

Getting Parents Involved: A Field Experiment in Deprived Schools

FRANCESCO AVVISATI

OECD

MARC GURGAND

Paris School of Economics

NINA GUYON

Department of Economics, National University of Singapore

and

ERIC MAURIN

Paris School of Economics

First version received November 2011; final version accepted June 2013 (Eds.)

This article provides evidence that schools can influence parents' involvement in education, and this has causal effects on pupils' behaviour. Furthermore, it shows how the impact of more involved parents on their children is amplified at the class level by peer group interaction. We build on a large-scale controlled experiment run in a French deprived educational district, where parents of middle-school children were invited to participate in a simple program of parent–school meetings on how to get better involved in their children's education. At the end of the school year, we find that treated families have increased their school- and home-based involvement activities. In turn, pupils of treatment classes have developed more positive behaviour and attitudes in school, notably in terms of truancy and disciplinary sanctions (with effects-size around 15% of a standard deviation). However, test scores did not improve under the intervention. Our results suggest that parents are an input for schooling policies and it is possible to influence important aspects of the schooling process at low cost.

Key words: Parental involvement, Cluster randomized trial, Classroom peer effects, Child support

JEL Codes: I21, J13, J18

1. INTRODUCTION

A vast body of research documents the importance of education for explaining a variety of life outcomes, including wages, employment, health, or participation in criminal activity. This literature has driven considerable interest among economists on how skills are acquired at school. Much research on the educational production function has focused on school-level inputs, such as class size or teacher quality, which are perceived as the most natural instruments for policy intervention. Attention has also been devoted to parental inputs, such as parents' involvement at school, and a very large body of work documents the relation between family background and educational outcomes (Sirin, 2005). As such, parental attitude and involvement at school can be

perceived as additional key inputs for educational success. However, little is known on whether these inputs can effectively be manipulated through policy initiatives. This study demonstrates that parental attitudes and school involvement can be significantly upgraded through simple participation programs and that such policies have a potential for reducing disciplinary problems in young teenagers.

Using a large-scale randomized control trial undertaken in Paris area middle schools, we show that a very simple and low-cost program of parent–school meetings increased the level and quality of school-related parental care; and that this enhanced involvement of parents translated into a significant reduction of reported truancy and misbehaviour in treatment classes, as well as into improved motivation for school work. The order of magnitude of impacts is that of average differences between families at the top and the bottom of the socio-economic scale in the control group. Moreover, while only a limited share of parents took part in the program, we build on the “partial population” design of our experiment (Moffitt, 2001) to show that the behaviour of all pupils within a class was affected, including those whose parents did not participate, and provide evidence that spillovers result from direct interactions between pupils.

The experiment was run during the 2008–2009 school year, in 34 middle schools of a relatively deprived educational district. In 183 classes, almost 1000 parents (22.5%) of sixth graders agreed to enroll in a program of three meetings with the school head on how to successfully manage the transition from primary school to middle school.¹ The program offered information on the functioning of schools and advice on how to support and monitor children with school work. Out of the 183 classes, 96 were randomly chosen to effectively run the program in November and December 2008. In each school and each class we are therefore able to clearly identify volunteer and non-volunteer families prior to the random decision of running the program. By comparing volunteer families in treatment classes to volunteer families in control classes, we capture the overall effect of the program on the treated parents and their children. By comparing children from non-volunteer families across treated and untreated classes, we capture the indirect effect of having treated families in the class.

The program is found to have a very positive impact on school-related involvement of enrolled parents. For instance, at the end of the year, the proportion of enrolled parents who actively participated in the school’s parents’ association is 37% in treatment classes, whereas it is only of 25% in control classes. As a consequence, the behaviour of pupils is affected along many dimensions. We find that the program is associated, by the end of the school year, with a decline of about 25% in truancy in treatment classes (from about four to about three half-days in a trimester), and an even larger decline in the probability of being sanctioned. At the same time, institutional reports of positive attitudes and behaviour improve, suggesting that the impact is not limited to the lower tail of the distribution. Generally, effects-sizes are in the range of 10–20% of an SD (standard deviation). Pupils in treatment classes also receive better marks from their teachers, especially in French, but we do not find impacts on externally set and marked tests of literacy and numeracy. Finally, changes in pupils’ behaviour are not limited to children of enrolled families: improvements spread out to their classmates. In particular, children from non-volunteer families are significantly less absent and receive fewer disciplinary sanctions when they are in treatment classes. Therefore, policies targeted to parents are not bound to benefit a small fraction of volunteer families only.

Teachers were not associated with the project and had no direct interest in its success. Nevertheless, we cannot exclude that they knew about pupils’ treatment status and that this knowledge influenced their judgement and marks. Even if they did not know who is treated, they could simply

1. The transition between primary and middle school represents one of the most critical stages of an educational career and this is why we chose to focus on 6th grade, that is, the first year of secondary education in France.

give credit for better effort from parents. In this scenario, our results could reflect changes in teachers' perceptions rather than changes in students' outcomes. In general, it is hard to entirely rule out concerns about the subjective nature of teachers' assessments. We nonetheless provide some evidence in favour of our interpretation: notably, impacts are not found early in the academic year, when the program is most salient, and remain perceptible at the end of 7th grade, 18 months after the end of the program, despite the fact that most pupils have moved to different classes with different teachers. Also, the measure of absenteeism is ruled by strong legal requirements and it is very unlikely that teachers or school administration may have misreported this information. Finally, spillover effects on non-volunteer pupils seem difficult to interpret as reflecting parents' efforts to elicit empathy from teachers, since we do not observe spillover effects on non-volunteer parents.

The most significant impacts are found on measures related to student's behaviour (truancy, discipline, work effort, good conduct). This is consistent with the psychology literature suggesting that behavioural dispositions are more malleable than cognitive skills, especially after the age of 10 and during adolescence (Knudsen *et al.*, 2006). Impacts on such outcomes are of particular interest because these factors have been documented to play a central role in explaining many economic outcomes (see, *e.g.*, Heckman *et al.*, 2006). In their study of the long-run effects of preschool programs, Heckman *et al.* (2011) and Chetty *et al.* (2011) suggest that most of these effects can be attributed to the programs' positive impacts on non-cognitive skills. As discussed by Brunello and Schlotter (2011), there exists much less evidence on the determinants of non-cognitive skills than of cognitive skills and, again, most existing work is about the role of teachers and other school-level inputs. Our article shows that increased parental involvement may represent an efficient alternative way to foster the acquisition of relevant attitudes.

These findings lie within the scope of several strands of the literature. They contribute first to the economic literature on the importance of parental input for children's education.² Existing economic studies analyse parental inputs using non-experimental data and structural approaches, (Todd and Wolpin, 2007; Cunha and Heckman, 2008), or fixed effects strategies (Aizer, 2004; Welsch and Zimmer, 2008). The education literature has also produced a number of estimates of the relation between parental involvement and pupil achievement but most of them are either very small scale or based on simple correlations.³ If we think of our intervention as providing information and training to parents, our results can also be connected to the early childhood literature, which suggests that home visitation interventions aiming at improving parents' teaching skills has positive effects on early childhood development (Grantham-McGregor *et al.*, 1994; Kagitcibasi *et al.*, 2001; Olds, 2002).

Our article is also related to the large and still growing literature on social interactions and spillover effects in education (see, *e.g.*, Angrist and Lang, 2004; Carrell and Hoekstra, 2010). Most interestingly, we find no spillover effects on non-volunteer parents, but strong spillovers on their children, especially on behavioural outcomes and, again, with peak improvement by the end of the school year. These findings 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.

The remainder of this article proceeds as follows. Section 2 provides background information on the context in which the program took place and describes the intervention and its objectives, as well as the experimental design. Section 3 introduces outcome measures, performs balancing tests on baseline data and discusses take-up and attrition. Section 4 presents the estimation strategy and the main results of the study. Section 5 provides robustness analysis to ensure that results

2. For a recent survey, see Avvisati *et al.* (2010).

3. See Desforges and Abouchaar (2003) or Hill and Tyson (2009).

are not driven by reporting biases. Section 6 discusses the generalizability of our results, and Section 7 concludes with implications for policy and future research.

2. PROGRAM AND EXPERIMENTAL PROTOCOL

2.1. *French middle schools*

Middle school (*college*) in France runs from grade 6 to grade 9; children enter sixth grade at the age of 11. There is no streaming by ability across schools, and parents are not free to choose the State school that their children will attend (otherwise than through residential location). For sixth graders, a typical week consists of 29 school hours, distributed across 9 different subjects. A different teacher teaches each subject.

Pupils stay in the same class throughout the school year, and in every subject. The class is therefore a very distinct and closed entity where most of the interactions with same age children take place. Classes are groups of 20–30 pupils. The school head allocates children to classes before the beginning of the school year, taking into account possible elective subjects. In contrast, the school head is not allowed to sort children by ability. Each class has one specific reference teacher (*professeur principal*). A class council (*conseil de classe*) is formed for each class and composed of all teachers of the class, and representatives of parents, pupils, and the school administration.

The year is divided into three terms. At the end of each term, the class council meets to discuss each pupil's work, achievement, and behaviour. The council bestows honours and warnings that are transmitted to families on the report card, together with teacher grades; at the end of the year, the council decides on grade repetition, and on the elective courses that each pupil will be allowed to take in the future. Only the best pupils 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).

2.2. *The intervention*

The experiment took place in the educational district of *Créteil*, which includes all suburbs located to the east of Paris, with about 4 million inhabitants. During the school year 2008–2009, this district wanted to test out a program to improve parental involvement in school, because it was strongly perceived that disadvantaged parents have inadequate knowledge of and confidence with schools, and that this situation could be potentially improved by a simple intervention. The program was targeted at families of 6th graders.

The experimental program consisted mainly of a sequence of three meetings, which took place every 2–3 weeks, between November and December (early January in some cases). Only parents were invited to these meetings and not their children. Sessions started at 6 PM at the school and lasted typically 2 h. Most of the time, they were conducted by the school head. He or she could draw on precise guidelines, designed by the districts' educational experts, and show 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 were distributed at these meetings to explain the functioning and opportunities of the school attended by the children.

The two initial sessions of the program focused on how parents can help their children by participating at school and at home in their education. The last session took place after the first class council and end-of-term report card. It offered parents advice on how to adapt to the results of the first term. At the end of the third session, the facilitator asked whether participants would like to attend additional sessions, but they only had a marginal turnout.

District-level guidelines insisted that the facilitator of the discussion 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 do well, 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 the administration. Generally, the topics developed during the meetings insisted on the ways (drawn from role-model and efficacy theories in psychology) in which parents can increase the effort exerted by pupils by giving them interest, attention, praise, and rewards related to the behaviour that leads to school success. The program and its material were developed by educational experts at the district level in accordance with state-of-the-art psychological literature on parental involvement (see, *e.g.*, Hoover-Dempsey and Sandler, 1995).

2.3. *Experimental protocol*

Over the summer before the start of academic year 2008–2009, 34 school heads volunteered to participate in the experimental study. They represent about 10% of the 352 State-run middle schools from the *Créteil* district. As a result, the universe to which the program was offered, and baseline and follow-up data were collected, consists of 34 schools, representing 183 classes, and the families of some 4300 pupils. Almost two-thirds of the schools that entered the study were located in a “priority education” zone—a label that distinguishes historically deprived areas. The district average is about one-third, which implies that disadvantaged schools were more likely to consider the program. Many families attending these schools are relatively poor: about 32% families received a (means tested) scholarship, compared with a district average of 18%.

Just after the start of academic year, during September 2008, the schools advertised the program to the families of their 6th graders. School heads were provided with a standardized leaflet to inform families. The program was presented as an outreach effort, distinct from usual parent–teacher meetings. Parents of 6th graders were told that the school would organize a series of three evening meetings 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. Given this information, parents were invited to volunteer for participation in the program. It was always explicit that actual eligibility to the program would occur only conditional on a random selection of eligible classes. The leaflet explicitly stated that the experimental nature of the program implied a limitation on the number of beneficiary classes. By mid-October each school listed all families who signed up, and closed the registration phase. This list defines the population that we call “volunteer families”, and has not been amended thereafter. Volunteer families constitute approximately one-fifth of the total population (970 out of 4308).

Overall, the initial information campaign defined two distinct populations within each school and each class: volunteers and non-volunteers. Appendix Table A.1 (based on administrative data collected before the start of the academic year) reveals that there are no strong observable pre-treatment differences between the two populations. This fact may be a consequence of the principals’ efforts to inform all categories of parents, even those whose involvement is usually very weak.

Randomization took place as soon as the registration phase was closed (mid-October). As all schools in the sample have several 6th grade classes, randomization was at the class level.

We randomized classes within each school at a uniform rate of: $m/2$ if the number of eligible classes in the school, m , was even; or $(m+1)/2$ if m was odd.⁴ Therefore, either half or slightly more than half of the classes was treated within each school. This procedure was carried out separately within each school, in the presence of the school head.⁵ It resulted in the selection of 96 classes in the treatment group and 87 classes in the control group. Appendix Table A.2 shows that observable characteristics, as measured at the beginning of the year, before treatment, are balanced across treatment and control groups. Following randomization, volunteer families belonging to treatment classes were informed by the administration of the exact dates at which the three meetings would take place.

The randomization procedure defines four basic groups of families within each school: volunteers in treatment classes; non-volunteers in treatment classes; volunteers in control classes; non-volunteers in control classes. Within this “partial population” framework (Moffit, 2001), any difference in outcomes between volunteers in treatment and control groups will capture the causal effect of eligibility to the program on the population of volunteers. In contrast, any difference in outcomes between non-volunteers in treatment and control groups will capture the causal effect of having eligible peers on the population of non-volunteers. This will be interpreted as a treatment spillover.

2.4. Program take-up

At the beginning of each session, we asked participants to sign in and collected the attendance lists. Table 1 shows the effective take-up rates across the four basic categories of families. Comfortingly, take-up is large for volunteers in treatment classes only, even though it remains far from 100%. Specifically, a little more than 50% of families in this group participated in at least one session, and about 14% attended all the three basic meetings. School heads were not to invite other parents, nor even to inform them about session times. As a result, only a very small share in the control group did attend some meetings.

As discussed above, families attending the last meetings in the initial program could determine whether, and in what form, to continue with additional sessions. Overall, the number of families which participated in at least one additional session beyond the initial three meetings makes up only about 18% of eligible parents. For most eligible parents, the program consisted in receiving additional invitations to the school over the school year and in participating to the initial three meetings only.

3. OUTCOME MEASURES

We are interested in three types of outcomes: (1) parental involvement attitudes and behaviour; (2) children’s behaviour, as reflected by truancy, disciplinary record and work effort; and (3) children’s academic results. We have several sources to measure those outcomes, which are described in this section altogether with an in-depth analysis of attrition rates.

4. 15 classes had 0 volunteer families: they were excluded from the whole process. In schools with an uneven number of classes, when a class had <4 volunteers, it was grouped with the class just above in terms of number of volunteers and the two formed a single randomization unit: the procedure for an even number of units was then applied. All empirical analyses use weights defined as the inverse of the probability of assignment.

5. 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.

TABLE 1
Take-up rate

	Treatment		Control	
	Volunteers	Non-vol's	Volunteers	Non-vol's
Initial workshop				
At least 1 debate	54.7	1.0	1.9	0.2
At least 2 debates	30.8	0.2	0.7	0.0
All 3 debates	13.5	0.1	0.7	0.0
Additional workshops	18.0	0.9	0.5	0.0

Note: All rates are expressed in percentage terms and are computed separately for each of the four groups.

3.1. Parent questionnaire

To assess the impact of the program on parental involvement attitudes, we distributed a questionnaire to all families at the end of the school year. The parent questionnaire is a self-administered short questionnaire, with 11 questions on school-based involvement, home-based involvement as well as on parents' perception of the school.⁶ A final question asked whether parents have been summoned by the school administration to discuss their child's behaviour or academic results. Being summoned is not only a symptom of the child's insufficient discipline or work but also a consequence of a lack of regular contact with the school.

Parents' answers to this questionnaire define 12 elementary indicators for parental involvement. Each indicator is standardized using the control group mean and SD. By summing these indicators across the full questionnaire, as well as within each section, these elementary measures have been converted into four synthetic scores—a global parenting score, a school-based involvement score, a home-based involvement score, and an understanding and perception score.

3.2. Administrative records

Our measures of pupils' outcomes are mostly based on administrative registry data. Truancy is measured as the number of half-days in the term that the child is not at school without a valid justification from his parents (an occasional hour skipped counts as a half-day). This is a very objective measure: French law is very specific about the legal responsibilities of schools and the way they should handle truancy.⁷ In case of accidents or problems occurring to an absent child, the school remains responsible until parents are informed of the absence. In such a context, it is not likely that recorded truancy could be affected by teachers' subjective perceptions or by the empathy that they may have for some parents or some children.

6. The questionnaire consisted of three questions on school-based involvement (participation in parents' organization—a necessary condition for being a representative in the class council—participation in parents/teachers information sessions, individual appointments with teachers), four questions on home-based involvement and parental control (help with homework, knowledge of grades, control overtime spent watching TV, and control overtime spent on videogames), and four questions on understanding and general perception (knowledge of available optional courses, plans about child's future, satisfaction with school, and anxiety about violence).

7. At the start of each course, teachers have to register attendance and must inform immediately the head supervisor of any unauthorized absence. After collecting information from all classes, the head supervisor must contact the parents as quickly as possible to identify the causes of absences. The head supervisor must also record absences for each child: if a child is absent without justification four times or more in a given month, the head supervisor has to contact the family and try to start a dialogue with parents. See Ministère de l'Éducation Nationale, *Circulaire 96-248* (October 25, 1996) and *Circulaire 2004-054* (March 23, 2004). See also *Le guide juridique du chef d'établissement* (2009).

We also define a “disciplinary sanction” dummy that takes value 1 if the child has received an official “disciplinary warning” or was temporarily excluded during each term. Temporary exclusions signal violent behaviour or repeated transgression of the rules. The school head sentences them.

Next, we can also use a “good conduct mark” (called *note de vie scolaire*) that is given each term in most schools and should reflect assiduity, contribution to the school and respectful behaviour.⁸ It usually has a much skewed distribution, with about 30% pupils having either the maximum (20/20) or next-to-maximum mark (19/20). Our variable “good conduct” is a dummy that takes value 1 for those top values and 0 otherwise.

Another type of measure, also recorded in school registries, reflects teacher judgement about attitude in class and work involvement. Each term, the class council meets and comments on pupils’ academic results: when its general impression is positive, the pupil gets honours (given to about 30%). They are not just a reflection of marks in each topic, but also a judgement on the general attitude and an encouragement to persevere. They can take three values: we code a variable named “honours” as taking value 0 if no honour is granted, 1 for the lower level (*encouragements*), 2 for the second level (*compliments*), and 3 for the highest level (*félicitations*).

Together, this set of variables sheds light on changes taking place at both ends of the distribution of behaviour and work attitudes. The “disciplinary sanction” variable separates the small proportion of pupils with heavy conduct problems, and can therefore measure an improvement occurring at the bottom of the distribution of behaviour. In contrast, “good conduct” and “honours” make it possible to capture improvement among better performing pupils. In turn, truancy increases sharply from an average of one half-day per term to nine half-days per term, as we move from pupils with honours and top marks in good conduct to pupils with a disciplinary sanction. Truancy can thus be considered as providing an objective and continuous measure of the quality of pupil behaviour over the whole distribution of pupil types.

Finally, we have data on children’s academic achievement. First, for each subject, we collected the teacher-given marks from end-of-term report cards. We use marks given in French and mathematics, as well as a weighted average of all subjects (with weights proportional to hours of instruction per week). Second, we can use a national test in mathematics and French conducted at the beginning of 6th grade in all middle schools. It is uniform across schools and classes, and was externally graded. We complemented that information with a second test in mathematics and French, specifically designed by the district pedagogical services, which was submitted to all classes at the end of the year, and also externally graded.

The administrative database also contains baseline information from a registration form collected in July 2008, when parents registered their children for the upcoming school year: this gives us demographic information on child and family (gender, date of birth, family situation, socioeconomic status, etc.), as well as an exhaustive count of the universe of pupils.

3.3. Reference teacher questionnaire

In each class one teacher acts as “reference teacher”. He or she can be a teacher of any subject, although main subjects (mathematics and French) are over-represented. Any teacher can be a reference teacher for one class only. This teacher is thus most knowledgeable about the class, and he is the one most likely to meet with parents. At the end of the year, we submitted a self-administered short questionnaire to the reference teacher of each class.

For every pupil in the class, we first asked reference teachers whether their dialogue with the child’s parents was satisfactory and whether they felt that parents did provide support to the child

8. Ministère de l’Education nationale, *Circulaire DGSCO* n°2006-105, June 23, 2006.

with school work. We also asked two questions about the child's attitude in class: how agreeable he/she is in class and whether he/she works diligently. Finally, we asked how much progress the pupil made over the year. The first two questions can be used to complement measures of parental attitudes, whereas the other questions provide additional measures for child's behaviour and school performance.

3.4. *Longer-term outcomes*

By the end of the following school year (2009–2010), we went back to the schools in order to download from school information systems the available administrative information for those same pupils who had been exposed to the program in the previous school year. Most of them had completed 7th grade and a few (2%) had repeated 6th grade. Some, however, had left the school. The information that could be easily retrieved is: teacher marks in all subjects, half-days of absenteeism and the good conduct mark. With this information, we define the same variables as in the main analysis to evaluate the impact of the program at the end of 7th grade (*i.e.* 18 months after the end of the program). As will be discussed below, this data can serve two objectives. One is to assess the existence of lasting impacts of the intervention; the other is to use the fact that pupils are grouped in different classes with different teachers to neutralize any possible reporting bias in favour of treated classes that may have affected estimated effects at the end of the 6th grade.

3.5. *Response rates*

The parent questionnaire was returned back within a week by approximately two-thirds of volunteer and non-volunteer families. On the subset of volunteer families, we made a special effort to improve the response rate: all volunteer families which did not return the questionnaire after a week were called to answer the same questions during a short phone interview. This has increased the response rate for volunteers to about 83% (Appendix Table A.3, Panel A). Comfortingly, non-response remains balanced for both volunteer and non-volunteer samples (Appendix Table A.3, Panel A column 6), so that attrition is not likely to introduce bias in our estimates.

Information on marks, sanctions, and absenteeism comes from school administration records. The detail in which it is available depends mostly on the specific software used by schools to record information. For example, at the end of 6th grade, information on disciplinary sanctions is recorded in 32 schools whereas information on the "good conduct" mark is recorded in 27 schools only (Appendix Table A.3, Panel B). Such variation in the availability of information across schools does not have the potential to introduce biases in estimation, as randomization was stratified by school.⁹ In contrast, within-school attrition may cause unbalance between treatment arms and generate biases. For absenteeism, sanctions, or marks, within-school attrition at the end of 6th grade is, however, very small (between 5% and 8%) and simply reflects that some registered pupils either never show up¹⁰ or change school during the academic year. For the "good conduct" outcome, a small fraction of reference teachers did not report the corresponding marks, which generates an additional source of (class level) attrition of about 8%. For all these outcomes,

9. It may nonetheless cause variation in the external validity of our estimates across outcomes. We have checked, however, that basic results remain qualitatively unchanged when we replicate our regressions on the core sample of the 25 schools for which information is jointly available on all basic outcomes, namely truancy, behavioural score, and average marks. This analysis is reported in Section A of the Appendix to this article available online (hereafter Web Appendix A).

10. Some students apply simultaneously to public and private schools and may be found in the registration files of a public school although they are actually attending a private school.

attrition rates are balanced across treatment arms (Appendix Table A.3, Panel B, column 6). Attrition for standardized tests conducted at the end of the 6th grade is somewhat larger than for administrative outcomes measured at the same date, due to teacher and pupil absences on the day of the test. However, the overall response rate is still above 80%, and balanced (Appendix Table A.3, Panel C).

We have performed the same analysis for administrative outcomes collected at the end of 7th grade (Appendix Table A.3, Panel D). These variables are subject to higher rates of within-school attrition because of school transfers and residential mobility between the end of 6th and the end of 7th grade.¹¹ Overall, attrition at the end of the 7th grade is comprised between 22% (absenteeism) and 30% (good conduct). It remains balanced across treatment arms.

Finally, the reference teacher questionnaire was filled and returned for more than 80% of pupils (Appendix Table A.3, Panel E). Most of the attrition is at the school and class level, reflecting the fact that some schools or some teachers forgot to (or objected to) answering this questionnaire. The individual attrition rate is small and similar to the rate for administrative measures, which indicates that, whenever returned, information on almost every pupil was provided. Again, non-response is balanced.

Overall, our different data sources provide information on a set of 17 basic outcomes and we do not find any significant variation in response rates across treatment arms. We have also checked for each outcome that the samples with non-missing values remain balanced on baseline observables. Specifically, Web Appendix B reports 306 treatment-control comparisons testing whether the 9 baseline characteristics in Appendix Table A.1 remain balanced on the 17 samples of volunteer and non-volunteer respondents. Comfortingly, only <2% of the tests (*i.e.* 5 out of 306) show a significant difference at the 5% level between treated and non-treated, among respondents.

We conclude from our analysis of non-response that attrition is ignorable and the baseline results presented in this article rely on this assumption. However, in Web Appendix C we also present bounds on treatment impacts under alternative assumptions about attrition. When no assumption is made on the relationships between treatment and response behaviour, we obtain worst-case bounds using recent estimators for partially identified linear models (Bontemps *et al.*, 2012). Generally speaking, these worst-case bounds are wide and uninformative about the sign of the effect of the program. In contrast, under the (less extreme) assumption that treatment affects response behaviour monotonically, we build on Lee (2009) to obtain bounds on outcomes that are much narrower and do not change signs (see Web Appendix C).

4. RESULTS

In this section, we analyse in turn the effect of the program on parental involvement and on pupils' outcomes. In most specifications, we estimate the following model on the volunteer and the non-volunteer populations separately:

$$Y_{ics} = \alpha T_c + X_i \beta + u_s + v_i$$

where, for each individual i in class c and school s , variable T_c is a dummy indicating whether class c is a treatment class, and X_i is a vector of control variables that includes dummies for gender, birth rank, high SES family, scholarship recipient, early grade repetition as well as controls for first-term marks in French and mathematics. Parameters u_s represent school effects (for which

11. Annual residential mobility is about 10% in France (Donzeau and Pan Ke Shon, 2009).

school dummies are included), whereas v_i represent unobserved individual random effects. The parameter of interest is α . Identification is a direct consequence of the experimental nature of the treatment assignment variable T_c , randomization having taken place after the information campaign was closed. The estimated α for non-volunteers can be interpreted as the effect of within-class spillover, whereas the estimated coefficient for volunteers encompasses direct effects of being invited to the meetings and possible feedbacks from treated and untreated peers, which cannot be non-parametrically disentangled.¹²

We assume that the outcomes in a given class are not influenced by the treatment status of pupils from other classes. This exclusion restriction is justified by the fact that the class being the constant structure of daylong school life in the French system, most peer interactions do take place within the class. Should also control classes be partly affected, we would most likely underestimate the true impacts.

To assess statistical significance, we provide robust standard errors allowing for correlated residuals within class. Web Appendix D also provides *adjusted p-values* for six families of estimates: 6th grade academic performance, 6th grade reported behaviour, and 7th grade outcomes, for volunteer and non-volunteers separately. Given that we observe several outcomes within each family, one possible concern is that some of the larger (absolute) *t*-statistics could have been observed by chance under the joint null assumption of no effects for all estimates in that family of parameters. To address this issue, we compute family-wise adjusted *p*-values, similar to Bonferroni corrections, but adjusted for the ordering of the tests and the covariance of the estimates. Reassuringly, tests based on family-wise adjusted *p*-values and on usual per-comparison *p*-values deliver similar conclusions about statistical significance.

4.1. *Increases in parental involvement*

The experimental evidence first suggests that the program was successful at significantly improving parental attitudes. Based on both parents' questionnaires and reference teachers' assessments, Table 2 reveals higher levels of parental involvement by volunteer parents in treatment classes, as well as a better perception and understanding of the school. Most estimated impacts on volunteer parents' are statistically significant at standard level and the 'global parenting score', averaging all questions in the parent questionnaire, increases by about 40% of a standard deviation.¹³ Appendix Table A.4 further describes the observed differences in parents' behaviour between treatment and control volunteers, across the 12 original dimensions measured by the questionnaire (from which the synthetic scores are computed). Parents' answers to such a questionnaire may suffer from desirability bias, but reference teachers' assessments confirm higher involvement by volunteer parents in treated classes, especially at home.

In contrast, having volunteer families in the same class does not affect the involvement of non-volunteer parents. When we restrict the analysis to non-volunteers, we find much smaller and equivocal differences between treatment and control groups. Estimated effects on non-volunteer parents are never significant at standard levels. This suggests that the effects of the principal's invitation do not spill over from volunteer to non-volunteer parents of treated classes. In contrast

12. Treatment on the treated parameters and social multipliers can be estimated based on stronger assumptions, as will be discussed later (see Section 4.3 and Web Appendix F).

13. Such differences are about the same order of magnitude as those that exist (independently of the treatment) between the 20% of families with higher SES and other families (see Web Appendix E). In theory, such a treatment-control gap could also be driven by a decline in parental involvement in the control group, even though it is unlikely that non-participation may have really disappointed volunteer parents, in the absence of explicit rewards associated with participation.

TABLE 2
Impact of the program on parental attitudes and behaviour

Dependent variable	Mean C	SD	T-C	(se)	n.obs.
Panel A: Volunteers					
Parent self-assessment					
Global parenting score	-0.010	0.344	0.137**	(0.023)	758
School-based involvement score	0.138	0.643	0.211**	(0.049)	757
Home-based involvement score	0.010	0.582	0.058*	(0.034)	758
Understanding and perceptions score	-0.100	0.538	0.134**	(0.038)	757
Reference teacher assessment					
Parent-school interaction	0.794	0.400	0.034	(0.029)	735
Parental monitoring of school work	0.206	0.420	0.067**	(0.034)	747
Panel B: Non-volunteers					
Parent self-assessment					
Global parenting score	0.003	0.344	0.011	(0.013)	1974
School-based involvement score	-0.048	0.643	0.018	(0.025)	1973
Home-based involvement score	-0.004	0.582	0.008	(0.022)	1973
Understanding and perceptions score	0.034	0.538	0.013	(0.022)	1967
Reference teacher assessment					
Parent-school interaction	0.806	0.400	-0.032	(0.021)	2449
Parental monitoring of school work	0.213	0.420	0.022	(0.022)	2464

Notes: Score variables are averages of normalized and centred answers to questions in the corresponding section of the parent questionnaire. The first column is the mean of the row variable in the control group for volunteers and non-volunteers, respectively; the second column is the standard deviation in the control group for the full population. Column T-C displays the coefficient from the regression of the row variable on a treatment class dummy, as well as controls for gender, birth rank, white collar, scholarship recipient, grade repetition, first-term marks in French and maths, and school-fixed effects. Each line corresponds to a separate regression. Robust standard errors allowing for correlated residuals within classes are in parenthesis. *, significant at 10% level; **, significant at 5% level.

to their children, parents of same class children have no specific reason to interact or even know each other.

In Web Appendix F, we further report estimates of the direct effects of actual *participation* into the program on the involvement of volunteer parents. We estimate these treatment-on-treated (TOT) effects using eligibility to the program as an instrumental variable for actual participation.¹⁴ These TOT parameters correspond to the intention to treat parameters divided by the regression-adjusted compliance rate. Our results suggest that participation into the program, as such, increases the global parenting score by 70% of an SD. Generally speaking, such TOT effects are of the same order of magnitude as those of some parent-focused early interventions where parents are trained to communicate with their young children and to teach skills to them (see, e.g., Kagitcibasi *et al.*, 2001). The Appendix also gives TOT effects for the other outcomes of the article, and orders of magnitude for pupils' outcomes, average marks, or outcomes observed after 18 months lie typically in the range of 15–30% of an SD. For outcomes that have spillover effects, the Appendix provides joint estimates of TOT and social multiplier, but their identification relies on the restrictive assumption of the linear-in-means model.

14. The identifying assumption is that eligibility to the program affects parents' outcome only through actual participation. This assumption is obviously debatable since being invited to participate in a program may, as such, affect outcomes. Our estimated TOT effects plausibly correspond to upper bounds for the effect of actual participation.

TABLE 3
Impact of the program on pupil's behaviour during Term 3

Dependent variable	Mean C	SD	T-C	(se)	n.obs.
Panel A: Volunteers					
Objective measure					
Absenteeism	4.173	7.524	-1.057*	(0.579)	726
Pedagogical team assessment					
Behavioural score	-0.028	0.738	0.153**	(0.038)	962
out of which					
Discipl. sanctions	0.110	0.300	-0.046**	(0.021)	917
Good conduct	0.290	0.483	0.065*	(0.033)	649
Honours	0.750	1.189	0.120**	(0.055)	923
Reference teacher assessment					
Behaviour in class	0.530	0.490	0.083**	(0.035)	750
School work	0.542	0.495	0.023	(0.030)	759
Panel B: Non-volunteers					
Objective measure					
Absenteeism	4.245	7.524	-0.539*	(0.294)	2421
Pedagogical team assessment					
Behavioural score	0.012	0.738	0.072**	(0.028)	3155
out of which:					
Discipl. sanctions	0.115	0.300	-0.024**	(0.012)	3014
Good conduct	0.348	0.483	0.046*	(0.025)	2190
Honours	0.794	1.189	0.061	(0.049)	2964
Reference teacher assessment					
Behaviour in class	0.591	0.490	0.022	(0.025)	2486
School work	0.574	0.495	0.011	(0.023)	2485

Notes: Absenteeism is counted in half-days; behavioural score is an average of normalized and centred dummies for sanctions, good conduct, and honours; all other variables are dummies. The first column is the mean of the row variable in the control group for volunteers and non-volunteers, respectively; the second column is the standard deviation in the control group for the full population. Column T-C displays the coefficient from the regression of the row variable on a treatment class dummy, as well as controls for gender, birth rank, white collar, scholarship recipient, grade repetition, first-term marks in French and maths, and school-fixed effects. Each line corresponds to a separate regression. Robust standard errors allowing for correlated residuals within classes are in parenthesis. *, significant at 10% level; **, significant at 5% level.

4.2. Improvements in pupils' behaviour and attitude

Turning to children, the measures taken at the end of the last school-term unanimously point to a better quality in children's relation to school in treatment than in control classes, across the complete range of available measures on behaviour and attitudes (Table 3). With respect to discipline, children of volunteer parents in treatment classes skip fewer classes (truancy is lower by 1.1 half-day, where the average is 4.2 half-days during the last term), are less likely to be punished for disciplinary reasons (6.4% against 11.0% in control classes), and are more likely to earn the top marks for their conduct (35.5% against 29.0%). Attitude in class also improves: honours are more often given by the class council and more often of the upper kind, and the reference teacher answers more often that the pupil is agreeable to deal within class (61.3% instead of 53.0%) and works diligently (56.5% against 54.2%), although this latter effect is not significant. Estimated improvements are perceptible at all levels of the distribution of behaviour: very bad behaviour is less frequent and very good one more frequent. Truancy, that was shown above to form a continuous, independent and objective measure, is also clearly affected.

Overall, our findings suggest that, as parents get more involved, children exhibit more positive attitudes and behaviour, with effects-sizes typically in the range of 10–20% of an SD. A simple school-based intervention can thus both improve parents' involvement in their children's education and foster changes in behavioural patterns.

4.3. *Peer-effects on behaviour and attitude*

Better behaviour happens to also influence non-treated children. The lower panel of Table 3 shows that children from non-volunteer families do behave better when they are in treatment classes. As a result of being in the same class, peers of treated children are less absent (by about 0.5 half-days), receive fewer disciplinary sanctions (the proportion of sanctioned children is 2.4 percentage points lower) and are more likely to get the top mark for their conduct (by 4.6 percentage points). All these effects on general discipline are only about half as large as for volunteers but remain statistically significant. Attitude in class as assessed by the pedagogical team or the reference teacher shows similar improvements of about one half the size of impacts on volunteer children, but none of those differences are statistically significant at standard levels.

As there is no evidence that untreated parents modified their behaviour, we exclude that spillovers originate from non-volunteer parents' influence on their own children. Effects on children of non-treated parents could rather result from progressive modification of the context of teaching and learning within the classroom, made possible by an improvement in the attitude of the volunteer share of the class. Significant spillovers on absenteeism or discipline also point to direct interactions between pupils, namely spillovers resulting from pupils imitating each other or taking action together.

To further test for the mechanisms of spillover effects, it is possible to compare classes with different shares of volunteers. The test relies on the idea that peer effects are likely to increase with increased level of interaction between volunteer and non-volunteer pupils. If interactions with peers are the explanation for observed spillovers on pupils' discipline, we should observe some dose–response relationship between the magnitude of these spillovers and the share of treated peers. As it turns out, this is what we find: when we focus on the four variables for which we obtain significant spillover effects on non-volunteer pupils, they are systematically larger when there are many volunteers in a treated class, even though the differential impact is never statistically significant at standard level (Table 4).¹⁵ 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. This exercise is, however, clearly suggestive of a dose–response relationship between the quantities of interactions with treated peers to which non-volunteers were exposed over the school year and the quality of their behaviour at the end of the year.¹⁶

As a complement to these reduced-form analyses, we provide (in Web Appendix F) estimates of the corresponding social multiplier (Glaeser *et al.*, 2003). To achieve identification of these more structural parameters, we build on simple linear-in-means models (Manski, 1993), where

15. In this analysis, “many volunteers” corresponds to a proportion of volunteers above the first tercile of the distribution of volunteers across classes (*i.e.* above 16% volunteers). Results are robust to change in this threshold.

16. If teachers' efforts or efficiency is sensitive to the number of volunteer children becoming involved in the program, such dose–response relationships may also be driven by teacher effects (as in Duflo *et al.*, 2011). In such a case, our estimated effects on non-volunteer pupils should be interpreted as an upper bound for the direct effects of mere peer interactions.

TABLE 4
Effects on non-volunteers: dose–response relationship

	Absenteeism	Behavioural score	Discipl. sanctions	Good conduct
T–C (NV): many vol’s	–0.746** (0.347)	0.118* (0.033)	–0.035** (0.016)	0.058* (0.030)
T–C (NV): few vol’s	–0.177 (0.662)	–0.003 (0.053)	–0.005 (0.016)	0.028 (0.049)
Many vol’s	0.515 (0.498)	–0.138** (0.047)	0.029 (0.020)	–0.020 (0.041)
n.obs.	2421	3155	3014	2190
p-values: differential effects many/few vol’s	0.464	0.059	0.192	0.609

Notes: This table presents an augmented version of regressions in Table 3 on the sample of non-volunteers only. Each column is a different regression. The treatment class variable is fully interacted with dummies for low (first tercile) and high (second and third tercile) proportion of volunteers in the class. Other variables (not reported) are controls for gender, birth rank, white collar, scholarship recipient, grade repetition, first-term marks in French and maths, and school-fixed effects. The model is thus $Y = aT * M + bT * (1-M) + cM + Xb + u$, where M is a dummy for high proportion class, T is a dummy for treatment, and X is the control variables. Robust standard errors allowing for correlated residuals within classes are in parenthesis. *, significant at 10% level; **, significant at 5% level.

individual outcomes are assumed to depend linearly on peer’s average outcomes. We identify the effect of peers’ average outcomes on the sample of non-volunteer pupils, using the treatment status of the class as an instrumental variable. These models provide point estimates of peer effects of about 0.68 for absenteeism and 0.82 for behavioural score (with s.e. of about 0.2), which imply, respectively, social multipliers of about 3 and 5. These estimates lie in the upper range of existing estimates of peer effects on students’ behaviour. Specifically, they tend to be larger than estimates obtained with non-experimental data (see, e.g., Gaviria and Raphael, 2001; Nakajima, 2007), but not significantly different from those obtained with experimental data (see, e.g., Duncan *et al.*, 2005; Bobonis and Finan, 2009). One must also bear in mind that there is an important margin of statistical uncertainty around those point estimates. As discussed by Glaeser *et al.* (2003), the values of the social multiplier have been found to vary a lot with the context. It is arguable that young adolescent behaviour is particularly sensitive to peers, especially in a schooling environment where pupils of the same class spend most of their time together.

4.4. Improvements in pupils’ cognitive achievement

Through its effects on pupils’ attitude and motivation, the program could be expected to extend its benefits to academic achievement measures. We have three sets of measures of achievement to test this claim: teacher marks, reference teachers’ assessment of progress, and standardized test scores. Teacher marks are essential in shaping pupils’ opportunities; they influence grade retention decisions, future high-school plans, and, in the medium term, the choice of optional subjects. One issue with this outcome, however, is that it may capture effort perception and not just learning levels. 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.

Using teacher marks, we do find a significantly higher achievement in French for pupils of the volunteer treatment group, relative to control pupils; the magnitude of the differential is about 12% of an SD (Table 5). We are not able to measure significant differences in mathematics, but the average of all subjects, weighted by class hours, shows an improvement of 8% of an SD. Confirming these findings, the reference teacher more often indicates progress for volunteer children in treatment classes, and the effect-size is similar (12% of an SD).

TABLE 5
Impact of the program on pupil's marks and test scores during Term 3

Dependent variable	Mean C	SD	T-C	(se)	n.obs.
Panel A: Volunteers					
Teacher marks					
French	10.736	3.717	0.456**	(0.192)	897
Maths	10.974	4.254	0.124	(0.221)	899
Average mark (all subjects)	11.586	2.878	0.240**	(0.118)	902
Reference teacher questionnaire					
Progress over the school year	0.548	0.488	0.057*	(0.033)	760
Uniform test					
French	-0.067	1.000	0.020	(0.060)	801
Maths	-0.077	1.004	-0.035	(0.064)	792
Panel B: Non-volunteers					
Teacher marks					
French	11.099	3.717	0.053	(0.144)	2938
Maths	11.087	4.254	0.047	(0.149)	2964
Average mark (all subjects)	11.717	2.878	0.032	(0.094)	2966
Reference teacher questionnaire					
Progress over the school year	0.612	0.488	0.00	(0.023)	2499
Uniform test					
French	-0.009	1.000	0.040	(0.043)	2614
Maths	0.006	1.004	-0.002	(0.046)	2607

Notes: Teacher marks are /20; average mark is an average of all subjects weighted by class hour; progress assessment is a dummy variable; uniform tests are normalized. The first column is the mean of the row variable in the control group for volunteers and non-volunteers, respectively; the second column is the standard deviation in the control group for the full population. Column T-C displays the coefficient from the regression of the row variable on a treatment class dummy, as well as controls for gender, birth rank, white collar, scholarship recipient, grade repetition, first-term marks in French and maths, and school-fixed effects. Each line corresponds to a separate regression. Robust standard errors allowing for correlated residuals within classes are in parenthesis. *, significant at 10% level; **, significant at 5% level.

In contrast, there are no effects in terms of test scores. As they provide objective and comparable measures, this would imply that neither parental involvement, nor pupils' behaviour and attitude did translate into higher cognitive abilities. This is not entirely surprising, because cognitive achievement should be less easily altered than attitudes or behaviour by an intervention targeting parents. But the fact that we find an effect on teachers' marks and no effect on test scores remains to be interpreted.

Test scores provide measures of cognitive achievement that are comparable across classes, but their limit is that pupils do not have any true incentive to succeed at these tests, as the results are not sent to parents and do not have any consequence for their future (this is especially true for end-of-the-year tests). In contrast, teachers' marks are sent to parents at the end of each term and represent an inherent part of the reward/sanction system in schools. As such, they depend not only on cognitive ability, but also very directly on pupils' motivation for achievement (see, *e.g.*, Borghans *et al.*, 2008). Web Appendix G provides direct evidence that teachers' marks tend to be significantly better for children who are less absent or receive fewer disciplinary sanctions, even after controlling for test scores. Assuming that absenteeism and sanctions reflect lack of motivation, these results are consistent with an interpretation of teacher marks as measuring both cognitive ability and motivation for achievement. All in all, the program seems to impact mostly behaviour and attitudes, and this is also apparent on teacher marks, to the extent that they reflect children's effort and motivations to succeed in school. Purely cognitive skills, at least when gauged with tests that come with no incentive to do well, are not affected.

Finally, turning to the non-volunteer group (lower panel of Table 5), there is no evidence of spillovers on any of the variables. For one thing, we could not expect spillover impacts on test scores when there is no direct impact on the volunteer population. For the other, spillovers on discipline and attitude in class are not strong enough to translate into significant spillover effects on teacher marks.

5. ROBUSTNESS ANALYSIS

The evidence presented in the previous section unanimously points to significant effects of the program across a wide range of outcomes. While absenteeism represents an objective, institutionally controlled and formalized measure, on which the program is consistently found to have both direct and indirect effects, many other outcomes are based on subjective assessment by teachers or by the school administration and it may be that such judgements were influenced by knowledge of pupils' treatment status. For example, if teachers want the program to be a success (although it does not provide additional resources), they may tend to better assess selected classes, regardless of pupils' true outcomes.

The existence of such Hawthorne effects, however, is less plausible than it seems. First, we have multiple view points and data sources on each kind of outcome, so that this sort of bias would need to affect everyone's judgement. Second, effects on both volunteer and non-volunteer pupils become perceptible and significant at the end of the school year only, after pupils in selected classes were exposed to repeated interaction with their treated parents and peers. Table 6, Panel A, compares treated and control samples for all measures that are available by the end of the first trimester. The end of Term 1 is in the middle of the sequence of meetings, and approximately 1 month after the assignment lottery took place. This is the moment where the experimental context of the program is most salient to teachers and school staff. There is nonetheless no detectable impact of the program on any of the outcomes at this point of time, regarding either discipline, attitude in class or teacher marks. This constitutes maybe the most direct set of evidence that teacher assessments are not influenced by class assignment status. Table 6, Panel B, shows similar outcomes at the end of Term 2: some of the impacts are already present among the volunteer group, but not all of them; and no impact is yet significant for the non-volunteer group. This is compatible with a progressive impact of parents on pupils and pupils on peers, whereas salience of the program to teachers should be decreasing with time.

Third, if knowledge about treatment status of the class did bias teachers' judgement, this knowledge would impact volunteer and non-volunteer pupils similarly, thus leading to systematic evidence of spillovers in all dimensions. This is not what we find. In particular, there are no significant spillovers on teachers' assessment of attitude in class or on marks, which are yet our most subjective outcomes. Peer effects are found mostly on objective outcomes (such as absenteeism) and this fact is not consistent with the reporting bias assumption. In fact, Hawthorne reporting effects provide a possible explanation for observed impact on teachers' assessments only insofar as teachers are able to identify not only treated classes, but also volunteer families within these treated classes. This could happen, but is very unlikely on a systematic basis: teachers were rarely involved in the meetings and, if so, only one or two of them were (in a school). Furthermore, such reporting bias should not form only after the first term.

A remaining possibility is that teachers form stronger empathy for children of parents that more actively seek teacher-parents interaction: the impact on parents would be real, but that on pupils would only reflect biased judgement in favour of involved families. However, teachers' own evaluation of the quality of parent-school interactions is not really improved when parents are invited to the program (Table 2), whereas their evaluation of parental monitoring of school

TABLE 6
Impact of the program on pupil's outcomes during Terms 1 and 2

Dependent variable	Mean C	SD	T-C	(se)	n.obs.
Panel A: Term 1					
Volunteers					
Absenteeism	1.090	2.761	-0.132	(0.151)	759
Discipl. sanctions	0.099	0.263	-0.030	(0.021)	874
Good conduct	0.370	0.493	0.016	(0.041)	690
Honours	0.956	1.187	-0.039	(0.074)	916
French mark	11.927	3.444	-0.363	(0.243)	907
Maths mark	12.252	3.931	-0.092	(0.253)	908
Average mark (all subjects)	12.472	2.626	-0.071	(0.165)	914
Non-volunteers					
Absenteeism	1.092	2.761	0.089	(0.113)	2712
Discipl. sanctions	0.080	0.263	-0.011	(0.011)	2915
Good conduct	0.412	0.493	0.024	(0.027)	2142
Honours	0.964	1.187	0.038	(0.038)	3034
French mark	11.963	3.444	-0.146	(0.156)	2962
Maths mark	12.541	3.931	0.089	(0.178)	2995
Average mark (all subjects)	12.584	2.626	-0.035	(0.094)	3008
Panel B: Term 2					
Volunteers					
Absenteeism	2.410	5.663	-0.125	(0.326)	753
Discipl. sanctions	0.139	0.279	-0.076**	(0.019)	899
Good conduct	0.364	0.486	-0.020	(0.041)	669
Honours	0.753	1.184	0.113*	(0.064)	916
French mark	11.207	3.635	0.102	(0.169)	943
Maths mark	11.564	4.007	0.165	(0.163)	910
Average mark (all subjects)	11.849	2.782	0.239**	(0.116)	948
Non-volunteers					
Absenteeism	2.746	5.663	-0.184	(0.205)	2573
Discipl. sanctions	0.088	0.279	-0.017	(0.012)	3029
Good conduct	0.362	0.486	0.033	(0.028)	2185
Honours	0.884	1.184	0.042	(0.045)	2924
French mark	11.268	3.635	0.017	(0.143)	3073
Maths mark	11.668	4.007	0.046	(0.146)	3003
Average mark (all subjects)	11.966	2.782	0.112	(0.099)	3100

Notes: This table presents variables identical to Tables 3 and 5, but *measured as of the end of the First and Second Terms*. The first column is the mean of the row variable in the control group for volunteers and non-volunteers, respectively; the second column is the standard deviation in the control group for the full population. Column T-C displays the coefficient from the regression of the row variable on a treatment class dummy, as well as controls for gender, birth rank, white collar, scholarship recipient, grade repetition, and school-fixed effects. Each line corresponds to a separate regression. Robust standard errors allowing for correlated residuals within classes are in parenthesis. *, significant at 10% level; **, significant at 5% level.

work (Table 2), pupil behaviour (Table 3) or progress (Table 5) is significantly stronger. There is thus no strong evidence that the increased school-based involvement reported by parents is actually perceived by teachers, to an extent likely to bias their judgement.

5.1. *Impact on first-borns*

Further evidence that differences between treated and untreated volunteers represent actual effects of the program derives from comparing first-born and other pupils. First-borns are the first children, in a family, to enter middle school. The school and its curriculum are new to their families and, as a consequence, parents of 6th graders who are the eldest child in the family

are likely to get more new information from the program than parents of other 6th graders. Furthermore, these parents do not have to share their involvement effort across several middle-school siblings, nor is parents' influence mitigated by that of older children in the same family. In fact a significant fraction of first-borns are lone children for whom parental influence is strong and unmitigated.

Overall, if the impact on volunteers' children reflects their repeated interactions with better-informed parents, we expect it to be stronger when the child is the only child in middle school. In contrast, even if the judgement of teachers or the administration was biased towards treated volunteers, it is not clear why it should vary according to pupils' birth rank. Therefore, stronger impacts on first-borns from volunteer families add strength to the evidence that the measured impacts are real and not the result of reporting bias. In the data, we can identify first-borns using information provided by families in the registration form before the start of the school year.

Table 7 presents the impact of treatment for the set of student outcome variables, where the treatment dummy is interacted with a first-born dummy. Very strikingly, for 11 outcomes out of 13, point estimates are larger when the child is a first-born. These larger effects do not hold for non-volunteers, but there is no reason why they should if, as we suggest, spillovers derive from peer interaction, not from parental influence. The difference between first-borns and other children among volunteers is most of the time marginally significant only, as we lack power for such a detailed analysis, but it is very systematic on several outcome variables that are very different in their scope and in their source. We take this result as compelling evidence that differences between treated and control children of volunteer parents are driven by the actual influence of better-informed parents rather than by reporting bias.

5.2. *Longer-term impacts*

Finally, we present impacts on treated children in June 2010, at the end of academic year 2009–2010, that is about 18 months after the end of the program. This is relevant per se, as it indicates whether the program has lasting effects. It also provides meaningful evidence on whether reported effects on treated pupils are robust to changes in class composition and pedagogical team.

As it turns out, the institutional practice in France is to reshuffle classes each year. Principals use this practice to influence future peer interactions and in particular to prevent dynamics of misbehaviour, by separating clusters of troublesome pupils across different classes. From 6th to 7th grade, choice of language options and other special programs can further imply reallocation. In our specific case, this practice implied that, with few exceptions, class-groups, as defined in the year of intervention, no longer exist as a unit of interactions in the second year. During 6th grade, non-volunteers in the treatment group were exposed to 23.9% of class-mates whose families had been invited to the meetings (the proportion in control classes is 0%) whereas, as a consequence of attrition and reshuffling, during 7th grade, the average non-volunteer from the treatment group had only 4.3% more invited classmates than the average control non-volunteer. Within this context, any effect which relies on peer exposure can no longer be detected with sufficient power. On the other hand, reshuffling makes it unlikely that teachers would be able to label pupils within mixed classes as belonging to the treatment group. Therefore, any impact that we can observe on former pupils from treated classes 1 year after treatment is not affected by systematic response behaviour.

Table 8 shows impact estimates for 7th grade variables. On the sample of volunteers, the effect of the program on French marks (and average marks) is almost as large and significant at the end of 7th grade as at the end of 6th grade, whereas the effect on "good conduct" is even slightly larger. The only outcome for which we obtain smaller (and less significant) effect

TABLE 7
Subgroup analysis: effects on first-borns and other children

Panel A: Behaviour									
	Absenteeism	Behavioural score	Discipl. sanctions	Good conduct	Honours	Behaviour in class	School work		
T-C (V): first-borns	-1.584** (0.737)	0.202** (0.050)	-0.069** (0.025)	0.072 (0.046)	0.163** (0.082)	0.149** (0.045)	0.102** (0.041)		
T-C (V): others	-0.481 (0.847)	0.095 (0.059)	-0.019 (0.030)	0.058 (0.047)	0.071 (0.086)	0.010 (0.050)	-0.065 (0.046)		
First-borns	0.082 (0.874)	-0.008 (0.064)	0.016 (0.029)	-0.001 (0.053)	0.038 (0.096)	-0.009 (0.046)	0.052 (0.047)		
n.obs.	726	962	917	649	923	750	759		
p-value: differential effects on first-borns/others	0.313	0.169	0.172	0.831	0.471	0.034	0.008		
Panel B: Marks and test scores									
		French mark	Maths mark	Avg. mark (all subjects)	Progress (ref.tech.ass.)	French test score	Maths test score		
T-C (V): first-borns		0.328 (0.233)	0.201 (0.260)	0.262* (0.144)	0.101** (0.043)	-0.053 (0.080)	0.006 (0.077)		
T-C (V): others		0.611** (0.253)	0.031 (0.295)	0.213 (0.166)	0.008 (0.045)	0.103 (0.081)	-0.080 (0.087)		
First-borns		0.525** (0.229)	0.057 (0.274)	0.141 (0.155)	0.066 (0.046)	0.189** (0.092)	-0.090 (0.081)		
n.obs.		897	899	902	760	801	792		
p-value: differential effects on first-borns/others		0.339	0.611	0.806	0.117	0.149	0.401		

Notes: This table presents an augmented version of regressions in Tables 3 and 5 on the sample of volunteers only. Each column is a different regression. The treatment variable is fully interacted with dummies for first-born within the family and a dummy for other children. Other variables (not reported) are controls for gender, birth rank, white collar, scholarship recipient, grade repetition, first-term marks in French and maths, and school-fixed effects. The model is thus $Y = aT + bF + cF + bT + dF + eT + u$, where F is a dummy for first-borns, T is a dummy for treatment, and X is the control variables. Robust standard errors allowing for correlated residuals within classes are in parenthesis. *, significant at 10% level; **, significant at 5% level.

TABLE 8
Longer-term effects (18 months after the intervention)

Dependent variable	Mean C	SD	T-C	(se)	n.obs.
Panel A: Volunteers					
Behaviour					
Absenteeism	4.456	8.085	-0.539	(0.493)	714
Good conduct	0.287	0.475	0.095**	(0.038)	612
Teacher marks					
French	10.775	3.859	0.340*	(0.176)	772
Maths	10.485	4.382	0.022	(0.226)	768
Average mark (all subjects)	11.343	2.958	0.193	(0.131)	774
Panel B: Non-volunteers					
Behaviour					
Absenteeism	4.649	8.085	0.175	(0.293)	2361
Good conduct	0.333	0.475	0.020	(0.022)	2083
Teacher marks					
French	10.706	3.859	-0.004	(0.151)	2552
Maths	10.418	4.382	0.158	(0.173)	2538
Average mark (all subjects)	11.379	2.958	0.015	(0.108)	2561

Notes: This table presents some of the variables in Tables 3 and 5 but *measured at the end of the year following the intervention*. The first column is the mean of the row variable in the control group for volunteers and non-volunteers, respectively; the second column is the standard deviation in the control group for the full population. Column T-C displays the coefficient from the regression of the row variable on a treatment class dummy (*as of the year of the intervention*), as well as controls for gender, birth rank, white collar, scholarship recipient, grade repetition, first-term marks in French and maths, and school-fixed effects. Each line corresponds to a separate regression. Robust standard errors allowing for correlated residuals within classes are in parenthesis. *, significant at 10% level; **, significant at 5% level.

is truancy. One possible explanation is that these reduced-form effects on 7th grade outcomes do not encompass peer effects anymore. Assuming that peer effects affect volunteers (during 6th grade) similarly as non-volunteers, the neutralization of these effects during 7th grade is able to explain most of the reduction in the overall effect of the program on truancy between 6th grade and 7th grade. As a matter of fact, Web Appendix F provides estimates of TOT effects on volunteers (*i.e.* effects net of spillovers), which turn out to be of similar magnitude at the end of the 7th grade as at the end of the 6th grade for marks, good conduct as well as truancy. These findings are quite in line with the hypothesis of persistent benefits from increased parental involvement. The persistence of such effects also confirms that teacher grades are not influenced by reporting biases.

6. EXTERNAL VALIDITY

Should this program be scaled up based on the results of this evaluation? Of course, we found very significant reduced form (or TOT) effects on the group of volunteer students, but these students are highly self-selected within each class and the subgroup of compliers in turn is self-selected within the volunteers. In such case, it is not clear whether we can extrapolate our reduced-form (or TOT) estimates to other groups of pupils or parents. In fact, given that treatment and randomization were defined at the class level, from a policy point of view, the relevant discussion of a generalization of the program should only be based on the effects of the program at that level; and on whether our school sample is representative of a larger set of schools.

With respect to the first question, Table 9 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

TABLE 9
Class-level analysis

	T-C	(se)	n.obs.
Parenting score	0.328**	(0.110)	171
Behavioural outcomes			
Absenteeism	-0.700**	(0.330)	142
Behavioural score	0.095**	(0.029)	183
Discipl. sanctions	-0.031**	(0.013)	175
Good conduct	0.040	(0.027)	130
Honours	0.089*	(0.046)	175
Behaviour in class	0.039	(0.027)	146
School work	0.015	(0.024)	147
Marks/test scores			
French mark	0.201	(0.160)	174
Maths mark	0.055	(0.168)	175
Average mark (all subjects)	0.092	(0.096)	175
French test score	0.029	(0.045)	165
Maths test score	-0.002	(0.049)	168

Notes: Column T-C displays the coefficient from the regression of the row variable's class level mean on a treatment class dummy. All regressions include school-fixed effects, as well as controls for average first-term marks in French and maths and for the class proportion of female, first-born, white collar, scholarship recipient, and grade repeating pupils. Mean parenting scores are normalized so that the distance between the first (P25) and third (P75) quartile of the distribution of class scores is set to 1. The remaining row variables are simple class level means of variables used in Tables 3 and 5. Each line corresponds to a separate regression. *, significant at 10% level; **, significant at 5% level.

or not. This analysis confirms that selecting a class into the program increases significantly the average involvement of parents and improves significantly the average behaviour of pupils. Specifically, we observe a reduction of about 20% in the average number of half-days of absenteeism in treated classes as well as a decrease of about 30% in the number of disciplinary sanctions and an increase of about 25% in the number of honours granted to pupils. Also, these class-level regressions confirm that the selection into the program generates improvements in average performance (as measured by teacher marks) that are positive, although not statistically significant at standard levels. Given the relatively small size of our sample of classes, we generally have sufficient precision at the class level only for outcomes that were hit by significant spillover effects.

Regarding the representativeness of our schools, we already mentioned that the 34 schools that volunteered to participate in the program represent about 10% of the district and most of them are located in “priority education” areas. Their results at the national exam¹⁷ taken at the end of middle school are actually very similar to those obtained on average by the schools classified as “priority education” in the district in 2008–2009 (p -value on a test for equal means: 0.304). The schools whose principals volunteered thus represent a significant share of what should be considered a primary target for such a program, and their academic results are actually representative of this target. Outside the *Créteil* district, in turn, “priority education” middle schools are not systematically different: the average passing rate at that same national exam was 72.5% in *Créteil* and 72.2% nationally. Our class-level estimates therefore provide a plausible prior for the average effect of the program should it be extended to all “priority education” schools in France that is, the 20% schools located in the most deprived areas in the country.

17. This exam (called the *Diplôme national du Brevet*) is externally set and marked.

With respect to the cost and benefit of such an extension, there is evidence that non-cognitive skills acquired during middle school have substantial impacts on subsequent educational achievement and labour market outcomes (Chetty *et al.*, 2011; Dee and West, 2011). Teacher and pupil well-being may also improve as a result of a better classroom climate. However, it is very difficult to place monetary values on such benefits. Our class-level analysis suggests nonetheless that the cost-effectiveness of a generalization of the program should be high, if only because the corresponding costs are very low. Because the district had to define guidelines and produce a DVD on middle-school issues, the program involved some fixed costs. But given that this investment has been made, the marginal cost of a generalization is estimated at only about EUR 1000 *per school* and per year.¹⁸

7. CONCLUSION

Research in the economics of education has put much emphasis on the role of school-level inputs, such as class size or teacher training, in the human capital production technology. Although family background has long been considered a strong determinant of schooling decisions and academic success, there is still limited evidence on whether parents' involvement in their children's education can be influenced by policy. This article shows that parents can be a significant input: when they receive invitations and support from the school, their involvement increases, and pupil's behaviour *at school* improves. Impacts are mostly on behavioural outcomes: they are difficult to map to specific monetary returns, but there is growing evidence that these are important determinants of long-term social and economic outcomes. Furthermore, compared to other school interventions, this program is remarkably low cost: this suggests that it may have higher rates of returns than most other interventions (such as class size reduction or teacher training), at least when implemented within the 10–20% low background schools which are the typical target of such initiatives.

The literature on early childhood interventions provides complementary evidence on how parental attitudes can be affected by training and home-visits, with demonstrated impacts on children cognitive and non-cognitive development. Given that cognition and behaviour are malleable at early ages, those interventions are usually considered efficient policies. According to our results, such findings generalize to pre-adolescent children, and parents are not only an input for at-home production of human capital, but also for the schooling technology itself. Importantly, the fact that mechanisms are elicited within the school environment gives rise to peer-pressure that shapes pupil's behaviour. As a result, we can define 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 efforts by some parents.

Overall, our results show that low levels of parental involvement are not a fatality in poor neighbourhoods. Schools have the critical ability to trigger higher levels of involvement among parents. It is likely, however, that some parents did not participate in the program even though they would have benefited from it. The net benefits of the program could probably be further increased if more parents could take it up. Further research is needed to better understand how the take-up of school-based interventions could be increased and the actual cost-effectiveness of improved take-up.

18. They correspond to overtime payment for the facilitator and a budget for communication with parents (*Ministère de l'éducation nationale*, Circulaire n°2010-106 du, July 15, 2010). For the sake of comparison, Dee and West (2011) find that it takes a 1 SD reduction in early class size to obtain a +10% of a SD improvement in non-cognitive skills during middle school, implying a cost of about 3400 US\$ *per student* and per year.

APPENDIX

TABLE A1
Differences between volunteer and non-volunteer characteristics

	Mean NV	V–NV	(se)	n.obs
Parent				
Employment status	0.86	–0.004	(0.013)	4276
Two-parents household	0.75	0.048**	(0.017)	4138
White collar	0.18	0.028*	(0.015)	4308
Scholarship recipient	0.35	0.022	(0.020)	3852
Children				
Girl	0.49	–0.036*	(0.019)	4308
First-born	0.54	0.002	(0.018)	4308
Past grade repetition	0.26	–0.021	(0.015)	4308
French test score (Sept. 08)	0.04	–0.073*	(0.041)	3820
Maths test score (Sept. 08)	0.02	–0.017	(0.044)	3831

Notes: All variables except test scores are dummies measured before the beginning of the year; French and maths test scores are uniform standardized test submitted at the beginning of the year. The first column is the mean of the row variable in the non-volunteer group. Column V–NV displays the coefficient from the regression of the row variable on a volunteer dummy, and school-fixed effects. Each line corresponds to a separate regression. Robust standard errors allowing for correlated residuals within classes are in parenthesis. *, significant at 10% level; **, significant at 5% level.

TABLE A2
Differences in pre-treatment characteristics

	Mean C	T–C	(se)	n.obs
Panel A: Volunteers				
Parents				
Employment status	0.85	–0.020	(0.021)	970
Two-parents household	0.80	–0.025	(0.025)	928
White collar	0.21	–0.015	(0.024)	970
Scholarship recipient	0.39	–0.043	(0.028)	863
Children				
Girl	0.46	–0.006	(0.030)	970
First-born	0.54	0.016	(0.028)	970
Past grade repetition	0.25	0.026	(0.028)	970
French test score (Sept. 08)	–0.07	–0.012	(0.079)	903
Maths test score (Sept. 08)	–0.04	–0.057	(0.074)	914
Panel B: Non-volunteers				
Parents				
Employment status	0.86	0.005	(0.012)	3306
Two-parents household	0.75	0.010	(0.015)	3210
White collar	0.18	0.006	(0.013)	3338
Scholarship recipient	0.35	0.024	(0.017)	2989
Children				
Girl	0.49	0.007	(0.012)	3338
First-born	0.54	–0.025	(0.018)	3338
Past grade repetition	0.26	–0.014	(0.017)	3338
French test score (Sept. 08)	0.04	–0.041	(0.049)	2917
Maths test score (Sept. 08)	0.02	0.032	(0.042)	2917

Notes: All variables except test scores are dummies measured before the beginning of the year; French and maths test scores are uniform standardized test submitted at the beginning of the year. The first column is the mean of the row variable in the control group. Column T–C displays the coefficient from the regression of the row variable on a treatment class dummy, and school-fixed effects. Each line corresponds to a separate regression. Robust standard errors allowing for correlated residuals within classes are in parenthesis. *, significant at 10% level; **, significant at 5% level.

TABLE A3
Response rates analysis

	Respondents		Within-school attrition			Balancing test	
	Individuals	Schools	Total	Class level	Individual	T-C	(se)
Panel A: Parent questionnaire							
Non-volunteers	1974	31	0.36	0.06	0.30	-0.029	(0.028)
Volunteers	627	31	0.31	0.08	0.23	-0.022	(0.037)
Volunteers (incl. call-back)	758	31	0.17	0.00	0.16	-0.023	(0.028)
Panel B: Behaviour administrative data							
Absenteeism	3147	26	0.05	0.00	0.05	-0.002	(0.006)
Behavioural score	4117	34	0.04	0.00	0.04	-0.001	(0.006)
Discipl. sanctions	3931	32	0.04	0.00	0.04	-0.003	(0.006)
Good conduct	2839	27	0.15	0.08	0.07	0.014	(0.023)
Honours	3887	33	0.07	0.01	0.06	-0.022	(0.014)
Panel C: Marks and tests scores							
French mark	3835	33	0.08	0.02	0.07	0.015	(0.017)
Maths mark	3863	33	0.08	0.01	0.07	0.001	(0.013)
Average mark (all subjects)	3868	33	0.08	0.01	0.07	0.002	(0.015)
French test score	3415	32	0.17	0.05	0.12	-0.020	(0.024)
Maths test score	3399	32	0.17	0.03	0.14	-0.011	(0.017)
Panel D: 7th grade measures							
Behaviour							
Absenteeism	3075	31	0.22	0.00	0.22	-0.003	(0.012)
Good conduct	2695	31	0.30	0.02	0.29	0.009	(0.021)
Teacher marks							
French	3324	34	0.23	0.00	0.23	-0.000	(0.012)
Maths	3306	34	0.23	0.00	0.23	0.005	(0.013)
Average mark (all subjects)	3335	34	0.23	0.00	0.23	0.000	(0.012)
Panel E: Reference teacher questionnaire							
Behaviour in class	3236	30	0.17	0.11	0.06	-0.039	(0.034)
School work	3244	30	0.17	0.10	0.06	-0.027	(0.033)
Progress	3259	30	0.17	0.10	0.06	-0.027	(0.033)
Parent-school interaction	3184	30	0.18	0.10	0.08	-0.008	(0.035)
Parental monitoring of school work	3211	30	0.18	0.11	0.07	-0.018	(0.032)

Notes: This table presents attrition rates of our different sources of data. Column 1 gives the number of individual observations available. Column 2 gives the number of schools for which we have some information. Column 3 gives the within-school attrition rate and columns 4 and 5 decompose this rate into the attrition that result from whole classes missing and the residual within-class individual attrition. Column 6 (T-C) displays the coefficient from the regression of a response dummy on a treatment class dummy within schools that are not entirely missing. Each line corresponds to a separate regression. Robust standard errors allowing for correlated residuals within classes are in parenthesis. *, significant at 10% level; **, significant at 5% level.

TABLE A4
Differences in parents' attitude and behaviour between treatment and control volunteers (raw indicators)

Question	Mean C	T-C	(se)
Global parenting score	-0.010	0.137**	(0.023)
School-based involvement score	0.138	0.211**	(0.049)
Several individual appointments with teachers	0.23	0.058*	(0.031)
Has attended parents/teachers meetings	0.80	0.076**	(0.026)
Has participated in parent' organizations	0.25	0.116**	(0.033)

(continued)

TABLE A4
Continued

Question	Mean C	T-C	(se)
Home-based involvement score	0.010	0.058*	(0.034)
Precise knowledge of child's grades	0.44	0.040	(0.034)
Sometimes helps with homeworks	0.89	-0.006	(0.022)
Child does not watch TV daily after 9 PM	0.81	0.036	(0.025)
Child spends <1 h/d on other screens	0.88	0.028	(0.020)
Understanding and perceptions score	-0.100	0.134**	(0.038)
Knowledge of optional courses offered	0.76	-0.101**	(0.028)
Has never been anxious about violence	0.26	0.015	(0.031)
Clear ideas about high-school plans	0.26	0.044	(0.033)
Satisfied with school	0.81	0.060**	(0.021)
Never been summoned to the school	0.72	0.097**	(0.029)

Notes: Score variables are averages of normalized and centred answers to questions in the corresponding section of the parent questionnaire; other variables are dummies from the raw questions asked in the parent questionnaire. The first column is the mean of the row variable in the control group. Column T-C displays the coefficient from the regression of the row variable on a treatment class dummy, as well as controls for gender, birth rank, white collar, scholarship recipient, grade repetition, first-term marks in French and maths, and school-fixed effects. Each line corresponds to a separate regression. Robust standard errors allowing for correlated residuals within classes are in parenthesis. *, significant at 10% level; **, significant at 5% level.

Acknowledgments. This research was supported by a grant from the French Experimental Fund for the Youth. We are very grateful for the support of the schools and administrative teams from the *rectorat de Créteil*, and particularly to Bénédicte Robert. We thank the many J-Pal Europe research assistants that worked on this project. We are also very grateful to the Editor and three referees for their comments, as well as to seminar participants at LSE (London), TSE (Toulouse), PSE (Paris), GREQAM (Marseille), Science Po (Paris), IZA, the University of Mannheim, SOLE/EALE conference (London), ESPE conference (Essen), and EEA congress (Glasgow). The analyses and the opinions in this article are those of the authors and do not necessarily reflect the views of the OECD and of its members.

Supplementary Data

Supplementary data are available at *Review of Economic Studies* online.

REFERENCES

- AIZER, A. (2004), "Home Alone: Supervision after School and Child Behavior", *Journal of Public Economics*, **88**, 1835-1848.
- ANGRIST, J. and LANG, K. (2004), "Does School Integration Generate Peer Effects? Evidence from Boston's Metco Program", *American Economic Review*, **94**, 1613-1634.
- AVVISATI, F., BESBAS, B. and GUYON, N. (2010), "Parental Involvement in Schools: A Literature Review", *Revue d'Economie Politique*, **120**, 761-778.
- BOBONIS, G. J. and FINAN, F. (2009), "Neighborhood Peer Effects in Secondary School Enrollment Decisions", *The Review of Economics and Statistics*, **91**, 695-716.
- BONTEMPS, C., MAGNAC, T. and MAURIN, E. (2012), "Set Identified Linear Models", *Econometrica*, **80**, 1129-1155.
- BORGHANS, L., DUCKWORTH, A. L., HECKMAN, J. J. and TER WEEL, B. (2008), "The Economics and Psychology of Personality Traits", *Journal of Human Resources*, **43**, 972-1059.
- BRUNELLO, G. and SCHLOTTER, M. (2011), "Non Cognitive Skills and Personality Traits: Labour Market Relevance and their Development in Education & Training Systems", IZA DP 5743.
- CARRELL, S. E. and HOEKSTRA, M. L. (2010), "Externalities in the Classroom: How Children Exposed to Domestic Violence Affect Everyone's Kids", *American Economic Journal: Applied Economics*, **2**, 211-228.
- CHETTY, R., FRIEDMAN, J. N., HILGER, N., SAEZ, E., WHITMORE SCHANZENBACH, D. and YAGAN, D. (2011), "How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project Star" (Working Paper No. 16381, NBER).
- CUNHA, F. and HECKMAN, J. J. (2008), "Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation", *Journal of Human Resources*, **XLIII**, 739-782.
- DEE, T. S. and WEST, M. R. (2011), "The Non-Cognitive Returns to Class Size", *Educational Evaluation and Policy Analysis*, **33**, 23-46.

- DESFORGES, C. and ABOUCHAAR, A. (2003), "The Impact of Parental Involvement, Parental Support and Family Education on Pupil Achievements and Adjustment: A Literature Review" (Research Report No. 433, Department for Education and Skills).
- DONZEAU, N. and PAN KE SHON, J. L. (2009), "Residential Mobility Trends in France, 1973-2006: New Estimates", *Population*, **64**, 687-703.
- DUFLO, E., DUPAS, P. and KREMER, M. (2011), "Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya", *American Economic Review*, **101**, 1739-1774.
- DUNCAN G. J., BOISJOLY, J., KREMER, M., LEVY, D. M. and ECCLES, J. (2005), "Peer Effects in Drug Use and Sex among College Students", *Journal of Abnormal Child Psychology*, **33**, 375-385.
- GAVIRIA, A. and RAPHAEL, S. (2001), "School-Based Peer Effects and Juvenile Behavior", *The Review of Economics and Statistics*, **83**, 257-268.
- GLAESER, E. L., SACERDOTE, B. I. and SCHEINKMAN, J. A. (2003), "The Social Multiplier", *Journal of the European Economic Association*, **1**, 345-353.
- GRANTHAM-MCGREGOR, S., POWELL, C., WALKER, S., CHANG, S. and FLETCHER, P. (1994), "The Long Term Follow-Up of Severely Malnourished Children Who Participated in an Intervention Program", *Child Development*, **65**, 428-439.
- HECKMAN, J. J., STIXRUD, J. and URZUA, S. (2006), "The Effects of Cognitive and Non Cognitive Abilities on Labor Market Outcomes and Social Behaviors", *Journal of Labor Economics*, **24**, 411-482.
- HECKMAN, J. J., PINTO, R., SHAIKH A. and YAVITZ, A. (2011), "Inference with Imperfect Randomization: The Case of the Perry Preschool Program" (IZA Discussion Papers 5625, Institute for the Study of Labor (IZA)).
- HILL, N. and TYSON, D. (2009), "Parental Involvement in Middle School: A Meta-Analytic Assessment of the Strategies that Promote Achievement", *Development Psychology*, **45**, 740-763.
- HOOVER-DEMPSEY, K. V. and SANDLER H. M. (1995), "Parental Involvement in Children's Education: Why Does it Make a Difference?", *Teachers College Record*, **97**, 310-331.
- KAGITCIBASI, C., SUNAR, D. and BECKMAN, S. (2001) "Long-Term Effects of Early Intervention: Turkish Low-Income Mothers and Children", *Applied Developmental Psychology*, **22**, 333-361.
- KNUDSEN, E. I., HECKMAN, J. J., CAMERON, J. L. and SHONKOFF, J. P. (2006), "Economic, Neurobiological, and Behavioral Perspectives on Building America's Future Workforce", *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 10155-10162.
- LEE, D. S. (2009), "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects", *Review of Economic Studies*, **76**, 1071-1102.
- MANSKI, C. F. (1993), "Identification of Endogenous Social Effects: The Reflection Problem", *The Review of Economic Studies*, **60**, 531-542.
- MOFFITT, R. A. (2001), "Policy Interventions, Low-Level Equilibria, and Social Interactions", in Durlauf, S. N. and Young, H. P. (eds) *Social Dynamics* (Washington, D.C.: The MIT Press, for Brookings Institution) 45-82.
- NAKAJIMA, R. (2007), "Measuring Peer Effects on Youth Smoking Behavior", *The Review of Economic Studies*, **74**, 897-935.
- OLDS, D. L. (2002), "Prenatal and Infancy Home Visiting by Nurses : From Randomized Trials to Community Replication", *Prevention Science*, **3**, 153-172.
- SIRIN, S. (2005), "Socioeconomic Status and Academic Achievement: A Meta-Analytic Review of the Research", *Review of Educational Research*, **75**, 417-453.
- TODD, P. E. and WOLPIN, K. E. (2007), "The Production of Cognitive Achievement in Children: Home, School, and Racial Test Score Gaps", *Journal of Human Capital*, **1**, 91-136.
- WELSCH, D. M. and ZIMMER, D. M. (2008), "After-School Supervision and Children's Cognitive Achievement", *The B.E. Journal of Economic Analysis & Policy*, **8**, art. 49.