterogenous, with the bulk of it being the initial meetings, additional sessions having a marginal turnout. For most of the following analysis, we estimate the impact of a policy consisting in offering a menu of sessions: impact is affected by whatever the intensity and structure of participation happens to be. This is the *intention to treat* approach, which has, in the present context, direct policy implications.

G. 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 table 6 and 7.

The parent questionnaire was returned back in due time by approximately two thirds of volunteer and non-volunteer families (table 6). This proportion is very similar across treatment arms. By conducting phone interviews, we have been able to increase 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 this 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 7). 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 Maths 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 8 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 8 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 difference across these same dimensions between white-collar families and non white-collar families in control classes[10]. Table 9 shows that the differences in school-involvement levels 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.

Table 10 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 . Less volunteers from test classes allow their children to watch television after 9pm 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 11): children in test classes skip less classes (absenteeism is lower by 0.7 half-days), are less likely to be punished for disciplinary reasons (10.9% against 13.4% in control classes), are more likely to get honors (38.6% against 34.2%) and are more likely to earn the top marks for their conduct (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. Importantly, effects are present at all sections of the distribution of behavior: very bad behavior is less frequent and very good one

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

more frequent. Truancy, that was shown above to form a continuous, independent and objective measure is also strongly affected.

We estimated the difference on data from the first term of the school year: resulting estimates confirm that the observed advantage was not already present at the beginning of the year (Table 12). This is not a pure placebo test, as the debates already began by the end of the first term; still, it is reasonable to admit that the observed behavior during the first term could not be influenced by a change in parental attitudes due to the debates. As table 12 shows, by the beginning of the year the advantage, if any, of test classes over control classes with respect to the quality of behavior is small and undistinguishable from random noise. The clear advantage observed by the end of the year is therefore the result of a cumulative process; intuitively, as the parents and school develop a partnership for education, the information and recommendations given to parents at the beginning of the school year eventually curb children's attitudes towards school in a positive way. From the parents' perspective, the program is limited in time; for children, in contrast, the program corresponds to a permanent change in the attitudes of their caring adults, with whom they interact on a daily basis. Change in parental attitudes can therefore induce significant consequences on children's attitudes in the medium and long term. To fix ideas, the observed difference in absenteeism or in the likelihood of having disciplinary sanctions is, again, of the same magnitude as the difference between children of white-collar families and other children. For the positive indicators, 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.

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 11 compare test volunteers with control volunteers, and test non-volunteers with control non-volunteers[11]. Most interestingly, 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

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

were effectively invited, but also on all other children belonging to the same classes. The existence of such large spillovers is remarkable and provides a major argument in favor of developing these policies: indeed, it would not be reasonable to expect that more than a fraction of all parents will ever be, or feel, able to attend evening meetings, debates, or training at their children's school. Despite directly involving only a fraction of all parents, this kind of policy has, nevertheless, the potential to extend its benefits to all children; accordingly, the objective of maximizing its effects seems not to be in contradiction with the objective of reducing classroom inequalities. Consistently, these spillovers are observed on the behavior of children which interact daily and influence each other at the class level, but not on the attitudes of parents.

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 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 greatly improves precision of the estimates, as persistence is very high for achievement measures.

Table 13 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 very different on volunteers and non-volunteers. The bulk of the effects is attributable to large impacts on the subpopulation of volunteers (gains for test volunteers over control volunteers are as large as +15.1% of a standard deviation for French marks), with small and insignificant spillovers on their classmates.

These uneven results, and particularly 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. The overall impact on pupil achievement is low, which is not surprising, given that achievement was very indirectly targeted as an outcome, and some of the evoked mechanism for a transmission of

[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.

positive behaviors and attitudes to higher achievement probably need more time to deploy than allowed here. It can be noted, also, that the measured impact on non-volunteers is positive, but smaller and not significant at standard levels for most outcomes. Again, in the short term, we would indeed expect stronger peer effects on behavior than on learning achievement.

E. Robustness and Subgroup Analysis

In Table 14, we add control variables to regressions (3) and (4) and formally test for differences in impacts on volunteers and non-volunteers. All previous results are robust to the addition of controls.

Table 15 further expands the analysis to allow for different impacts across subgroups. Treatment dummies in Table 15 are fully interacted with group indicators for SES status (whitecollar 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.

Regarding 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 behavior. 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), neither directly nor indirectly. Our interpretation of improvements in the behaviour of non-volunteer pupils is thus classroom interactions and peer effects. The influence of peers can trickle along many channels, including direct influence of peers' behavior on own behavior or more indirect influence through progressive modification of the context of teaching and learning within the classroom. These different channels are very difficult to isolate, but there are nonetheless testable differences between such peer effects and simple placebo effects.

A possible placebo explanation for spillovers would link the observed improvement in the behaviour of all pupils to a change in teachers' attitude towards selected classes that is independent from the programme itself and that affects volonteer and non-volonteer students indifferently. Indeed, telling the staff that some classes were selected and other control could have had, as such, an impact on their behaviour 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 compared teachers' subjective marks during term 1 with initial test scores. Most interestingly, we find that the difference between teachers' marks and externally marked test scores is not significantly different in test and control classes[13].

Moreover, as discussed above, test and control classes to not differ either in honors and sanctions granted at the end of the first trimester by the teachers of the class during the class council. 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. Yet its impacts are only detectable at the end of the school year, after the non-volunteers in selected classes were exposed to repeated interaction with their treated peers. This constitutes direct 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 volonteer and non-volonteer pupils, whereas placebo effects are, by construction, independent from the "dose" of social interactions received by non-volonteer pupils. Any dose-response relationship between the magnitude of spillover effects and the level of classroom interactions is compatible with peer effects only.

Interestingly enough, 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 cross-sectional evidence is suggestive of a dose-response relationship between the quantity of interactions with treated peers to which non-volunteers were exposed and the quality of non-volunteers' behavior. 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 16). 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 16 is, however, the observational counterpart which comes closest to the thought experiment of creating a "placebo group" for this treatment, whose classes are announced to be part of the treatment, before the number of volunteers is exogenously set to 0, so that in effect nobody gets any treatment.

[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 Maths.

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 17 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 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 their ability to pass the easiest part of the French test. Given 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).

normalization used, the results suggest that the difference in parental involvement between the 25% classes where the involvement is maximum and the 25% classes where it is minimum 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 point 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 17 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. Most interestingly, 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.

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 positive and cooperative environment in those classes. With greater support from their family, school results of pupils in these classes show already some improvement by the end of the school-year: pupils' command of the most basic reading skills improves, possibly because loss of motivation among low-achieving pupils could have been prevented.

To our best knowledge, this is the first experimental evaluation of a school-based parental involvement program targeting education outcomes for middle-school children[15]. It remains thus an open question whether similar results would be obtained in different contexts: more experimentation is needed before the conditions for successful parenting programs are known. Also, the elements of design which lead to successful interventions need to be more thoroughly explored and rigorously tested.

Despite these limitations, the results of our study distil some encouraging implications for educational policy and can be used to inform the next wave of experiments.

Our results show that in poor neighborhoods low levels of parental involvement are not a fatality. 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 results 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 behaviors have 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 and what 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 providing support only to at-risk families. Targeting entire communities where risky behavior is more prevalent, rather than individuals, has the advantage of minimizing the stigma associated with individual

[15] Scott *et al.* (2006) is the only other experimental evaluation study in this literature; its focus is on a parenting program for parents of five- and six-year-olds in poor boroughs in London.

targeting of remedial programs. In the context of parenting programs, this does not come 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. Even if all take-up costs faced by parents are considered in addition to the program's costs for the school, the net benefits from this school based parent support scheme could probably be increased if more parents took up this program. This is because even the most rational among parents do not internalize the large positive externalities on classmates of their efforts. If this is the case, the provision of small and targeted incentives to compensate some parents for their effort to attend school-based meetings can be justified.

More largely, parental involvement decisions are the result of an individual arbitrage between costs and benefits. An experiment which manipulates exogenously some of the elements on either side of this arbitrage might not only provide interesting policy lessons on how to reach the socially optimal level of involvement, but also cast new light on the interplay of private costs and benefits, and thereby contribute to the large literature in sociology and education on the determinants of parental involvement.

References

AIZER, ANNA. 2004. Home alone: supervision after school and child behavior. *Journal of Public Economics*, **88**, 1835–1848.

ANGRIST, JOSHUA, & LANG, KEVIN. 2004. Does School Integration Generate Peer Effects? Evidence from Boston's Metco Program. *American Economic Review*, **94**(5), 1613–1634.

BANERJEE, ABHIJIT, BANERJI, RUKMINI, DUFLO, ESTHER, GLENNERSTER, RACHEL, & KHEMANI, STUTI. 2008 (September). *Pitfalls of Participatory Programs: Evidence from a Randomized Evaluation in Education in India.* NBER Working Paper No. 14311.

BJÖRKMAN, MARTINA, & SVENSSON, JAKOB. 2009. Power to the People: Evidence from a Randomized Field Experiment on Community-Based Monitoring in Uganda. *Quarterly Journal of Economics*, **124**(2), 735–769.

CUNHA, FLAVIO, & HECKMAN, JAMES J. 2008. Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation. *The Journal of Human Ressources*, **XLIII**(4), 739–782.

DESFORGES, CHARLES, & ABOUCHAAR, ALBERTO. 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., & SANDLER, HOWARD M. 1995. Parental Involvement in Children's Education: Why Does it Make a Difference? *Teachers College Record*, **97**, 310–331.

HOOVER-DEMPSEY, KATHLEEN V., & SANDLER, HOWARD M. 1997. Why Do Parents Become Involved in Their Children's Education? *Review of Educational Research*, **67**, 3–42.

HOXBY, CAROLINE. 2000. *Peer Effects in the Classroom: Learning from Gender and Race Variation.* NBER Working Paper No. 7867.

MOFFITT, ROBERT A. 2000. Policy Interventions, Low-Level Equilibria, and Social Interactions. *Pages 45–82 of:* DURLAUF, STEVEN N., & YOUNG, H. PEYTON (eds), *Social Dynamics.* The MIT Press, for Brookings Institution.

OECD. 2004. *Learning for Tomorrow's World – First Results from PISA 2003.* Organisation for Economic Co-Operation and Development.

SCOTT, STEPHEN, O'CONNOR, THOMAS, & FUTH, ANNABEL. 2006. *What makes parenting programmes work in disadvantaged areas?* Joseph Rowntree Foundation.

TODD, PETRA E., & WOLPIN, KENNETH E. 2007. The Production of Cognitive Achievement in Children: Home, School, and Racial Test Score Gaps. *Journal of Human Capital*, **1**(1), 91–136.

WELSCH, DAVID M., & ZIMMER, DAVID M. 2008. After-School Supervision and Children's Cognitive Achievement. *The B.E. Journal of Economic Analysis & Policy*, **8**, art. 49.

Table 1: Relation Across Outcome Measures for Pupil Behavior (Term 3)

Relation between indicators

	sanction = 1				*sanction = 0*		
	$h = 0$	$h = 1$			$h = 0$	$h = 1$	
$c = 0$	0.11	0.00	*0.11*	$c = 0$	0.39	0.13	*0.52*
$c = 1$	0.00	0.00	*0.00*	$c = 1$	0.12	0.24	*0.36*
	0.11	*0.00*	***0.11***		*0.51*	*0.38*	***0.89***

average absenteeism

	sanction = 1				*sanction = 0*		
	$h = 0$	$h = 1$			$h = 0$	$h = 1$	
$c = 0$	9.75	.	*9.57*	$c = 0$	4.60	1.82	*3.87*
$c = 1$	.	.	*.*	$c = 1$	1.81	1.19	*1.38*
	9.60	*.*	***9.71***		*4.01*	*1.43*	***3.27***

legend: h = *honors*; c = *good conduct*

89% of pupils don't have sanctions during the third term. 36% of pupils earn the top conduct mark, and 38% are bestowed honors from the *conseil de classe*. 24% of pupils have both honors and the top conduct mark. Pupils without sanctions have skipped on average 3.27 half-days of school without justification during the third term. Among those with the top conduct mark, this average falls to 1.38 half-days; it is of 1.19 half-days for pupils with both the top conduct mark and honors. Average absenteeism varies in a predictable manner with other indicators of behavior.
Notes: Only observations for which all measures are available are used in this table. Sample size is 2399.

Table 2: Difference in Pre-Treatment Characteristics (Volunteers vs. Non-Volunteers)

	NV	std	V - NV	(se)	n.obs.
parents					
Employment status	0.85	0.35	**-0.002**	(0.013)	4660
Intact family	0.72	0.45	**0.038****	(0.015)	4728
White collar	0.18	0.39	**0.025***	(0.014)	4529
children					
Girl	0.48	0.50	**-0.023**	(0.018)	4728
6th grade repetition	0.06	0.24	**-0.013***	(0.007)	4728
Age (sept 2008)	11.46	0.57	**-0.060****	(0.020)	4728
tests					
French test (sept 2008)	0.01	1.00	**-0.044**	(0.040)	4165
Maths test (sept 2008)	0.00	1.00	**0.004**	(0.042)	4167

85% of pupils from non-volunteer families have at least one caring adult which is employed. This proportion is 0.2% lower among volunteer pupils.
Notes: Column "V-NV" displays the coefficient from the regression of the row variable on a volunteer dummy and school fixed effects. Robust standard errors allowing for correlated residuals within classes are shown in parentheses.
*: Significant at the 10% level. **: significant at the 5% level.

Table 3: Difference in Pre-Treatment Characteristics across Treatment Arms.

	mean	std	T - C	(se)	n.obs.
parents					
Employment status	0.85	0.36	**-0.002**	(0.010)	4660
Intact family	0.73	0.44	**-0.002**	(0.013)	4728
White collar	0.19	0.39	**0.001**	(0.011)	4529
children					
Girl	0.48	0.50	**0.006**	(0.010)	4728
6th grade repetition	0.06	0.23	**0.000**	(0.005)	4728
Age (sept 2008)	11.45	0.57	**-0.002**	(0.019)	4728
tests					
French test (sept 2008)	-0.01	1.00	**-0.048**	(0.044)	4165
Maths test (sept 2008)	-0.01	1.00	**0.008**	(0.036)	4167
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	**0.004**	(0.014)	3672
White collar	0.18	0.39	**0.006**	(0.012)	3492
children					
Girl	0.48	0.50	**0.012**	(0.012)	3672
6th grade repetition	0.06	0.24	**-0.001**	(0.006)	3672
Age (sept 2008)	11.46	0.57	**-0.002**	(0.021)	3672
tests					
French test (sept 2008)	0.01	1.00	**-0.058**	(0.047)	3180
Maths test (sept 2008)	0.00	1.00	**0.020**	(0.040)	3173

Notes: Column "T - C" displays the coefficient from the regression of the row variable on a test dummy and school fixed effects. Robust standard errors allowing for correlated residuals within classes are shown in parentheses. *: Significant at the 10% level. **: significant at the 5% level.

Table 4: 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

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

Table 5: Program Variants and Intensity

	schools	(sessions)	families	(sessions)
initial workshops				
at least 1 debate	37		384	
at least 2 debates	36		219	
all 3 debates	29		104	
additional workshops				
parenting	17	(3.1)	80	(2.7)
internet	15	(5.7)	57	(4.0)
French as foreign language	8	(6.0)	19	(4.7)
any of the above	26	(7.2)	118	(3.9)

37 schools organized at least 1 debate. 384 families attended at least one debate. Additional workshops on parenting issues were organized by 17 schools (3.1 sessions on average). They were attended by 80 different families, each family attending on average 2.7 sessions.

Table 6: 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

Notes: Column "T - C" displays the coefficient from the regression of a response dummy on a test dummy and school fixed effects. Robust standard errors allowing for correlated residuals within classes are shown in parentheses.
*: Significant at the 10% level. **: significant at the 5% level.

Table 7: Pupil Outcomes: Response Rate

outcome	mean C	T - C	(se)	resp.	sch	cl	ind
	response rate				*source of attrition*		
Pupil Behavior (Term 3)							
absenteeism	0.74	0.004	(0.006)	3401	28	0.05	0.07
behav. score	0.94	0.003	(0.006)	4467	37	0.07	0.06
discipl. sanctions	*0.88*	*-0.001*	*(0.005)*	*4198*	*35*	*0.07*	*0.05*
good conduct	*0.61*	*0.010*	*(0.021)*	*2971*	*28*	*0.16*	*0.07*
honors	*0.90*	*-0.016*	*(0.013)*	*4234*	*36*	*0.08*	*0.07*
Teacher Marks & Tests (Term 3)							
all							
Teacher Marks	0.89	0.026	(0.017)	4271	37	0.02	0.08
French Tests (June 2009)	0.81	-0.018	(0.022)	3734	35	0.05	0.14
Maths Tests (June 2009)	0.80	-0.009	(0.015)	3707	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)	3262	37	0.02	0.10
French Tests (June 2009)	0.79	-0.019	(0.023)	2853	35	0.05	0.15
Maths Tests (June 2009)	0.79	-0.016	(0.016)	2837	35	0.04	0.17

Notes: Column "T - C" displays the coefficient from the regression of a response dummy on a test dummy and school fixed effects. 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: Impact of the Program on Parental Attitudes and Behavior

Dependent Variable	mean C	T - C	(se)
	all		
Global Parenting Score	-0.072	0.119**	(0.035)
School-Based Involvement Score	-0.053	0.127**	(0.035)
Home-Based Involvement Score	-0.024	0.057*	(0.031)
Understanding & Perceptions Score	-0.028	0.064*	(0.037)
Never been summoned to the school	0.79	0.028*	(0.015)
	volunteers		
Global Parenting Score	-0.141	0.266**	(0.071)
School-Based Involvement Score	0.172	0.320**	(0.076)
Home-Based Involvement Score	0.015	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	-0.050	0.044	(0.040)
School-Based Involvement Score	-0.124	0.001	(0.036)
Home-Based Involvement Score	-0.036	0.019	(0.040)
Understanding & Perceptions Score	0.021	0.024	(0.041)
Never been summoned to the school	0.81	0.009	(0.016)

Notes: 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. 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 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 (see table 6).
*: Significant at the 10% level. **: significant at the 5% level.

Table 9: Parental Attitudes and Behavior Scores: Difference by Socio-Economic Status

outcome	mean oth.	whitecollar	(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)

Notes: Column "whitecollar" displays the coefficient from the regression of the score variable on a whitecollar dummy and school fixed effects. Only control classes are used in these regressions. 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 10: Impact of the Program on Parental Attitudes and Behavior: Volunteers Only (Raw Indicators)

Question	mean C	T - C	(se)
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)

Notes: 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. 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 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 834 (see table 6).
*: Significant at the 10% level. **: significant at the 5% level.

Table 11: Impact of the Program on Pupils' Behavior (Term 3)

outcome	mean C	T - C	(se)	std	n.obs.
		all			
absenteeism	4.324	-0.711**	(0.296)	7.737	3401
behav. score	-0.013	0.106**	(0.037)	1.024	4467
discipl. sanctions	*0.109*	*-0.025***	*(0.011)*	*0.296*	*4198*
good conduct	*0.326*	*0.048***	*(0.024)*	*0.481*	*2971*
honors	*0.345*	*0.040***	*(0.016)*	*0.482*	*4234*
		volunteers			
absenteeism	4.217	-0.771	(0.549)	7.737	786
behav. score	-0.012	0.117*	(0.066)	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	*0.352*	*0.018*	*(0.029)*	*0.482*	*1006*
		non volunteers			
absenteeism	4.351	-0.600*	(0.337)	7.737	2615
behav. score	-0.014	0.098**	(0.042)	1.024	3422
discipl. sanctions	*0.110*	*-0.021**	*(0.012)*	*0.296*	*3223*
good conduct	*0.336*	*0.048**	*(0.026)*	*0.481*	*2295*
honors	*0.343*	*0.046***	*(0.019)*	*0.482*	*3228*

Notes: Column "T - C" displays the coefficient from the regression of the dependent 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 12: Difference in Pupils' Behavior in Term 1

outcome	mean C	T - C	(se)	std	n.obs.
		all			
absenteeism	1.125	0.015	(0.112)	2.975	3825
behav. score	-0.115	0.057	(0.040)	1.160	4605
discipl. sanctions	*0.086*	*-0.016*	*(0.011)*	*0.265*	*3869*
good conduct	*0.385*	*0.020*	*(0.028)*	*0.491*	*2903*
honors	*0.453*	*0.011*	*(0.016)*	*0.499*	*4302*

Notes: Column "T - C" displays the coefficient from the regression of the dependent 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 13: Impact of the Program on Pupils' Achievement Gains

	impact (std pts)		*base points (T3)*		
outcome	T - C	(se)	avg (C)	std	n.obs.
		all			
Teacher Marks					
French	0.065*	(0.036)	10.8	3.8	4271
Maths	0.005	(0.038)	10.9	4.3	4271
Tests					
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			
Teacher Marks					
French	0.151**	(0.048)	10.7	3.8	1009
Maths	0.024	(0.054)	10.9	4.3	1009
Tests					
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			
Teacher Marks					
French	0.040	(0.038)	10.9	3.8	3262
Maths	0.003	(0.038)	10.9	4.3	3262
Tests					
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

Notes: The dependent 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 dependent variable on a test dummy. 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. All regressions include 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. Columns 4 and 5 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 14: Impact Controlling for Baseline Characteristics

Panel A: Parental Attitudes and Behavior

	Global Score		**School-inv. Score**		**Home-inv. score**		**Und. & Perc. Score**		**Never Summoned...**	
	(3)	(4)	(3)	(4)	(3)	(4)	(3)	(4)	(3)	(4)
T - C	.100**		.113**		.055*		.070*		.018	
	(.032)		(.035)		(.032)		(.037)		(.014)	
T - C (V)		.264**		.376**		.103*		.169**		.060**
		(.064)		(.075)		(.053)		(.069)		(.027)
T - C (NV)		.031		.001		.035		.028		.000
		(.038)		(.038)		(.039)		(.043)		(.016)
N	3026	3026	3024	3024	3025	3025	3018	3018	3013	3013
p-value		.002		.000		.294		.080		.059

Panel B: Pupils' Behavior

	Absenteeism		**Behav. Score**		**Sanctions**		**Good C'duct**		**Honors**	
	(3)	(4)	(3)	(4)	(3)	(4)	(3)	(4)	(3)	(4)
T - C	-.741**		.142**		-.032**		.066**		.059**	
	(.296)		(.033)		(.011)		(.023)		(.015)	
T - C (V)		-1.116*		.176**		-.044**		.080**		.043*
		(.571)		(.055)		(.019)		(.036)		(.025)
T - C (NV)		-.622*		.131**		-.027**		.062**		.064**
		(.319)		(.037)		(.012)		(.024)		(.018)
N	3401	3401	4467	4467	4198	4198	2971	2971	4234	4234
p-value		.419		.467		.398		.596		.499

Panel C: Pupils' Achievement Gains

	French Mark		**Maths Mark**		**French Test**		**French Test: Observation**		**Maths Test**	
	(3)	(4)	(3)	(4)	(3)	(4)	(3)	(4)	(3)	(4)
T - C	.073*		.017		-.002		.086*		-.011	
	(.038)		(.036)		(.036)		(.044)		(.036)	
T - C (V)		.171**		.001		-.020		.216**		-.024
		(.051)		(.056)		(.052)		(.060)		(.051)
T - C (NV)		.041		.022		.003		.042		-.006
		(.039)		(.035)		(.038)		(.048)		(.038)
N	4301	4301	4328	4328	3734	3734	3734	3734	3707	3707
p-value		.004		.648		.666		.006		.712

Notes: Columns labeled (3) report estimates from an augmented version of equation (3); columns labeled (4) report estimates from an augmented version of equation (4). All regressions 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. Each column in each panel corresponds to a separate regression. Robust standard errors allowing for correlated residuals within classes are shown in parentheses. For model (4), the p-value for the hypothesis of equal effects on volunteers and non-volunteers is reported.
*: Significant at the 10% level. **: significant at the 5% level.

Table 15: 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	

Notes: The table reports estimates from an augmented version of equation (4), where the variables of interest are fully interacted with subgroup dummies. All regressions 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.
*: Significant at the 10% level. **: significant at the 5% level.

Table 16: 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

Notes: The table reports estimates from an augmented version of equation (2), where the variable of interest is fully interacted with dummies for low ($<.16\%$) and medium to high proportion of volunteers in class ($\geqslant 16\%$). 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. 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 17: 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)
sch. f.e.	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)
sch. f.e.	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)
sch. f.e.	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

Notes: 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.

*: Significant at the 10% level. **: significant at the 5% level.

Figure 1: Flow diagram for the Field Experiment.