The Long-Term Impact of Social Skills Training at School Entry: A randomized controlled trial[1]

Yann Algan (Sciences Po), Elizabeth Beasley (Sciences Po),

Frank Vitaro (Univ Montréal, Grip), Richard E. Tremblay (UCD, Univ Montréal, Grip)

July 2013

Abstract

Early childhood social skills are strong predictors of adult socio-economic success, but little is known about how to improve these social skills, in particular among children the most at risk of engaging in criminal behavior or failing at self-sufficiency as adults. We use data from an intensive social skills training randomized program for children with long term follow-up to answer this question. We match detailed data on behavior from childhood and adolescence with administrative criminal and educational records and self-reported socio-economic outcomes. As adults, the subjects in the treated group are less likely to have a criminal record (an 11 percentage point reduction from an average of 33%), more likely to have a secondary diploma (a 19 percentage point increase from an average of 32%), have significantly stronger labor market performance than the non-treated group, with an increase in the likelihood of employment at age 26 of 10 percentage points, and higher social capital and community involvement. We distinguish the different channels through which this intervention operates, and find that among the non-cognitive channels, an increase in social trust explains a significant share of lower adult criminality and improved economic outcomes, and increased self-control explains a significant share of increased secondary completion and reduced criminality. Using conservative assumptions in a simple framework, we estimate overall rate of return of this program in terms of expected lifecycle income ranges from 282% to 452%, implying that every $1 invested yields $2.8 to $4.5 in benefits.

---

[1] We thank for very helpful comments Sam Bowles, Ernst Fehr, Nicole Fortin, Jens Ludwig, Philip Oreopoulos and Emmanuel Saez. All the data from the MLES experiment are located at GRIP: Groupe de recherche sur l'inadaptation psychosociale chez l'enfant, Montreal University, Montréal Canada. The research leading to these results has received funding from the European Research Council for Yann Algan under the European Community's Seventh Framework Program (FP7/2007-2013) / ERC grant agreement n° 240923. Elizabeth Beasley gratefully acknowledges the support of the Abdul Latif Jameel Poverty Action Lab (J-PAL). The experimentation was funded by a research grant to Richard E. Tremblay from the Canadian Institutes of Health Research. The longitudinal follow-up of the study was funded by the following Canadian research agencies : Social Sciences and Humanities Research Council of Canada, Canadian Institutes for Health Research, Québec Research Fund for Society and Culture (FRQSC) and Québec Research Fund for Health (FRQS).

1. Introduction

Is it possible that counting to ten when angry at age seven can increase someone's chances of getting a job at age 30 by 10%? Or that role-playing the disappointment of being tagged out in grade school can reduce the chance that a child will go to prison when he grows up by 30%? More generally, can public policy improve outcomes for adults through interventions targeted at social skill development in childhood, especially among the most disruptive children?

This question is important for education policy in general, and it is critical for preschool children with low levels of social skills who are likely to face serious social adjustment difficulties as adults and be locked into poverty. Substantial evidence shows that non-cognitive skills, such as impulse control, social-information processing or social cooperation are strongly correlated with favorable economic and social outcomes, and that formal education is important not only for training in cognitive skills (for example, how to do long division) but also for the development of these non-cognitive skills (Jencks, 1979, Heckman and Rubinstein, 2001, Bowles et al. 2001, Segal, 2013, Dewey 1944, Algan, Cahuc and Shleifer 2013). Recent assessments of early childhood experiments aimed at improving cognitive skills, such as the Abecedarian project, the Perry Preschool program, Head Start or project Star, also suggest that their long-term impact flows through some sort of non-academic or non-cognitive skills channels (see Almond and Currie 2010 for a review). This raises the question of what those non-cognitive skills are and how much of the impact comes through them. It also raises the question on how policy intervention centered on social skills trainings could remediate deficits in non-cognitive skills, particularly for those children who have the most need and are considered "high risk" of anti-social behavior.

This paper contributes to this important literature by estimating the long-run impact of the Montreal Longitudinal Experimental Study (MLES), a randomized experiment of social skills training for children and improvement of parenting skills targeted at disruptive behavior. In the spring of 1984, the MLES evaluated 1037 boys at the end of kindergarten in schools in low socio-economic areas. From this original sample, 250 boys were targeted for the experiment based on teacher ratings of disruptive behavior. These 250 boys were randomly assigned to either participate in the social skills/parent training program or to be part of the control group.

The program lasted two years, when the subjects were between seven and nine years old. The training program was intensive and involved different points of contact: the child

himself, his parents, his peers and his teachers. Over the two-year period 19 social skills sessions were given by a team of professional childcare workers, a social worker and a psychologist. The sessions included disruptive boys and boys with very good social skills to facilitate learning of the social skills by the disruptive boys. Sessions included social behavior/skills training (such as how to ask why, how to invite a bystander to play, or how to help) and self-control strategy lessons (like how to react to teasing or how to react when angry, how to correctly identify the intentions of others). The parents received a home-based training program in child rearing to promote the social skills of their children. Parents were taught to observe and recognize their child's problem behaviors, to define clear guidelines for their children respond in a way that favored their children's acquisition of social skills, and to supervise the homework and relationships outside home of their child. The control group did not have access to this program but had access to all of the standard programs and resources available to the Montreal public school children in this period.

Over the last three decades detailed longitudinal data were collected on the MLES children concerning their non-cognitive and cognitive skills, and later adult outcomes. The outcomes include whether or not the subject received a high-school diploma from administrative data, criminal record at age 24 from administrative data, self-reported economic outcomes at ages 17-26 years, including labor market history, employment status and income, as well as self-reported social outcomes related to social capital such as group membership and civic activities.

This paper analyzes the long-term impact of this early-childhood social skills intervention on school achievement, adult criminality, and socio-economic outcomes at ages 17-26 years. We find that the treated group reports a much higher rate of employment (10 percentage points more likely to be employed than the control group) and more favorable economic outcomes. Administrative records also document a higher rate of secondary-school graduation for the treatment group (19 percentage points higher than the control group average of 32%), as well as a large reduction in the probability of having a criminal record (11 percentage points lower than the control group average of 33%). The treatment group also displays higher level of social capital, in particular group membership (such as civic or sports groups).

We then evaluate the channels producing the long-term economic impact of early childhood social skills training. We look at the relationship between treatment and a large set of intermediate variables concerning cognitive and non-cognitive skills that the MLES

collected on a yearly basis at ages 10-17 years. We exploit the comprehensive set of direct measures for non-cognitive skills collected by the MLES to distinguish two channels related to social skills (Social Trust and Altruism), two channels related to school (School Performance and School Attitude), two channels related to conduct problems (Delinquent Behavior and Hyperactivity) and one channel related to self-image (Self Esteem). We find that the program had significant impacts on Social Trust, Delinquent Behavior, and Hyperactivity. The last part of our analysis seeks to identify which of the intermediate variables are likely to be functioning as channels for treatment impact, following the seminal methodology proposed by Heckman et al. (2012).

We find that Social Trust is important for reducing criminality, improving economic outcomes, and increasing social capital, while it does not seem to play an important role in obtaining a secondary school diploma (once in-school performance is accounted for). Reducing hyperactivity and delinquent behavior is important for reducing criminality and increasing the chance of obtaining secondary school diplomas. Improved in-school performance is important for explaining reduced criminality, increased secondary diplomas, and improved economic outcomes, and we argue that at least some part of in-school performance is a proxy for non-cognitive skills that we are unable to measure.

Our study contributes to the growing literature on the impact of education and early childhood development programs on adult outcomes. Substantial evidence show that early childhood educational intervention, like the Abecedarian (Campbell et al. 2002) and Perry Programs (Heckman et al., 2010), the Head Start program (Currie and Duncan, 1995; Ludwig and Phillips 2007) or the Project STAR (Krueger 1999, Chetty et al. 2011) leads to a variety of favorable long-term outcomes. Those outcomes range from improved earnings and employment prospects, increase in educational attainment to large drop in criminal rates or mortality rates.

The MLES project differs in three main dimensions from other experiments. First, this intervention is specifically targeted at children with very high deficit in social skills while the selection criterion of previous studies was mostly based on low IQ and low-income status of parents. There is a particular need for research on programs that focus on high-risk children (i.e., the "tail" of the distribution of behavior and ability), for equity reasons regarding the wellbeing of the worst off, but also for more practical reasons. Individuals at the end of the spectrum of anti-social behavior during their childhood are more likely to incur costs for society than those towards the middle of the spectrum. For example, untreated individuals in

our sample, who were above the 70th percentile of the MLES 1037 children sample in terms of disruptiveness, spent twice as many years receiving public transfers as the bulk of the larger sample (7% vs. 14%) and were more than twice as likely to have a criminal record (33% vs. 15%).

Second, the MLES project is centered exclusively on social skill trainings with a comprehensive set of interventions targeting children, peer interactions, parents and teachers, while the other early childhood interventions were mainly based on additional academic support (such as extra teaching and quantity of schooling, new teaching methods, class size, teacher quality). Although elements of new pedagogy were introduced[2], the non-cognitive skills were not the only primary targets of those interventions and the boost in those skills came initially as a surprise. The experiment that is closest to ours is analyzed by Heller et al. (2012). The authors look at a randomized field experiment targeted at the development of social-cognitive skills of pre-adolescents in Chicago Public Schools and find a significant effect of this intervention on schooling outcomes and violent crime arrests during the year of the intervention. This intervention is remarkable by the size of the sample (around 3000 children in grade 5-7th) and the social skills they look at (impulse control, emotional regulation, social information processing). However, Heller et al (2012) estimate the impact after one year while the MLES follows an entire cohort of preschool children and allows to estimate the long-term impact of social skills training during early elementary school on a comprehensive set of adult outcomes.

Third, the MLES provides comprehensive measures of non-cognitive skills during the entire childhood and adolescence period, while the evaluation of the experimental programs described above lacked detailed assessments of non-cognitive skills, taking them as a "black box" for any non-academic behaviors rewarded by the labor market (see the discussion by Bowles 2013). In particular, previous studies focused on specific proxies such as academic motivation and externalizing behavior (Chetty et al. 2011, Heckman et al. 2012). We show

[2] The Perry Preschool program mainly consisted of additional preschool sessions and home visits by teachers for children at ages 3- 4 years. The curriculum also included some non-cognitive tasks with a plan-to-review sequence to develop a sense of responsibility and independence (see Schweinhart, Barnes and Weikart, 1983). The Abecedarian program was an intensive educational daycare program which started in infancy and lasted until age 5 years. The curriculum was largely based on educational games designed during infancy to promote cognitive, language, perceptual-motor, and social development (Sparling & Lewis, 1979). When children reached the preschool years, the curriculum was more focused on language development and pre-literacy skills (Sparling & Lewis, 1984). African-Americans were the main participants in both programs.

that the access to a variety of non-cognitive measures is key to identify the channels of impact of an intervention on different outcomes. For example, we are able to differentiate social trust from altruism, gregarious behavior, delinquent behavior, school attitude, and self-esteem to show that it is in fact the ability to connect with and trust other people that is the mechanism for improvements in economic outcomes and adult social capital (group membership), rather than, for example, improved networks (which would be more likely to be captured by gregariousness) (see Banfield 1956 and Coleman 1990 for seminal discussion about the various dimensions of social capital). We are also able to show that increased self-control (reduction of hyperactivity) is important for reducing criminality and improving educational achievement. Similarly, some part of the delinquent behavior (also referred to as externalizing behavior in the literature) that leads to poor adult outcomes is due to low levels of self-control. Outcomes may thus be improved by interventions that specifically focus on specific non-cognitive skills.

Our paper also relates to recent studies on the MLES project. Boisjoli et al. (2007), Tremblay et al. (1992, 1995, 2000) and Vitaro et al. (2001) have extensively analyzed the MLES effect on the development of aggressive behavior during childhood and on school achievement and crime during adolescence. Our paper adds to these evaluations by evaluating the long-term impact of this comprehensive intervention on various socio-economic adult outcomes: employment and income, social capital, education and crime. We also provide a framework to decompose the mediating channels and estimate the interactions between those channels.

The paper is organized as follows. In section 2, we review the MLES design and address the potential issues surrounding the validity of the experiment. Section 3 estimates the adult economic and social impacts of the experiment. In section 4 we analyze the effect of the program on intermediate cognitive and non-cognitive skills during childhood. Section 5 decomposes the channels through which the program affects adult outcomes. Section 6 estimates the rate of return of the MLES experiment. Section 7 provides robustness checks, in particular with respect to attrition. Section 8 concludes.

2. Experimental design and Data

The Montreal Longitudinal Experimental Study (MLES) began in 1984 and piloted a prevention program for disruptive children with three components: social skills training in a

small group format involving a majority of pro-social peers, family visits, and teacher management skills (see Tremblay et al. 1992 for a detailed description of the experiment).

2.1 Selection

Kindergarten teachers in 53 schools in Montreal, Canada, were asked to rate the behavior of their male students at the end of the 1984 school year. These schools were located in areas of low socioeconomic status. Almost all (87%) of the teachers provided ratings for a total of 1,161 boys. To create a homogenous sample, only subjects whose parents were Canadian-born with French as a first language and 14 years or less of schooling were included in the longitudinal study, which reduced the number to 1037 boys.

These assessments used the disruptiveness scale of the Social Behavior Questionnaire, measuring the frequency of physical aggressions, oppositional behavior and hyperactive behavior (Cronbach $\alpha = 0.93$). Boys with a score above the 70th percentile (N = 250) on those disruptiveness scales were considered to be at very high risk of later antisocial behavior (Vitaro et al., 2001). The sample for this paper consists of these 250 individuals.

These 250 subjects were randomly assigned to a group treated by a preventive program (69 boys) and a control group (181 boys). Random assignment was carried out by drawing names from a box. The control group was initially divided into two groups, one of which was a no-contact group during the adolescent years. The no-contact group and the control group showed no difference in outcomes and so the two groups were collapsed into one control group. Some families (78 out of 250) from both the treatment and the control groups refused to participate in some elements of the study, particularly in the elements involving parent participation, but were included in the longitudinal data collection. The rate of non-participation was the same across groups. These subjects are included in the analysis as belonging to their initially assigned treatment groups (intention-to-treat analysis) in order to avoid any bias due to differential. We use intention-to-treat analysis and use initial group assignment as the indicator for treatment. Figure 1 provides a visual depiction of the recruitment and selection process of the MLES.

2.2 Intervention Program

The intervention program was implemented over a 2-year period, from ages 7 to 9 years (Grades 2 and 3). The prevention program included three main components to improve children's non-behavioral skills, particularly social interactions and self-control: child

training, parent training, and teacher training. The teacher training component was not fully implemented and so we focus on the child and parent training here. Appendix section B provides detailed information on the program and contents of the training sessions.

2.2.1 Children social skills training

The first aspect of the intervention consisted of direct training on social skills to children. The experiment drew on non-randomized studies from psychology showing correlation between pro-social behavior and specific skills such as emotional regulation and impulse control, social-information processing and how to interpret other's intent (Cartledge and Milburn, 1980; Kettlewell and Kausch, 1983; Michelson et al., 1983; Schneider and Byrne, 1987; Weiss et al., 1992; Dodge 2003 and 2007).

The social skills training took the form of sessions at school (outside the classroom), which grouped four to seven children together, of which one or two would be the treatment subjects, and the rest would be boys identified by their teachers as pro-social. This arrangement was intended to avoid stigmatizing the treatment subjects and also to provide positive role models for the treatment subjects. The sessions were held once a week for 45 minutes, during lunch or after school. During the first year, nine sessions of social behavior training were offered. Sessions included how to invite a bystander to play, how to ask "why", how to give a compliment, and how to help. The second year included 10 sessions of problem solving and self-control strategies (Camp et al 1977; Goldstein et al, 1980). Some stimulus situations for these sessions were how to react to teasing, how to react when angry, and what to do if other children refuse to play. For each situation, the children reviewed ways to define the problem, identified the intentions of the other person, analyzed their feelings if they were in the role of the victim, suggested different action plans to solve the problem, anticipated their consequences, selected one action plan and, finally, reinforced themselves for their cognitive work. Verbal instructions, coaching, modeling, behavior rehearsal, and positive (verbal and material) reinforcement were all used. Children were encouraged to use their newly learned skills before the next training session. At the following meeting, the children were reinforced for having performed their new skills in the interim. Teachers and parents were informed through one-page letters of the new skills learned by the children during each session. They were encouraged to praise the child for using these new skills as often as possible.

For example, one session covered Self Control. The facilitator introduced the topic,

and talked about situations where children are upset and might want to make an angry outburst, like a spilled glass of milk or a disappointment. The facilitator then modeled a situation: she has been playing tag, and she just got tagged and is now out. She's upset because she is the first person to be tagged out, and she's angry and disappointed. The facilitator demonstrates how children can respond in this situation: noticing clues in her body that she is going to lose control (clenching fists, feeling hot), she thinks about what happened to make her feel this way (she got tagged first, is worried other kids will laugh at her), she chooses a way to avoid making an angry outburst (count to ten, move away, say to herself "calm down", breathe), and then she acts and praises herself. The facilitator then invites children to perform additional role-plays based at school (one child bumps another's desk and their pen falls down), at home (someone suddenly turns off the TV because it's time for dinner) or while playing (a friend takes a ball that was dropped). Together, the group makes observations about what the actors are doing, how they are following the steps, and gives feedback. At the end of the session, the facilitator fills out a workbook with the children that identifies how they will practice self-control until the next session as "homework".

2.2.2 Parental Social skills training

The second component of the MLES consisted of parent training in child rearing. This intervention was modeled on the Oregon Social Learning Center model (Patterson, 1982). General goals of the family intervention were to reduce coercive interactions, increase consistency of consequences for behaviors, and improve family conflict interactions. The program focused specifically on setting clear rules, supervising the child, positive reinforcement for pro-social behaviors, and problem solving skills. The family intervention was tailored to individual family needs and capacities. It included: (1) a reading program for parents; (2) training parents to monitor their child's behavior; (3) training parents to positively reinforce their child's prosocial behavior; (4) training parents to punish effectively without being abusive; (5) training parents to manage family crises; and (6) helping parents to generalize what they have learned.

Work with the parents was planned to last for 2 school years with one session every 2 weeks. The sessions were intended to be weekly and last for one hour, with the frequency of the sessions decreasing over time. Parents participated in an average of 17.4 sessions, with a maximum of 47 sessions.

An example of the parental social skills training is given by the "Observation period".

When focusing on encouraging pro-social behaviors, the facilitator would first ask the parents to undertake a period of observation. The parents would spend some time each day observing, and recording on a worksheet, the interactions between the child and parents (for example, "Mother asks for help setting table, child whines and grumbles, mother threatens punishment, child complies."). These observations would then form the basis for identifying patterns of behavior, and the facilitator could propose some strategies for addressing these behaviors. Role-playing might reinforce these strategies; particularly as playing the role of the child allows the parent a different point of view on their own actions. For example, in the case of pro-social behaviors, the facilitator would suggest letting the child know what is expected of him at the beginning of the day, so that there is no room for argument, and this interaction might be role-played. Finally, when the facilitator and parents had identified a few behaviors for work during the coming week, a "contract" between the parent and child might be drawn up, where the desired behaviors or tasks are listed ("Set the table", "Say hello when someone comes in" etc) and children would receive points each time the behavior manifested itself or the task was completed. Children could then use points for some reward: A game with dad after dinner, an extra half hour before bedtime, or some other special treat. In the next session, the facilitator would review the enforcement of the contract with the parents, and adjust goals and strategies as necessary.

A third component of the intervention consisted of information and support for teachers that focused on responding to the behavior of the at-risk boys, which was intended to reinforce the parent training. The intention of this third component was to improve teachers' management skills of behavior problems in the classroom and set up individualized behavior management programs for the participants. However, compliance with this aspect of the intervention was weak.

The training and support activities were carried out by two university-trained child-care workers, one psychologist, and one social worker, both working full time. Each of those therapists had a caseload of 12 families. The team was coordinated by a fifth professional who worked on the project half time. The intervention program lasted 2 school years, from September 1985 to June 1987. Boys were 7 years of age when the intervention started and 9 years of age when it ended.

2.3 Timeline and Data

Figure 2 shows the timeline for data collection. Following the two-year intervention program between ages 7-9 (1985-1987), the MLES collected measures of cognitive and non-cognitive skills during the entire childhood and adolescence on a yearly basis (from age 10 to age 17, period 1988-1995). Those measures will be used to identify the channels of impact of the MLES experiment on various outcomes. The outcome variables are measured with administrative data and questionnaires. When participants were 21 and 26 years old (2001 and 2006), the MLES run adult questionnaire with detailed measures of the respondent economic and social situation. Around age 23-24 (year 2003), the MLES collected administrative data on criminal record and school achievement. We present the main outcome variables of interest below.

Economic Outcomes. The MLES collected a wide range of adult economic outcomes. Subjects reported their activities (employment, school, recipient of government transfers) at age 21 and 26, and their recalled activities each year since age 17. These data are used to construct several variables, which we use to test the impact of the program on economic outcomes. The first variables are dummies indicating whether they were employed and whether they occupy a full-time job. The second set of variables are the percent of years from 17-26 either working or studying full-time, the percent of years from 17-26 working, the percent of years from 17-26 receiving government transfers, the percent of years from 17-26 spent inactive. We also measure the wages at age 26.[3] Finally, we construct a dummy that indicates whether at age 21 they were engaged in either school or work, and a dummy that indicates whether they were engaged full time.

We construct a composite of these economic variables in two ways. First, following Kling et al (2007) we test an index of employment status at age 26, hourly wage, full time (in either work or study) status at age 21, percent of years from 17-26 receiving transfers, and the percent of years from 17-26 occupied full time in either work or study (or both). This index is a simple average of the z-scores of the different variables; reoriented so higher values are more desirable outcomes.[4] Second, we construct a factor using maximum likelihood estimation as with the measurement variables, below, which allows the different components

[3] We impute wages for the unemployed based on previous employment and secondary graduation. However, since there is a large number of unemployed, the results are sensitive to the assumptions used to impute wages for the unemployed (such as the replacement rate). For this reason, we do not focus on hourly wages as a principal outcome, though we do not exclude it from the index. Excluding it does not substantially change the results.

[4] As in Kling et al (2007), missing values are imputed using the average value for the assigned group, provided that the observation has a valid value for at least one variable.

to have different weights in the composite variable. The distributions of resulting composite variables are fairly similar regardless of the technique.

Educational achievement. The GRIP laboratory also collected administrative information on whether each subject had received a secondary-school level degree from the Ministry of Education of Quebec. This information was collected when the participants were around 23-24 years old (2003).

Criminal record. The Ministry of Justice in Quebec provided administrative data on the Criminal Records, including the number of offenses, whether the offense was violent or nonviolent, whether a weapon was involved, and year of each offense. This information was collected when the participants were around 23-24 years old (2003) and includes adult criminal records prior to age 23.

Social capital. Subjects reported on multiple facets of their interactions with others in questionnaires at age 21 and 26. In particular, we have various traditional measures of social capital including whether the person voted in the last election, is the member of an organized group (for example, community groups or sport clubs), and volunteers (see Coleman, 1970 or Putnam, 2000 for a discussion of these measures). As with the economic data, we construct two alternative indexes, one is the average of the z-scores of each variable within a category; the other is a factor using maximum likelihood estimation with the measurement variables.

Table 1 shows the summary statistics for the administrative data on criminal record and high-school diploma, and summary statistics for economic and social capital outcomes from the adult questionnaires at ages 21 and 26.

2.4 Validity of the experiment design and baseline controls

The validity of the causal inference from the MLES rests on the quality of the randomization. To test for successful randomization, we test for balance in pre-determined variables between the treated and the control group.

Table 2 shows the baseline values of two groups for several critical variables measured prior to randomization. Among the 32 baseline variables, there are significant differences on 4 variables: initial anxiety measures, age of father at birth of subject, prestige of the mother's employment (all at the 10% level), and number of sisters (at the 1% level). The fact that there are some differences does not indicate that the selection process was non-random. It is not surprising to find imbalances for a handful of variables, given that the sample size is small

and 32 variables were tested for differences and so we do not fear that the randomization protocol was violated. However, since these variables might impact the channels and outcomes we wish to examine, we control for them in the analysis that follows. Since some of the subjects are missing these control variables, we impute missing values for the control variables, using available baseline data, to maintain sample size.

Missing data is a problem for many of the outcomes and intermediate variables. One potential solution is dropping observations with missing data; the other is to impute the missing data points using data from other individuals in the sample to predict the missing data point. The former brings the risk of differential attrition and bias, as well as the possibility of substantially reducing sample size (which is already small). The second is not possible when the variable of interest is itself an outcome of interest.

We have imputed certain baseline measurements that are necessarily to eliminate an imbalance between the groups (see above) but omit observations when the intermediate or outcome variable of interest is missing. Father's age, prestige of the mother's work, and number of sisters were imputed using all other available baseline information. We include a dummy equal to one when the value is imputed. In the maximum likelihood estimation, where all channels must be included to control for one another, omitting observations with missing values would lead to a substantial reduction in sample size, we impute missing values using maximum likelihood.

3. The MLES Effect on Adult Outcomes

3.1 Crime

Figure 3 shows the percent of people in each group who have a criminal record of any kind. We find a large negative point estimate of the coefficient, significant at 10%: the treatment group is one-third (11 percentage points) less likely to have a criminal record than the control group to have a criminal record (Table 3 – Column 3).

3.2 Secondary Diploma

Figure 4 shows the percent of people in each group who completed secondary school according to administrative records. Table 3 - Column 2 shows a significant and very large coefficient: subjects in the treatment group are 19 percentage points more likely to have received a secondary school diploma than subjects in the control group, significant at the 5% level (note that only 32% percent of control group subjects completed secondary school).

3.3 Economic outcomes

Figure 5 shows the kernel density for the economic index of the treated and control groups. Table 3 - Column 1 gives the effect of the early childhood intervention on the composite economic index, which is 0.2 standard deviations, and significant at the 1% level (controlling for the baseline variables discussed in section 2). Table 11 gives the coefficient on treatment for the variables that compose the economic index, and we find both a sizeable and statistically significant effect for several individual variables. In particular, treatment subjects are 0.39 standard deviations (10 percentage points) more likely to report being employed in 2006 than control subjects, significant at the 5% level, and 0.50 standard deviations (18 percentage points) more likely to be employed full-time in 2006 than control subjects, significant at the 1% level. Over the period from age 17-26, subjects from the treatment group spent 0.36 standard deviations (10 percentage points) more time working than subjects from the treatment group, significant at the 5% level. Subjects from the treatment earn also a higher hourly wage than the control group, however the magnitude and significance of the effect depends on the assumptions used to calculate the replacement rate among the unemployed (15% among the control group against 4 % among the treated).

3.4 Social Capital

Figure 6 shows the kernel density of the social capital index. Table 3– Column 4 gives the regression estimates of the effect of the MLES on social capital. We find that the treatment has a positive effect on social capital (over 0.2 standard deviations, statistically significant at the 10 percent level). Table 12 shows that this result is mainly driven by group membership: subjects were almost 20 percentage points more likely to belong to a social group, significant at the 10% level.

4. Identification of the Channels during adolescence

This section estimates the effect of the MLES experiment on changing individual traits measured during adolescence, identified later as the channels of impact on adult outcomes. Measures relative to the different channels were collected annually from ages 10 and 17 years, as illustrated in Figure 2. We use the average over all years available for those variables that are constant over time, and for those that are not, we use the most recent observations. Because this paper does not aim to compare this population to standardized measures of the channels, we use individual questions from the psychological scales rather

than the calculated scales themselves. This approach allows for the possibility that individual questions might cluster together effectively and also allows data from different sources (self-reported, administrative, teacher-reported) to be used together.

We base our identification of channels on an exploratory factor analysis. Combining all data available and averaging over the years available, we use factor analysis to examine how the factors combine into groups potentially measuring the same latent variable. We include questions from several psychological inventories (Jesness and Wedge, 1983; Kovacs, 1983; March, 1990; Rosenberg, 1965; Lacourse et al., 2002; Tremblay et al, 1992). Parallel analysis identified 13 potential factors, and after dropping variables with weak loadings and iterating on the process, we identified 8 factors. The factor extracted by the EFA procedure are linked to channels that have been already documented in economics and psychology, which gives greater confidence in the EFA procure. We discuss the channels below.

Previous studies on this sample have found that subjects are on different trajectories of behavior, that is, subjects may be grouped as to whether their disruptiveness is increasing, decreasing, or stable over time with respect to their initial disruptiveness levels (Tremblay et al., 1991). For this reason, it may not be appropriate to use averages over time to calculate the channels when several years are available. We examined the correlation over time of the different variables identified for the factors and, for those that varied substantially over time, used the most recent data (generally 1995) and for those that did not vary substantially we used averages over all years available (some data are available for only one year). For education outcomes (specifically grades) we used data from 1992-1995 since some subjects have dropped out prior to 1995, and by averaging over this period we could preserve some data on grades for these individuals.

Table 4 shows summary statistics for the construction of each channels: number of components (Col. 1), the Cronbach alpha index for internal consistency (Col. 2), the number of observations (Col. 3), p-value for the test of equality of mean (Col. 4) and distribution (Col. 5) between the treated and the control groups, and the p-value of the channel impact on economic outcome (Col. 6). Table 5 shows the correlation between the identified channels. Tables 14-20 list the measurement variables, their contribution to the channel being measured, and the relationship of the measurement variable to treatment.

Social Trust. The social trust channel is measured by variables that ask the subject about their ability to trust others (friends, strangers, the police), the quality and closeness of their social relationships, and their ability to correctly interpret the actions of others. The recent

literature in economics and sociology demonstrate that social trust is strongly associated with beneficial economic outcomes (see Banfield, 1958; Coleman, 1990; Putnam, 1993 and 2000 for seminal research in sociology; and Fehr 2009 or Algan and Cahuc 2013, for recent surveys on social capital in the economic literatures). Social-information processing, and in particular the hostile attribution bias (the way one may treat social cues and assume the worst of other's people intention such as interpreting an accidental bump as a purposeful shove), have been shown to be strongly correlated with the ability to trust and cooperate with others (Dodge 2011).

School Performance. The school performance channel is measured by information on grades (French and math), status in school (having been held back a year, assigned to special education class, and dropping out), and final achievement of a secondary school diploma. The first variables were reported by the school, and the final variable was obtained from the Ministry of Education. We construct school performance as both a latent channel and as an index, and we find no substantial difference in results though the distributions tend to be visually quite different because of the large contribution of binary variables in the maximum likelihood solution.

Delinquent Behavior. The literature in criminology and psychology shows that early anti-social behavior, or externalizing behavior, is highly correlated with adolescent and adult criminality (Boisjoli et al. 2007; Tremblay 2000). Recent literature in economics shows that childhood misbehavior is associated with poor later education and labor market outcomes (Segal 2013). Recent evidence also shows that antisocial and externalizing behavior, measured by behaviors related to lying, stealing and swearing, is a significant channel of impact of the Perry Preschool program on economic outcomes (Heckman et al. 2012). The Delinquent Behavior channel is measured by 37 questions (self-reported, teacher-reported, and peer reported) and one administrative record (the existence of a juvenile criminal record). Delinquent behavior is behavior directed towards others that can be hostile or aggressive, including physical or verbal aggression, vandalism, theft, and defiance. The questions that are self-reported include questions about theft, violence, and vandalism, and the questions that are teacher-reported include questions about the extent to which the subject fights, tells lies, dominates others with physical force, threatens others, etc. The peer-rating evaluations were run from ages 10 to 13 years using the Pupil Evaluation Inventory. This inventory includes 34 behavior descriptions grouped in three scales: aggressiveness, social withdrawal and

likability, of which we use the aggressiveness scale with items where peers compares the participant as someone "who is mean and cruel to others" or "who can beat everybody up".

Hyperactivity. Hyperactivity, or deficit hyperactivity disorder (ADHD), has been shown to be a statistically significant predictor of academic achievement (Szatmari et al. 1989). In particular recent research demonstrates that part of the positive gender gap between girls and boys in college attendance is due to higher hyperactivity among boys (Jacob 2002; Becker et al. 2010, Fortin et al. 2011; Bertrand and Pan 2011). Our Hyperactivity channel is measured by six questions answered by teachers. These questions are from the Social Behavior Questionnaire. They include the extent to which the student fidgets, is easily distracted, disobedient, and jumps from one activity to another.

Self Esteem. The Self-Esteem channel is measured by a 41 self-reported questions from multiple psychological scales that were grouped together by the EFA. This channel is composed of measures of self esteem: subject's agreement with statements such as "I often feel sad and alone", "I don't succeed at anything", and "I fail at everything I try."

Altruism. The EFA methodology suggests a distinction between Social trust and Altruism factors. This is consistent with the experimental economics literature, whereby social trust entails a component of strategic social interactions skills absent from altruism (Fehr 2009). The Altruism channel is composed of 10 questions from the Social Behavior Questionnaire, answered by teachers. The questions address behavior that is pro-social and specifically that demonstrates generosity towards others: the extent to which the subject would comfort a crying child, volunteer to clean up a mess, show sympathy, or offer to help a child in trouble.

School Attitude. School attitude and academic motivation have been found to be a statistically significant predictor of performance in school and adult economic outcome. Most of the future employment gain from the Perry Preschool program for girls has been attributed to a boost in academic motivation, although no effect was found for boys (Heckman et al. 2012). Similarly, the long-run effect of the Tennessee STAR class size experiment on adult income is mainly attributed to its positive impact on academic motivation measured by questions on effort and motivation (Chetty et al. 2012). The School Attitude channel is measured by 16 self-reported questions on how the subject felt about school and his performance at school, including how much they liked school, whether they felt they could talk to the teacher, thought grades were important, and whether they felt they could succeed at school.

5. Testing the Channels of impact on Adult outcomes

This section estimates the impact of the induced changes in channels, identified above, on changes in adult economic and social outcomes. We first present the methodology to decompose the mean treatment effects on outcomes into components corresponding to changes in each channel. We then estimate decompose the main channels of impact for crime, education and economic outcomes. This methodology draws on Heckman et al (2012) approach.

5.1 Methodology

Our model of impact relies on Heckman et al (2012). An outcome Y depends on the treatment group assignment D, where D=1 if a subject is in the treatment group, so that

$$Y = DY_1 + (1\text{-}D)Y_0$$

Y_d is the outcome for treatment group D, and is determined by a vector of baseline characteristics X, and a vector of J channels, θ, as follows

$$Y_d = \varkappa_d + \alpha_d\theta_d + \beta_d X + \zeta_d \qquad (1)$$

where ζ_d is an error term assumed to be uncorrelated with θ_d and X. Even if the MLES collected a very large set of measures during childhood and adolescence, it is very likely that not all channels are measured. In addition, even among the skills that are measured, some measurements are likely to be imperfect. This means that the term $\alpha_d\theta_d$ is made up of measured and unmeasured skills. Following Heckman et al (2012), we isolate the impact of

the channels we measure, denoted by the vector θ_d^{obs} , and channels we do not, θ_d^{inobs} by rewriting equation (1)

$$Y_d = \tau_d + \alpha\theta_d^{obs} + \beta X + \varepsilon_d \qquad (2)$$

where τ_d is equal to $\varkappa_d$ plus the unmeasured component of $\alpha_d\theta_d$; and ε_d is a zero-mean error equal to ζ_d plus the error term of the unmeasured component of $\alpha_d\theta_d$. We have further assumed in equation (2) that α_d and β_d are the same for treatment and control groups, which means that the treatment affects the level of the skills, but not the effect of the skills and background variables on outcomes. We have tested and did not reject this assumption. Note that the error term is correlated with the measured channels if those measured channels are correlated with the unmeasured ones.

In order to decompose the treatment effects into the contributions of the measured skills, we write the impact of the program, $E(Y_1\text{-}Y_2)$ as

$$E(Y_1\text{-}Y_0) = (\tau_1\text{-}\ \tau_0) + \alpha E(\theta_1^{obs} - \theta_0^{obs}), \qquad (3)$$

so $(\tau_1\text{-}\ \tau_0)$ is the proportion of the impact that is not explained by changes in measured channels, and $\alpha E(\theta_1^{obs} - \theta_0^{obs})$ is the proportion of the impact that is explained by changes in measured channels. Thus the model allows for a direct impact of treatment on the outcome and also an indirect effect via θ.

To measure the channels, we use a factor analysis. This method allows us to form interpretable and low-dimensional latent variables that summarize the covariability among the observed measures. This method is less arbitrary than just taking the average of the different measures since the weight for a given measure is data-driven, using information from the measure's covariance with other measures. This method also accounts for measurement errors (Heckman et al., 2012).

We model J underlying channels, θ_j as measured by a group of k measurement variables m_j^k, where each group of measurement variables applies to only one channel. Each channel is then measured using the following system of equations:

$$\begin{aligned} m_{jd}^{1} &= \mu_j^{1} + \phi_j^{1}\theta_{jd} + \eta_j^{1} \\ m_{jd}^{2} &= \mu_j^{2} + \phi_j^{2}\theta_{jd} + \eta_j^{2} \qquad (4) \\ &\cdots \\ m_{jd}^{k} &= \mu_j^{k} + \phi_j^{k}\theta_{jd} + \eta_j^{k} \end{aligned}$$

where parameters ϕ_j are the factor loadings for the channels θ_j and parameters μ_j are measure-specific intercept. The parameters η_j are zero-mean errors assumed to be independent of θ_j and i.i.d. We assume that the parameters μ_j, ϕ_j and η_j are independent of d, implying that the impact of the treatment on the channels do not depend on the measurement system for those channels.

To estimate the total impact of treatment on outcomes, we use OLS with robust standard errors for the continuous variables (economic and social indices) and logit for the dummy variables (criminal record and secondary school completion), reporting marginal effects, though the Logit specification does not differ substantially from a linear probability model.

Our principal method for decomposing the variance is maximum likelihood estimation, which we carry out in two steps. First, we estimate the set of equations (4) for each θ_j in order to obtain a predicted value of each channel based on the measurement variables. We allow for covariance between the measurement variables. Second, we simultaneously estimate equation (2) for all θ_j together, again allowing for covariance between the channel variables. We use the residuals from linear regressions of the outcome and measurement variables on the baseline variables in order to control for those baseline variables. This vastly simplifies computation and does not change substantially when the control variables are included directly.

5.2 The effects on channels of impact

Determining whether each channel explains part of the treatment effect requires first that the channel be changed by the treatment and second that the channel itself impacts the outcome. In addition, we should find an impact on channels that are anticipated to be mediators – that is, channels that the program was designed to change, mainly related to social skills (Maric et al., 2012; MacKinnon et al., 2004).

Figures 8a and 8b show the treatment has a significant positive impact on social trust (p=0.02) and on school performance (p<0.01). Figure 9a and 9b shows that the Delinquent Behavior channel (p<0.05) and the Hyperactivity channel are also both significantly reduced by treatment (p<0.10). In contrast, Figures 10a and 10b how that neither the Self-Esteem channel (p=0.19), nor the altruism channel (p=0.19) are significantly related to treatment, nor is the School Attitude channel or the Gregarious Behavior channel (p=0.15 and p=0.63). We will

thus focus on four channels of impact: social trust, school performance, delinquent behavior and hyperactivity.

5.3 Decomposition of the channels of impact on outcomes

We estimate the proportion of the impact of treatment on economic outcomes explained by changes in the channels using maximum likelihood estimation of simultaneous equations. This methodology allows flexibility in the assumptions about the nature of the covariance of the channels, and whether the channels are estimated simultaneously with the treatment effect. We run several different specifications under different assumptions about the covariance of the channels and their components, the construction of the school performance channel and the outcome composite variable, excluding outliers, and estimating the channel simultaneously with the estimation of the mediation. For each specification, we estimate the total impact of treatment on the outcomes and the indirect impact that is associated with changes in each channel. We use this to calculate the percent of the total treatment impact that is explained by changes in the channels. For detailed coefficient results from the maximum likelihood estimation under different assumptions see the Appendix.

Figures 11-15 show the decomposition of the explained treatment effect on Crime, Secondary Diploma, Economic Index and its sub-components, and the Social Index and its component. Each bar represents the explained treatment effect, normalized to 1, and the decomposition represents the percent of the explained treatment effect that is explained by changes in the channels. We also report the residual non-explained variance at the bottom of each figure. The unexplained share corresponds to the direct treatment effect on outcomes that is non-mediated by the identified channels. This unexplained treatment effect ranges from 7% for Crime, 43% for Secondary Diploma, 55% for the Economic index, to 84% for the Social index. Tables 6-9 show the detailed decomposition and the associated p-value for each channel, and Tables 21-34 provide the coefficients from the maximum likelihood solutions. We report both significant and non-significant channels, since the value of these estimates is of interest as well as changes in portion explained between different specifications.

In this decomposition exercise, we test four main specifications depending on the outcome:

Simple I. This first specification allows: (a) treatment to predict the behavioral channels (Delinquent Behavior, Social Trust and Hyperactivity) and school, where school is a combination of performance during secondary school and secondary school completion

(secondary diploma, as an outcome, is not included in this specification) and (b) the behavioral channels and school to predict the outcomes.

We observe that social trust accounts for a considerable part of the treatment effect for all three outcomes: 15% for Criminal Record (p= 0.12), 50% for Economic Index (p=0.086), and 26% for Social Capital (p=0.22). Delinquent Behavior explains a portion of treatment impact only for Criminal Record (about 30%, p=0.03). School accounts for a large portion of treatment effects across all three outcomes: nearly 60% for Criminal Record (p=0.01), 50% for Economic Index (p=0.093), and 26% for Social Capital (p=0.42), where percentages reflect the percentage of the explained treatment effect.

As the intervention did not include tutoring or additional class time, which might have improved grades or school performance, the importance of schooling is unexpected, and raises the question of what role school is playing, and what the school variable is measuring.

Simple II. Since performance in high school might play a different role than actually having a diploma (particularly with respect to economic outcomes, where a diploma may be an important signal), we separate school into performance (a composite variable including grades, dropout, special education classroom assignment and failing a grade) and diploma in specification 2, which is otherwise identical to Simple I. We also add Secondary Diploma as an outcome variable (which has one fewer predictor). Here we observe that for the Economic Index and for the Criminal Record, school performance accounts for more of the treatment effect than the secondary school diploma itself: 39% vs 20% for Criminal Record (p=0.02 and 0.06), and 30% vs 17% for the Economic Index (p=0.19 and 0.267). For the Social Capital index, Secondary Diploma captures the entirety of the school effect, and captures more of the effect alone than when combined with school performance (51%, p=0.26). When the Secondary Diploma is considered as an outcome itself, in this simple specification, 80% (p=0.00) of the explained treatment effect comes from the School Performance variable, 14% (p=0.11) from Hyperactivity, and 7% (p=0.19) from Delinquent Behavior (and 0 from Social Trust).

Cross-Channel I. In specification 3, we allow a cross-channel impact, using behavioral channels to predict school performance and Secondary Diploma, as one possibility is that the changes in school performance and achievement are due to changes in behavior, and permitting this cross-channel impact properly allocates those impacts to the behavioral

channels. We estimate this specification only for Criminal Record and Secondary Diploma. For Criminal Record, we find that the portion of impact explained by Delinquent Behavior increases slightly with this specification from 30% to 38% (p=0.02), and Social Trust increases slightly from 17% to 20% (p=0.08). For Secondary Diploma, the portion of impact explained by Delinquent Behavior increases substantially from 4% to 15% (p=0.04) but Social Trust remains at zero.

Cross-Channel II. In our final specification, we go one step further and allow the behaviors likely to have been directly impacted by the intervention (hyperactivity and social trust) to explain Delinquent Behavior as it explains Criminal Record and Secondary Diploma. For Criminal Record, some of the treatment previous explained by Delinquent Behavior is now explained by Hyperactivity: Delinquent Behavior drops from 38% to 26% (p=0.06) and Hyperactivity increases from 3% to 17% (p=0.09). For Secondary Diploma the change has a similar pattern but is smaller: Delinquent Behavior drops from 25% to 17% (0.09) and Hyperactivity increases from 19% to 27% (p=0.08). At this stage, our cross-channel analyses examine the mediators simultaneously (e.g in parallel) instead of sequentially. In particular we do not impose time lags between Delinquent behavior and its explained factors Hyperactivity and Social trust. Future research will look at whether a chain of mediators could explain the results better than parallel and thus competing mediators.

5.4 Interpretation of Results

We find that the changes in criminality in the treatment group are explained by changes in Delinquent Behavior and Schooling, both of which are to some extent explained by changes in Hyperactivity, and also Social Trust. Changes in the likelihood of obtaining a Secondary Diploma are due to improvements in in-school performance, about one-third of which is explained by reductions in Hyperactivity and Delinquent Behavior. Improvements in Social Trust account for about half of the explained impact on Economic Outcomes, and Schooling, particularly in-school performance, accounts for the rest.

Schooling thus remains important in explaining the impact of the program, and we offer several possible interpretations for this: First, higher levels of schooling might lead to greater later-life success (for example on the labor market) which could in turn reduce criminality. This seems unlikely, since actually obtaining Secondary Diploma is less important than performance while in secondary school. Second, higher levels of in-school performance

might reflect greater time spent at school, rather than skipping school and spending more time in environments that could lead to more exposure to delinquent behavior and crime. Third, performance in-school is in some sense a measurement by the teachers (particularly assignment to special education class), and the teachers may be measuring non-cognitive skills with a precision not available from the questionnaires we use.

Finally, recall that obtaining a Secondary Diploma is a weak channel once performance in-school is accounted for. This is related to the literature that argues that it is the non-cognitive skills acquired in secondary school, rather than the diploma itself, that help people in the labor market (Heckman and Rubinstein 2001, Chetty et al. 2012).

6. Rate of Return

In order to estimate Δ^I, the change in income due to the program, we first estimate the increase in hourly income due to increased employment. We compare the hourly income of the employed to the unemployed without secondary education, controlling for family characteristics and behavior indices at age 6 ($I_E^{NS} - I_U^{NS}$), and multiplying this by the increase in probability of being employed due to the program ($\Pi(E^T) - \Pi(E^C)$). We then estimate the increase in hourly income due to secondary school completion contingent on employment by comparing the hourly income of secondary school graduates to drop outs, contingent on being employed controlling for family characteristics and behavior indices at age 6 ($I_E^S - I_E^{NS}$), and multiplying this by the increase in the probability of completing secondary school due to the program ($\Pi(S^T) - \Pi(S^C)$). We add these two benefits together to obtain the overall increase in hourly income due to the program. That is,

$$\Delta_I = [(I_E^{NS} - I_U^{NS}) * (\Pi(E^T) - \Pi(E^C))] + [(I_E^S - I_E^{NS}) * (\Pi(S^T) - \Pi(S^C))]$$

We use data from the control group and the normative (non-disruptive) group to calculate these estimates in order to maintain a larger sample size. Since the increase in hourly income due to employment is large in this sample, we provide alternative estimates using the minimum wage. Table 10 documents the different assumptions.

We assume that people work 20 hours per week for 52 weeks per year for 40 years. The additional income stream over the next 40 years is discounted by 3% in order to obtain the present value in 2004 dollars. This yields a present value of $ 42,660 in 2004, or $40,862 in the CPI reference year (2002) using the wage data, or $29,511 and $28,267 using the minimum wage. Unfortunately, we do not have information on the actual cost of the

program. However, we do know the composition of the implementation team (1 full time social worker, 2 full time childcare specialists (BA level), 1 psychologist, and 1 half-time program administrator). We use reference salaries for these positions in Montreal in 2011, and we assume an additional 20% of salaries in administrative costs. We then calculate how much the program would cost, per person, to be implemented in 2011 (for two years). We calculate that the total cost per person would be $8,736 in 2011, or $7,406 in 2002.

These estimates give an overall rate of return of 282% using the minimum wage and 452% using the wage data, implying that every $1 invested yields $2.8 to $4.5 in benefits. While high, this estimate is in line with estimates of the rate of return to investments in early education. Note that this is a low estimate of the benefit of this program, for example, it does not take into account costs of social transfers and criminality (see, among others, Heckman et al. 2010 and Heller et al. 2013 for a detailed synthesis of the literature).

7. Robustness checks: Missing data and attrition

7.1 Attrition rate

In longitudinal studies, attrition rate might bias the result. This is not an issue here for the outcomes crime and secondary school completion that come from administrative data. The attrition rate is 0 percent for criminal record (we have data for the entire sample of 250 subjects) and 3 percent for secondary diploma (8 missing observations). But this is much more of an issue for the self-reported socio-economic outcomes. Table 35 shows that from the original 250 subjects, roughly 44% are missing employment data at age 21, and 59% are missing employment data at age 26. This attrition comes both from those that did not respond to the questionnaires, and those that did not complete that question on the questionnaire. Proportionally more treatment than control subjects are missing, although the difference is not statistically significant (p-value= 0.3662)

This level of attrition raises concern that our estimate of the impact of treatment on self-reported socio-economic outcomes may be biased. It is relatively easy to propose reasons that attrition might not be random and uncorrelated to treatment, and not clear which direction the resulting bias would be. For example, suppose that the higher level of attrition in the treatment group is due to higher job mobility in the treatment group due to greater economic success: then the most successful of the treatment group might have moved away from Montreal, which would result in a downward bias of the impact estimate. On the other

hand, suppose that the treatment subjects who were the least successful felt more ashamed of their lack of success than the control subjects, because they felt that the University of Montréal had invested in their development, and so they were less likely to respond to the questionnaire – this would result in an upward bias on the impact estimate.

7.2 Determinants of Attrition

We test whether attrition is related to 25 variables collected prior to the program: both parents living at home; age of mother and father at birth of subject; education of mother and father; work status of mother and father; number of children in the household; adversity index of household; age of subject in 1984; rank in birth order of child; number of brothers, sisters, half-brothers, and half-sisters; whether or not the subject went to pre-school; and initial scores on aggression, anxiety, opposition, pro-sociality, combativeness, inattention, hyperactivity, and anti-social behavior. We construct three dummy variables equal to 1 if data on employment is missing and 0 otherwise (one for age 21, one for age 26, and one for percentage of time worked full time). For each baseline variable to be tested, we construct an interaction term between treatment and the baseline variable (in order to capture any possibility of differential attrition between the groups along that variable). We conduct 26 logit regressions for each year (25 baseline variables plus treatment alone), where the dummy for missing is regressed on treatment, the interaction term, and the baseline variable, as follows:

$$MISSING_i = \alpha + \beta\, TREAT_i + \gamma\, (TREAT_i * BASEVAR_i) + \delta\, (BASEVAR_i) + ERROR_i$$

We do not combine all of the variables into one regression because of the small sample size. Of greatest concern are the coefficients β and γ, because if they are significant they indicate that attrition is not orthogonal to treatment (random) along that baseline variable, and so at a minimum, that baseline variable should be controlled for in the regression specifications.

Table 37 shows all the significant channels. Father's age at subject's birth and schooling, are negatively correlated to attrition, but the correlation is the same in treatment and control groups. Mother's employment is weakly positively correlated to attrition in the treatment group for employment at age 26. Number of sisters is weakly positively correlated to attrition on the percent full time variable, but only in the treatment group. Note that the number of sisters is one variable that is not equal across treatment and control groups and is already

controlled for due to that imbalance. Father's employment is very significantly correlated with attrition and treatment status at age 21 and 26. The combination of coefficients implies that, once father's employment is taken into account, treatment subjects are more likely to be missing than control subjects, and control subjects are more likely to be missing if their father was employed. Note that the checks on randomization showed no correlation (near zero and non significant) of father employment with treatment status. Once aggression or fighting is controlled for, the treatment group is more likely than the control group to be missing for employment at age 21 (note that aggression was not correlated with assignment to treatment). Pro-sociality is positively correlation with attrition only for the treatment group at age 26 (neither group at age 21). Hyperactivity is correlated with attrition and treatment group at age 21. Once hyperactivity is controlled for, the treatment group was more likely to be missing, but this difference decreased with hyperactivity. Note that the checks on randomization showed no correlation between hyperactivity and treatment assignment (near zero and non-significant).

Following these results, we include father's employment, mother's employment, pro-sociality, aggression, fighting and hyperactivity as controls in our preferred specification, along with the variables that were significant in the random assignment check (father's age at subject's birth, prestige of mother's work, initial anxiety, and number of sisters).

7.3 Using administrative data on schooling to estimate bias

In order to get an idea of the possible direction and magnitude of the bias, we estimate the bias introduced by the same level of attrition into the estimate of the program impact on secondary school education.[5] Administrative data on secondary school completion is available for nearly the entire sample (242 out of 250 subjects, or 97%). We use the following procedure. We create a false secondary school variable that takes the value of missing if the subject is missing data for employment at age 21 (or 26) and the value of the true secondary school variable if they have data for employment at age 21 (or 26). In this way we mimic the level of attrition in the economic data in the secondary school data (we call

[5] It is sometimes possible to create bounds on the estimate by calculating the coefficient under two hypothetical cases, one where all of the missing from the treatment group are artificially assigned the "bad" outcome and all of the missing from the control group are artificially assigned the "good" outcome to obtain the lower bound, and the converse to obtain the higher bound. This is not useful in this case because the high rate of attrition would result in extremely wide and uninformative bounds.

this the "attrited" dataset). We then estimate the impact of the program on secondary school completion using our preferred paper specification. We control for all variables related to treatment assignment and that are related to attrition across treatment groups. We compare this to an estimate of the impact of the program using data from the entire sample (the "true" dataset. The validity of this comparison rests on the assumption that the direction of the bias induced in the secondary school completion data is likely to be the same as the direction of the potential bias in the employment data.

The first three columns of Table 36 show this test for attrition at age 21. The point estimates of the attritted (column 1) and true (column 2) datasets are very close in value, suggesting little attrition bias. The second three columns of Table 36 show this test for attrition at age 26. The point estimate for the attrited dataset for age 26 (column 3) is 0.133, while the point estimate for the true dataset is 0.164. There is a bias due to attrition (the point estimate in the attrited sample is 23% lower than in the true sample) but note that it is a downward bias. If the estimate of impact for secondary school completion and economic success would biased in the same way by attrition, the estimates used in this paper are likely to be an underestimate of the economic impact of this program.

8. Conclusion

We find that an elementary school program conducted in inner-city Montreal in the 1980s and focused on improving social skills had substantial long-term positive effects on economic and social outcomes: crime rates decreased, secondary school graduation increased, economic outcomes improved, and social capital increased in response to the program. We estimate a large rate of return to this program – over 250% – even under conservative assumptions, which is in line with other estimates of the profitability of early childhood development programs. We use a rich longitudinal dataset that includes subjective and teacher, parent and peer reported-data, as well as psychological inventories, to provide evidence on the channels of impact.

The mechanisms of impact were improvements in Social Trust, and reductions in Hyperactivity and Delinquent Behavior. We also find that in-school performance plays a large role, even once secondary school completion is controlled for, and we believe that this is in part because in-school performance can measure non-cognitive behaviors that are not captured by the other psychological scales. Secondary school completion (diploma) does not

explain much of the impact on economic outcomes and criminality once in-school performance and the measured non-cognitive skills are controlled for.

The policy implication of this paper is that increased investment in early elementary school programs for poor children with social skills deficits is likely to be an efficient and profitable public policy, especially where such programs explicitly incorporate strategies to foster the development of non-cognitive skills, and in particular social skills and self-control. These investments would be particularly cost-effective when made for the children most at risk of unemployment and criminality as adults.

References

Y. Algan, P. Cahuc, Trust, Growth and Well-Being: New Evidence and Policy Implications. *Handbook of economic growth*, in Aghion and Durlauf (eds), Elswier (2013).

Y. Algan, P. Cahuc, A. Shleifer, Teaching practices and Social Capital. *American Journal of Economics: Applied*, (2013).

D. Almond and J. Currie. Human Capital Development Before Age Five, Handbook of Labor Economics, Volume 4, Elswier (2010).

E. Banfield, *The Moral Basis of a Backward Society*. New York: Free Press (1958).

G. Becker, W. Hubbard, K.Murphy, The Market for College Graduates and the Worldwide Boom in Higher Education of Women. *American Economic Review*, **100(2)**, 229-33 (2010).

M. Bertrand, J. Pan, The Trouble with boys: social influences and the gender gap in disruptive behavior, NBER working paper, 17541 (2011)

R. Boisjoli, F. Vitaro, E. Lacourse, E.D. Barker, R. Tremblay, Impact and clinical significance of a preventive intervention for disruptive boys: A 15-year follow-up. *British Journal of Psychiatry*,**191,** 415-419 (2007).

S. Bowles, A (Science-Based) Poor Kids' manifesto, (Review of Heckman James, Giving Kids a Fair Chance: A strategy that works, Cambridge MIT), *Science,* **340,** 1044-1045 (2013).

S. Bowles, H. Gintis, M. Osborne, The determinants of earnings: A behavioral approach. *Journal of Economic Literature*, **39,** 1137-1176 (2001).

B.W. Camp, G.E. Blom, F. Hebert, W.J van Doorminck, Think aloud: A program for developing self-control in young aggressive boys. *Journal of Abnormal Child Psychology* **5**, 157-169 (1977).

F.A. Campbell, C.T. Ramey, E. Pungello, J. Sparling S. Miller-Jonhson, Early childhood education: Young adult outcomes for the Abecderian project. *Applied Development Science*, **6**, 42-57 (2002).

G. Cartledge, J. F. Milburn, *Teaching social skills to children. Innovative approaches* (Pergamon Press, New York, 1980)

N. Castellanos-Ryan, J. Séguin, F.Vitaro, S.Parent, Richard E. Tremblay, A multimodal intervention for disruptive kindergarten children reduces substance-use across adolescence: a randomized control trial, *British Journal of Psychiatry*, forthcoming (2013).

R. Chetty, J. Friedman, N. Hilger, E. Saez, D. W. Schanzenbach, D. Yagan, How does your kindergarten classroom affect your earnings? Evidence from project Star. *Quarterly Journal of Economics* **126(4)** (2011)

D.R. Clasen, B.B Brown, The multidimensionality of peer pressure in adolescence. *Journal of Youth and Adolescence* **14**, 451-468 (1985).

J. Coleman, *Foundations of Social Theory*. Cambridge, MA: Harvard University Press (1990).

J. Currie, T. Duncan, Does Head Start make a difference? *American Economic Review*, **85**, 341-364 (1995).

J. Currie, R. Khan, *The future of Children*, Princeton-Brookings, Washington DC (2012).

J. Dewey, *Democracy and Education*, New York, NY: Dover (1944).

K. Dodge, Godwin J., Conduct Problems Prevention Research Group, Social-information-processing patterns mediate the impact of preventive intervention on adolescent antisocial behavior. *Psychol Sci* **24(4)**, 456-65 (Apr 2013).

K. Dodge, Processes in the Prevention of Crime and Delinquency. *Conrolling Crime: Strategies and Tradeoffs*, P. Cook, J. Ludwig, J. McCrary (eds), University of Chicago Press (2011).

K. Dodge, *Do social information-processing patterns mediate aggressive behavior?* (2003)

E. Fehr, On the Economics and Biology of Trust, Presidential address at the 2008 meeting of the European Economic Association, *Journal of the European Economic Association*, **7 (2-3)**, 235-266 (2009).

N. Fortin, P. Oreopoulos, S. Phipps, Leaving Boys Behind: Gender Disparities in High Academic Achievement, Working Paper Department of Economics UBC (2011).

A. P. Goldstein, R. P. Sprafkin, N. J. Gershaw, P. Klein, The adolescent: Social skills training through structured learning. In G. Cartledge & J. F. Milburn (Eds.), *Teaching social skills to children: Innovative approaches* (Pergamon Press, New York, 1980) pp. 249-277.

J. Heckman and Y. Rubinstein, The Importance of non-cognitive skills : lessons from the GED testing program, *American Economic Review*, **91**, 145-149 (2001).

J.J. Heckman, S. H. Moon, R. Pinto, P. A. Savelyev, A. Q. Yavitz, Analyzing social experiments as implemented: A reexamination of the evidence from the High Scope Perry Preschool. *Quantitative Economics* **1(1)**, 1-46 (2010)

J.J Heckman, S.H. Moon, R. Pinto, P. A. Savelyev, A. Yavitz, The rate of return to the HighScope Perry Preschool Program. *Journal of Public Economics* **94(1)**, 114-128 (2010).

J.J. Heckman, L. Malofeeva, R. Pinto, P. Savelyev, Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review*, 2012.

S. Heller, H. Pollack, R. Ander and J. Ludwig, Improving social-cognitive skills among disadvantaged youth: A randomized field experiment, Working Paper Chicago (2013).

B. Jacob, Where the Boys Aren't: Non-cognitive Skills, Returns to School and the Gender Gap in Higher Education. *Economics of Education Review*, **21(6)**, 589–98 (2002).

C. Jencks, *Who gets ahead? The determinants of Economic Success in America*,(Basic Book, New York, 1979).

C. F. Jesness, R. F. Wedge, *Classifying offenders: The Jesness inventory classification system* (California Department of Youth Authority, Program Research and Review Division, 1983).

L. Katz, J. Ludwig, G. Duncan, L. Gennetian, R. Kessler, J. Kling, and L. Sanbonmatsu, Long-Term Neighborhood Effects on Low-Income Families: Evidence from Moving to Opportunity," American Economic Review, 103, 226-31 (2013).

L. Katz, J. R. Kling and J. B. Liebman, Moving to Opportunity in Boston: Early Results of a Randomized Mobility Experiment, Quarterly Journal of Economics, 116, 607-654 (2001)

J.R Kling, J. B. Liebman and L. F. Katz, Experimental Analysis of Neighborhood Effects, Econometrica, 75(1), 83-119 (2007)

P. W. Kettlewell, D. F. Kaush, The generalization of the effects of a cognitive behavioral treatment program for aggressive children. *Journal of Abnormal Child Psychology* **11**, 101-114 (1983).

M. Kovacs, *The Children's Depression Inventory: A self-rated depression scale for school-aged youngsters*. (University of Pittsburgh School of Medicine, Department of Psychiatry, Western Psychiatric Institute and Clinic, 1983).

A. Krueger, Experimental Estimates of Education Production Functions. *Quarterly Journal of Economics*, 114, 497-532 (1999).

E. Lacourse, S. Cote, D. S. Nagin, F. Vitaro, M. Brendgen, R. E. Tremblay, A longitudinal-experimental approach to testing theories of antisocial behavior development. Development and Psychopathology **14,** 909-924 (2002).

Lipsey, M., "The primary factors that characterize effective interventions with juvenile offenders: A meta-analytic review," Victims and Offenders, 4 (2009), 124-147.

Lipsey, M, and Cullen, F. "The Effectiveness of Correctional Rehabilitation: A Review of Systematic Reviews," Annual Review of Law and Social Science, 3 (2007).

J. Ludwig, D.L. Miller, Does Head Start improve children's life chances? Evidence from a regression-discontinuity design. *Quarterly Journal of Economics* **122(1)**, 159-208 (2007).

M. Maric, R.W. Wiers, P. J. M. Prins, Ten ways to improve the use of statistical mediation analysis in the practice of child and adolescent treatment research, *Clin Child Fam Psychol Rev* **15,** 177-191 (2012).

D.P. MacKinnon, C.M. Lockwood J. Williams, Confidence limits for the indirect effect: distribution of the product and resampling methods, *Multivariate Behavioral Research* **39(1)** 99-128 (2004).

H.W. Marsh, A multidimensional, hierarchical model of self-concept: Theoretical and empirical justification. *Educational Psychology Review* **2(2),** 77-172 (1990).

D. Meichenbaum, *Cognitive-behavior modification: An integrative approach*. (Plenum Press, New York, 1977).

L. Michelson, D.Sugai, R.Wood, A.E Kazdin, *Social skills assessment and training with children*. (Plenum Press, New York, 1983).

G.R. Patterson, *Coercive family process*. Eugene, OR: Castalia. (1982).

G.R. Patterson, J. B. Reid, R. R. Jones, R.W. Conger, *Families with aggressive children Volume 1 : A social learning approach to family intervention* (Castalia, Eugene, OR, 1975).

R. Putnam, *Making Democracy Work*. Princeton, NJ: Princeton University Press (1993)

R. Putnam *Bowling Alone: The Collapse and Revival of American Community*. New York: Simon & Schuster (2000).

M. Rosenberg, Rosenberg self-esteem scale (RSE). *Acceptance and Commitment Therapy. Measures Package* **61** (1965).

B. H. Schneider, B. M. Byrne, Individualizing social skills training for behavior disordered children. *Journal of Consulting and Clinical Psychology* **55,** 444-455 (1987).

L.J Schweinhar , J. Montie, Z. Xiang, W. S. Barnett, C. R. Bel_eld, and M. Nores (2005) "Lifetime Effects: The High/Scope Perry Preschool Study Through Age 40". Ypsilanti, MI: High/Scope Press.

C. Segal, Misbehavior, Education and Labor Market Outcomes, *Journal of European Economic Association*, in press (2013)

P. Szatmari, D. Offord, M. Boyle, Ontario Child Health Study: Prevalence of attention deficit disorder with hyperactivity. *Journal of Child Psychology and Psychiatry*. **30(2)**: 219-230 (1989).

R. Tremblay, The development of aggressive behavior during childhood: what have we learned during the past century. *International Journal of Behavioral Development*, **24**, 129-141 (2000).

R. Tremblay, F. Vitaro, L. Bertrand, M. Le Blanc, H. Beauchesne, H. Boileau, H., Parent and child training to prevent early onset of delinquency: The Montreal longitudinal experimental study. In J. McCord & R. E. Tremblay (Eds.), *Preventing antisocial behavior: Intervention from birth through adolescence* (Guilford Press, New York, 1992) pp. 117-138.

R. E. Tremblay, F. Vitaro, C. Gagnon, A Prosocial Scale for the Preschool Behavior Questionnaire: Concurrent and Predictive Correlates. *International Journal of Behavioral Development* **15(2)** 227-245 (1992).

R. J. Tremblay, E. McCord, J. Boileau, P. Charlebois, C. Gagnon, and M. LeBlanc, Can disruptive boys be helped to become competent? *Psychiatry*, **54(2)** 2, 148-161 (1991).

R. E. Tremblay, L. Desmarais-Gervais, C. Gagnon, P. Charlebois, The Preschool Behavior Questionnaire: Stability of its Factor Structure Between Cultures, Sexes, Ages and Socieconomic Classes. *International Journal of Behavioral Development* **10(4)** 467-484 (1987).

F. Vitaro, D. Larocque, M. Janosz, R. Tremblay, Negative social experiences and dropping out of school. *Educational Psychology* **21**, 401-415 (2001).

T.J Wansbeek, and E. Meijer. *Measurement error and latent variables in econometrics*. Vol. 37. (Elsevier, Amsterdam, 2000.

B. Weiss, K. A. Dodge, J. E. Bates, G. S. Pettit, Some consequences of early harsh discipline: Child aggression and a maladaptive social information processing style. *Child Development* **63**, 1321-1335 (1992).

APPENDIX A. MAIN TABLES AND FIGURES

Figure 1. Illustration of the Experimental Design of the MLES

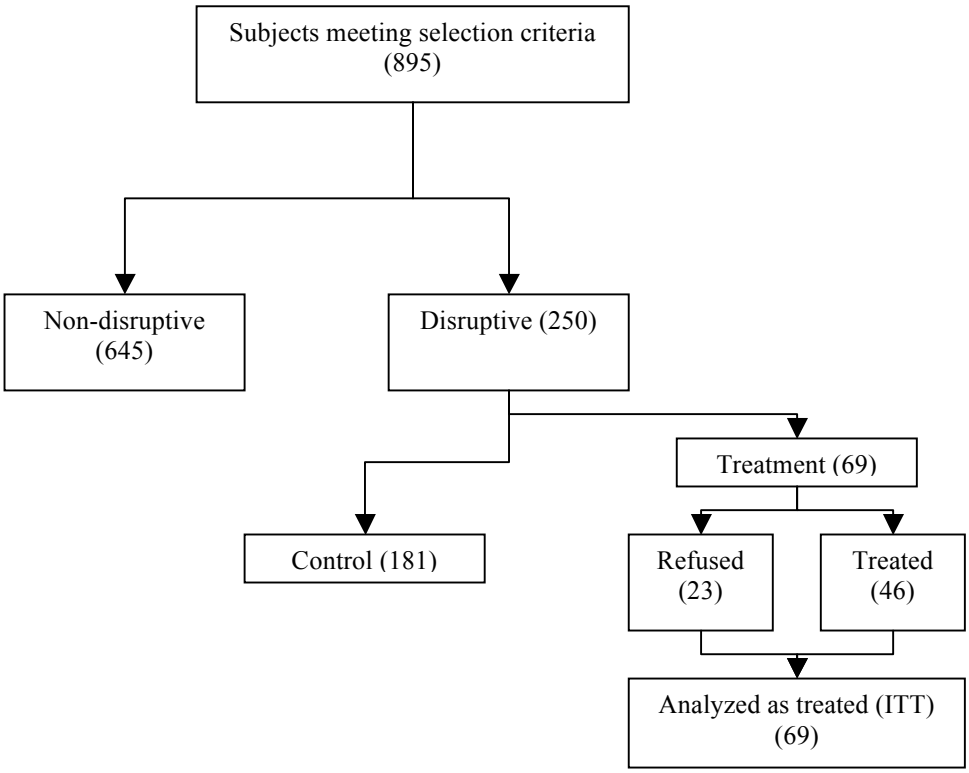


Figure 2. Timeline of the Experimental Study

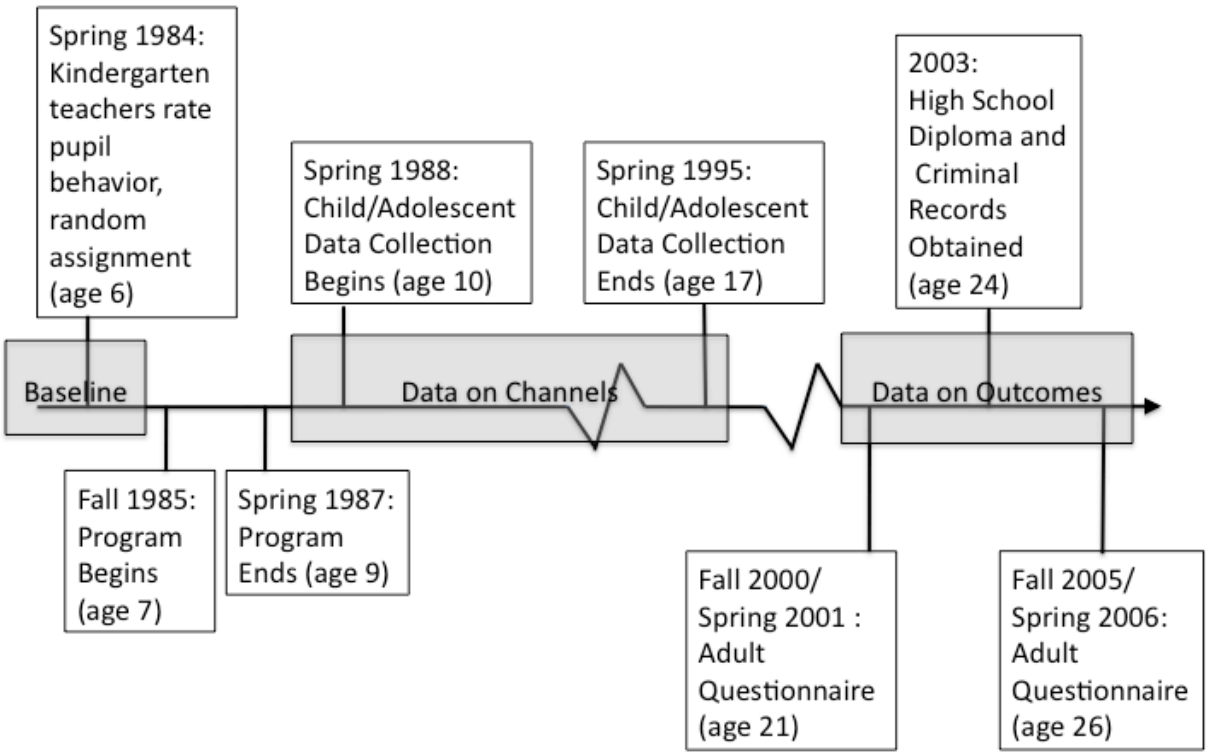
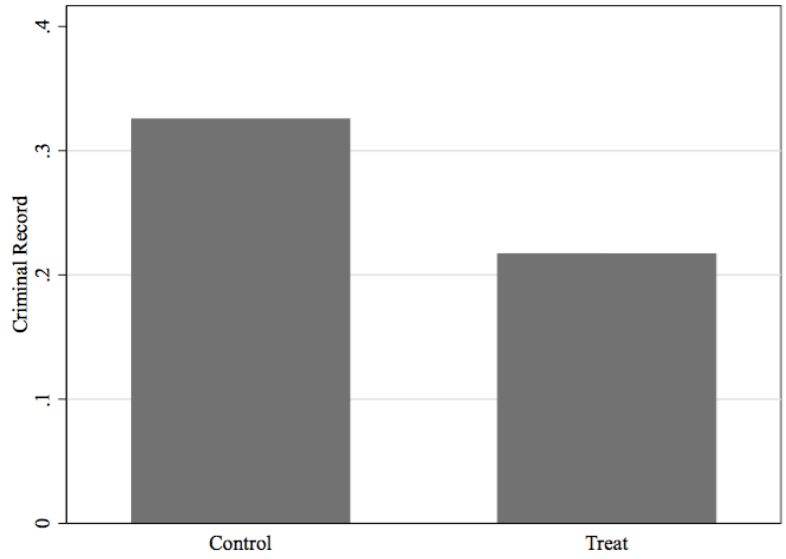
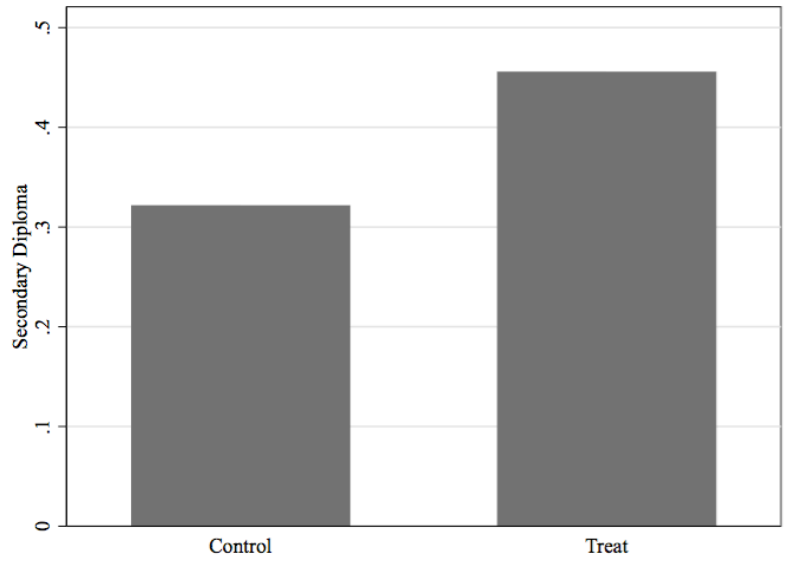


Figure 3. Percent of subjects with any criminal record



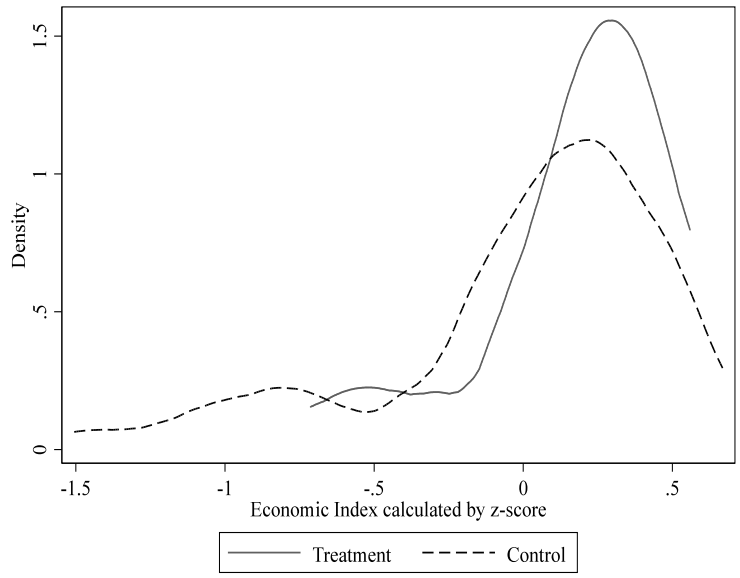
Two-sided p-value of difference between treatment and control groups = 0.09. Criminal record is a dummy equal to 1 if the subject ever had a criminal record. (Administrative data).

Figure 4. Rate of obtaining a Secondary School Diploma



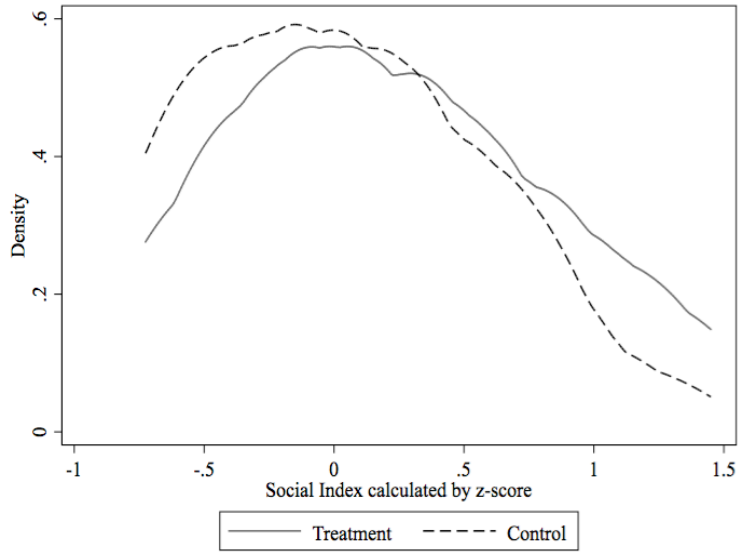
Secondary School is a dummy equal to 1 if the subject graduated from secondary school. P-value of difference between treatment and control groups = 0.05. (Administrative data)

Figure 5. Distributions of the Economic Index, by Treatment and Control (KS test p-value=0.002)



Economic Index is the unweighted average of the z-scores of an employment dummy at age 26, a full-time employment dummy at age 26, hourly wage, the percent of years from 17-26 occupied full time in work and/or study, whether subject was in school or working at age 21, and the percent of years from 17-26 receiving government transfers. (All data in the economic index are self-reported.)

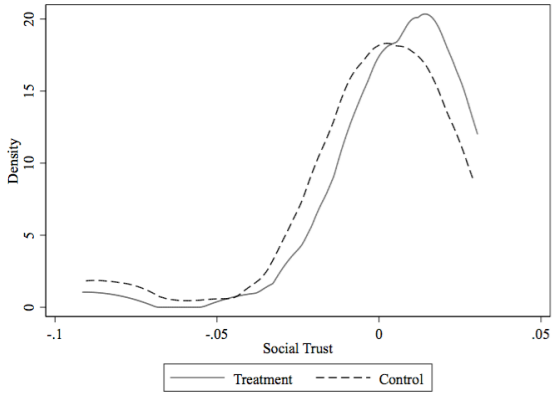
Figure 6. Distribution of Social Capital Index (KS test p-value=0.10)



Social Index is the unweighted average of the z-scores for whether the person voted in the last election, is the member of an organized group (for example, community groups or sport clubs) ,and volunteers (self-reported data at ages 21 and 26).

Figures 7a and 7b. Distribution of Social Trust and School Performance Channels

p-value=0.02



p-value=0.01

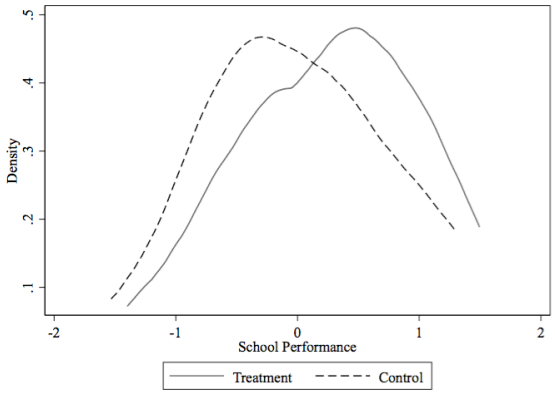
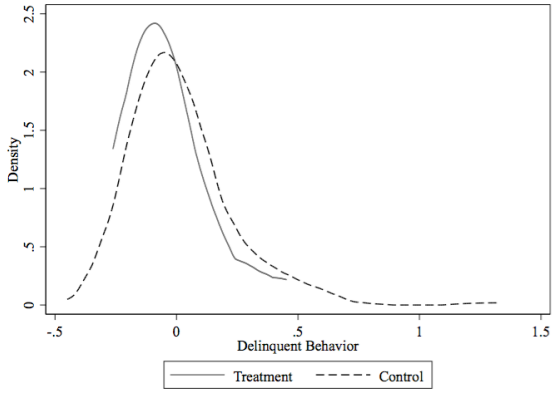


Figure 8a and 8b. Distribution of Delinquent Behavior and Hyperactivity Channels

p-value=0.02



p-value=0.05

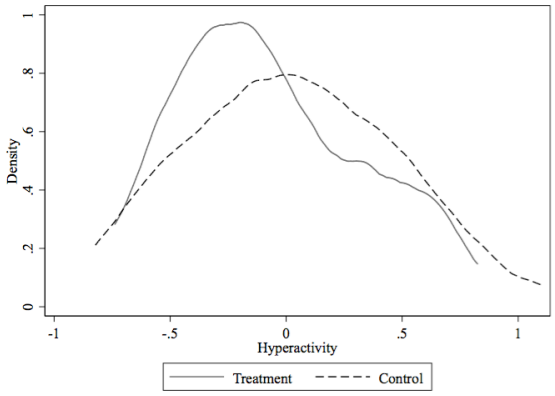


Figure 9a and 9b. Distribution of Self-Esteem and Altruism Behaviors

p-value=0.19 p-value=0.19

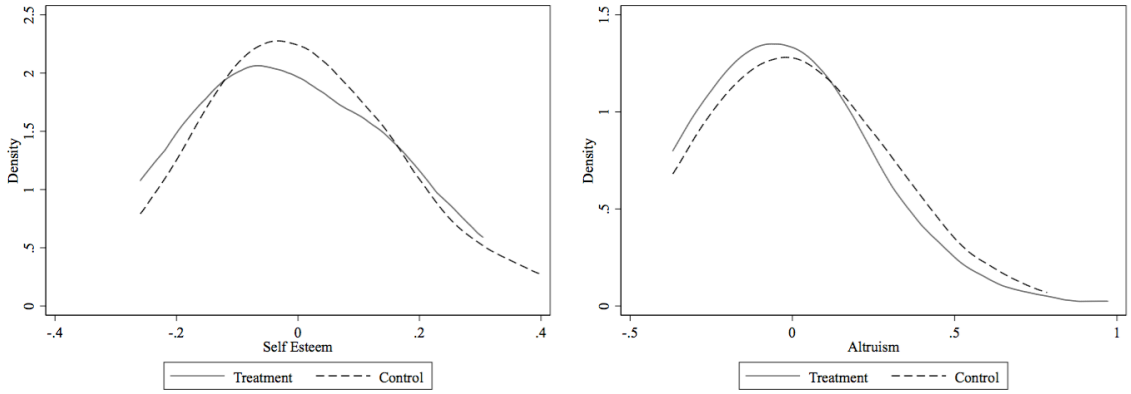


Figure 10a and 10b. Distribution of School Attitude and Gregarious Behavior Channels

p-value=0.15 p-value = 0.24

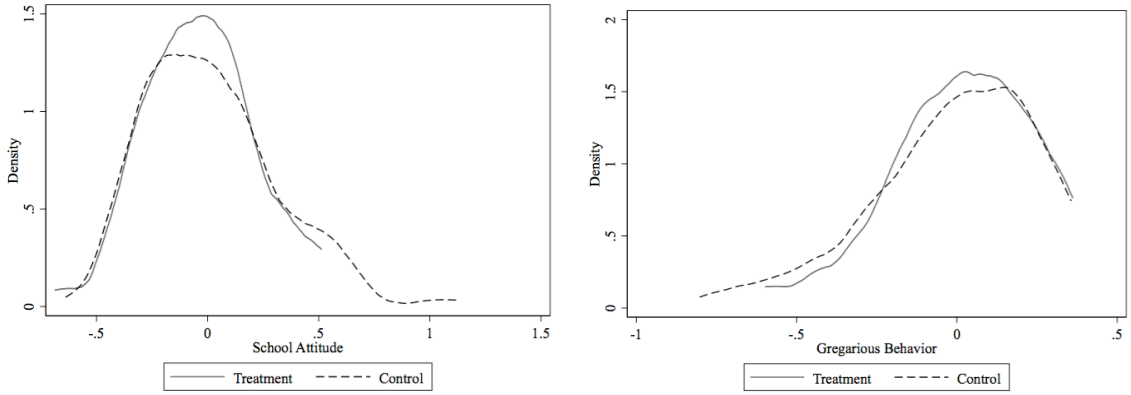
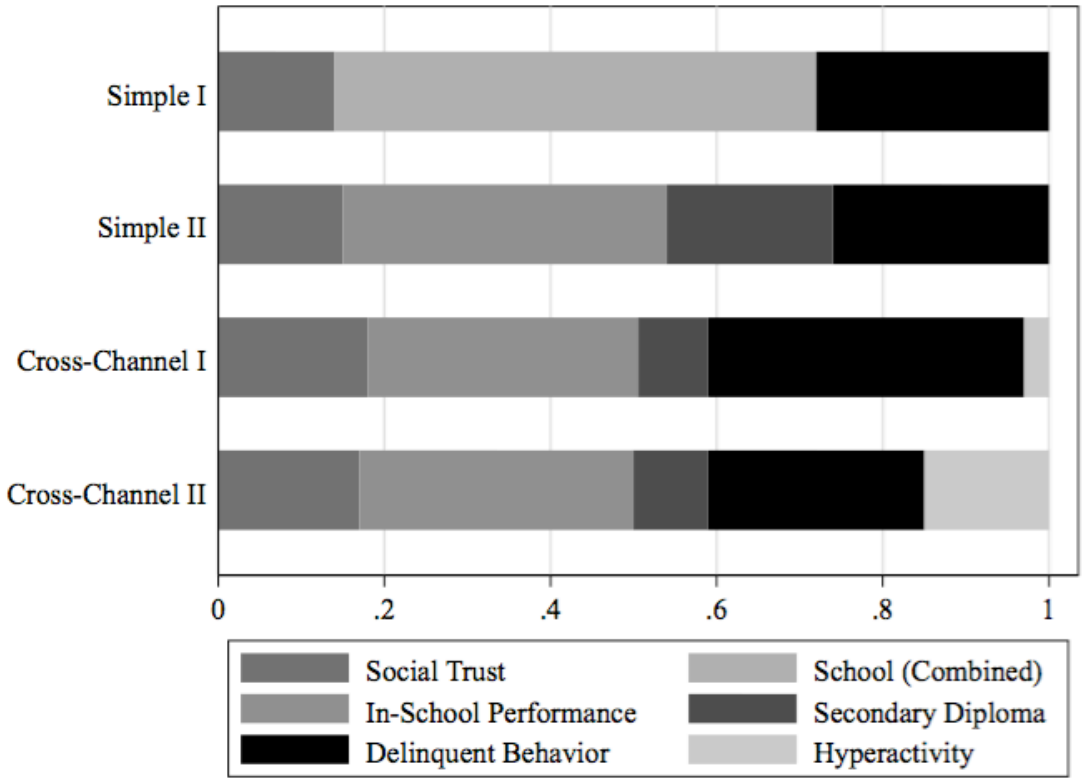
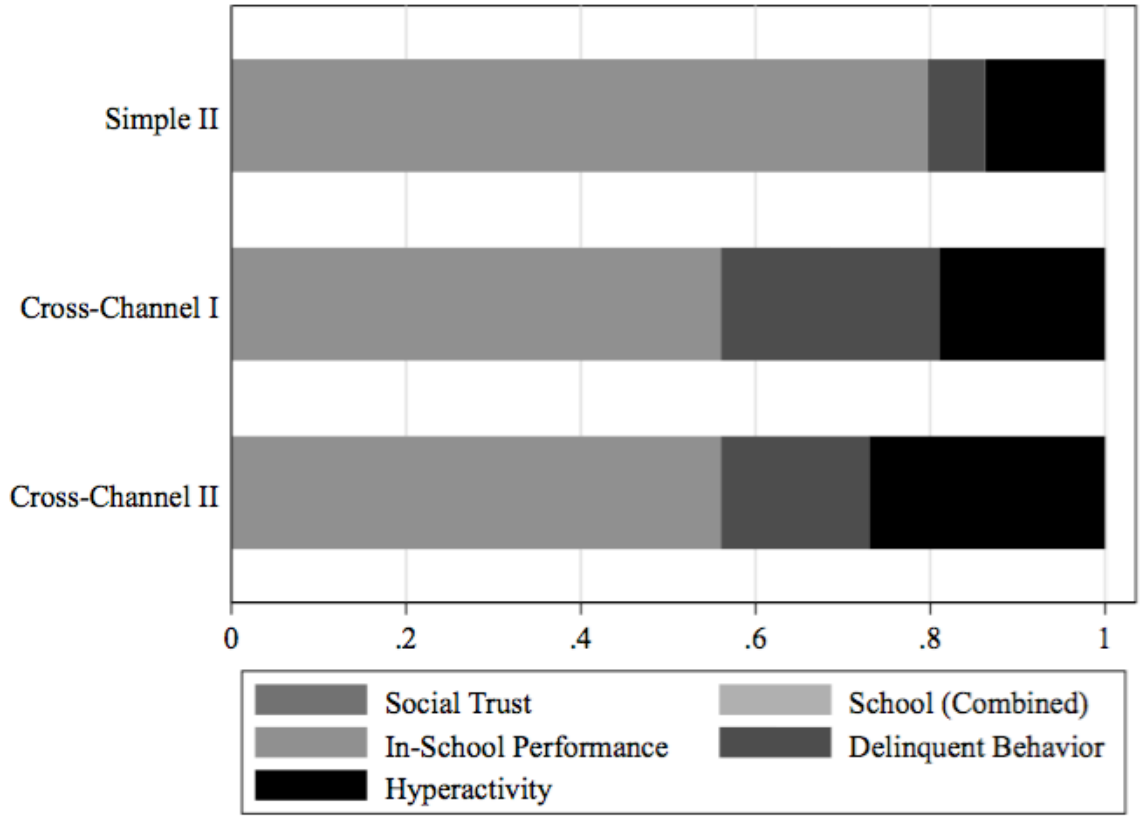


Figure 11. Decomposition of the explained treatment effects on Criminal Record



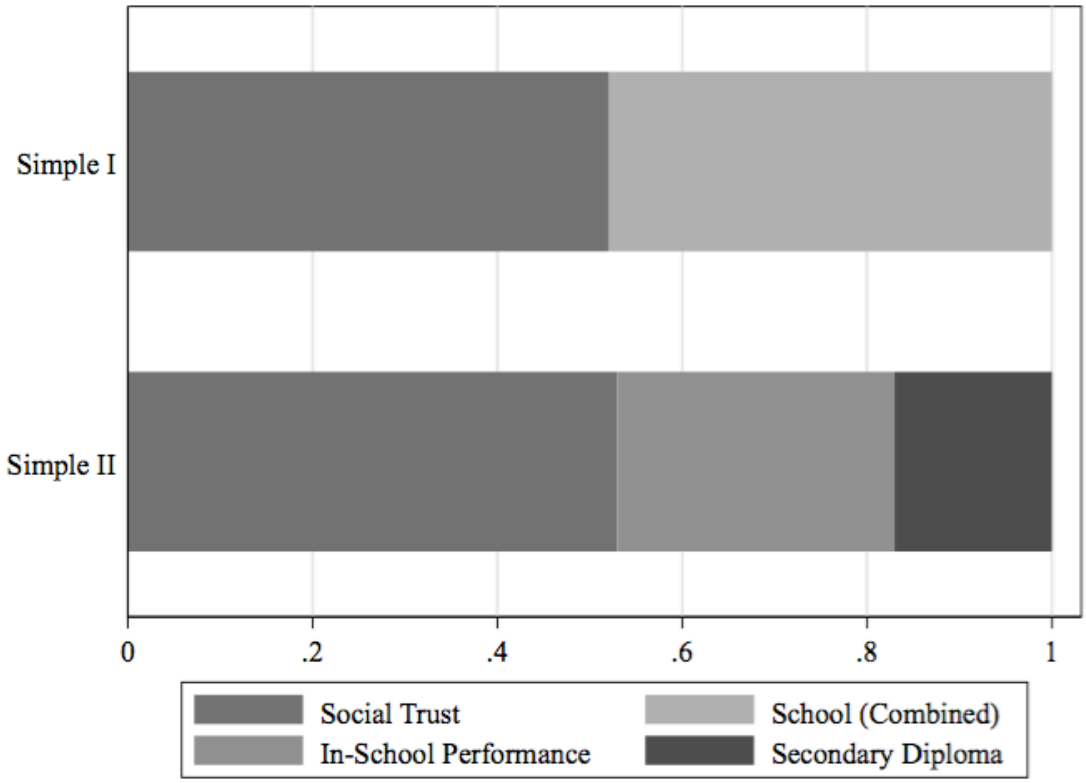
Each bar represents the explained treatment effect, normalized to 1, and the decomposition represents the percent of the explained treatment effect that is explained by changes in the channels. SimpleI is the specification that does not allow any cross-channel effects and combines all school variables into one channel. Simple II is the specification that does not allow any cross-channel effects, and distinguishes between school performance (class assignment, grade, and dropout) and Secondary Diploma. Cross-Channel I is the same as Simple II, but allows for behavioral channels (Social Trust, Delinquent Behavior, and Hyperactivity) to explain the school channels. Cross-Channel II is the same as Cross-Channel I, but allows Social Trust and Hyperactivity to explain Delinquent Behavior. Negative contributions are omitted. The channels explain the entire impact of treatment

Figure 12. Decomposition of the explained treatment effects on Secondary Diploma



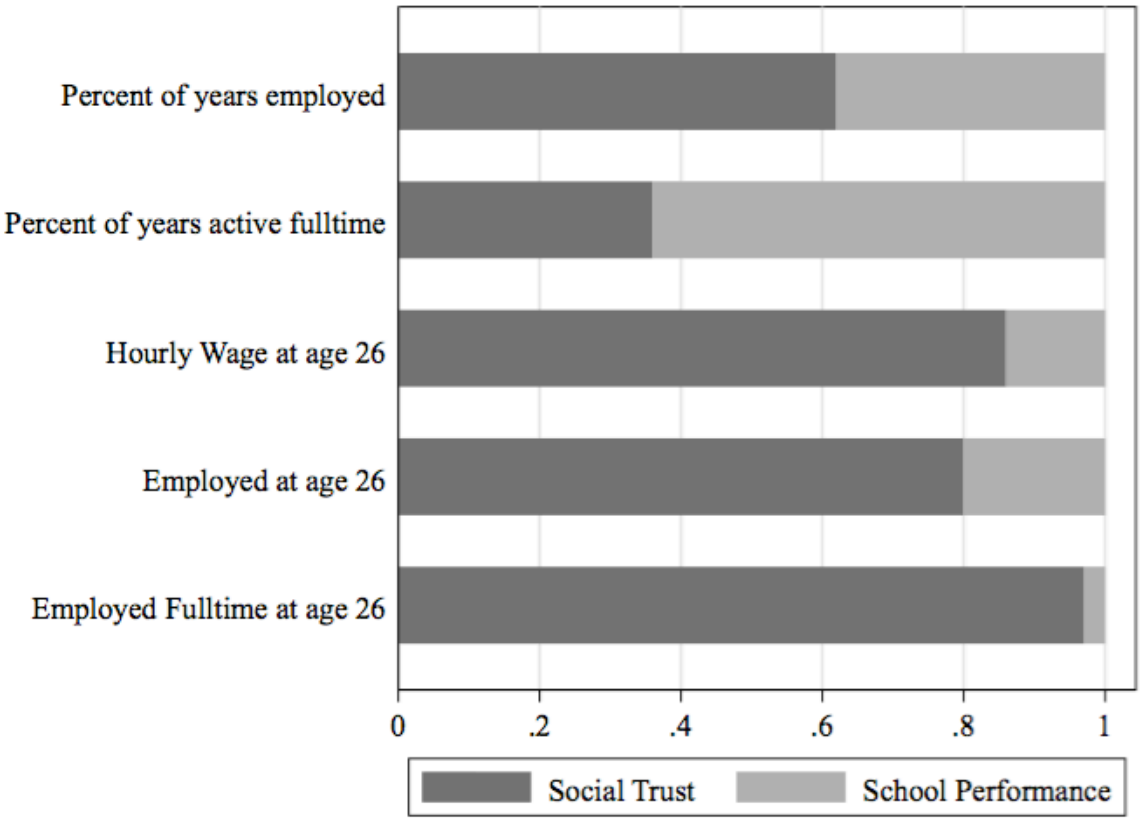
Each bar represents the explained treatment effect, normalized to 1, and the decomposition represents the percent of the explained treatment effect that is explained by changes in the channels. Simple II is the specification that does not allow any cross-channel effects. Cross-Channel I is the same as Simple, but allows for behavioral channels (Social Trust, Delinquent Behavior, and Hyperactivity) to explain School Performance. Cross-Channel II is the same as Cross-Channel I, but allows Social Trust and Hyperactivity to explain Delinquent Behavior. Negative contributions are omitted. The channels explain 57% of the impact of treatment.

Figure 13. Decomposition of the explained treatment effects on Economic Index



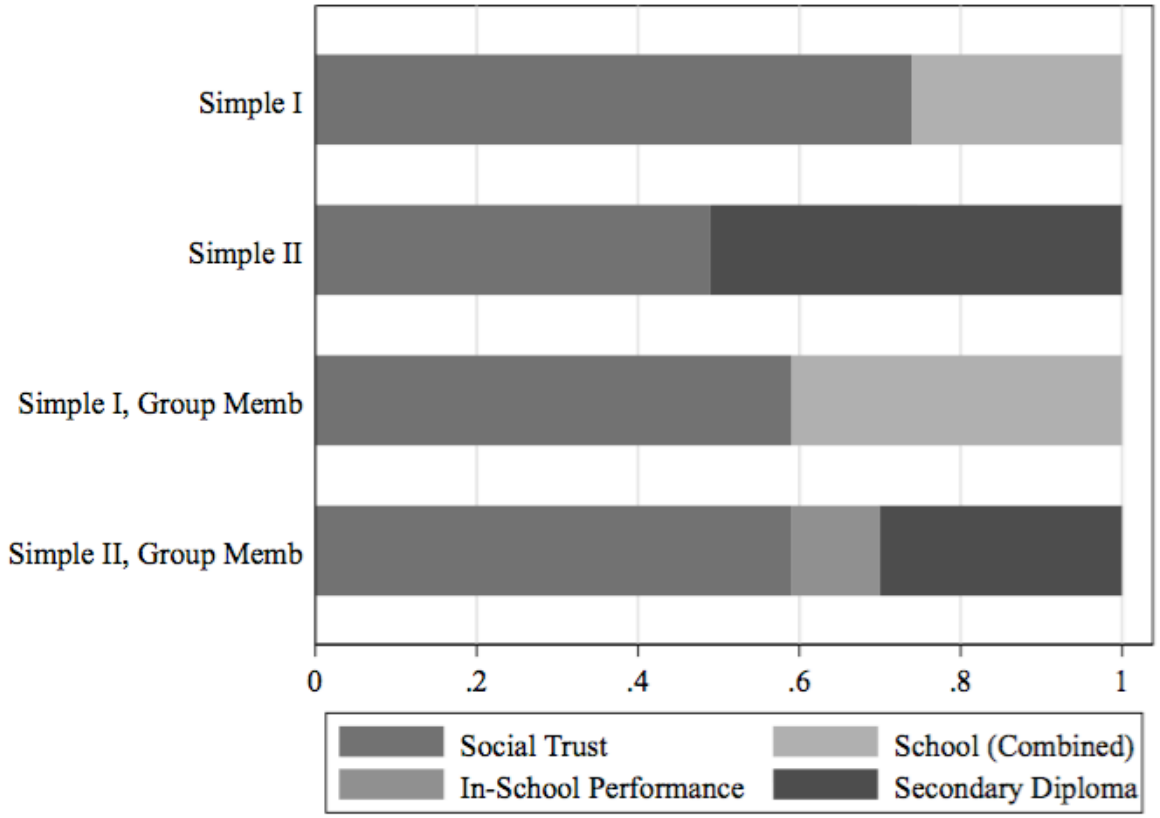
Each bar represents the explained treatment effect, normalized to 1, and the decomposition represents the percent of the explained treatment effect that is explained by changes in the channels. Simple II is the specification that does not allow any cross-channel effects and combines all school variables into one channel. Simple II is the specification that does not allow any cross-channel effects, and distinguishes between school performance (class assignment, grade, and dropout) and Secondary Diploma. Hyperactivity and Delinquent Behavior are not significantly related to treatment in the sample for which economic data are available and are excluded from the estimation. Cross-channel specifications do not substantially change the results. Negative contributions are omitted. The channels explain 44% of the treatment effect.

Figure 14. Decomposition of the explained treatment effects on selected Economic Index Components (Specification Simple I)



Each bar represents the explained treatment effect, normalized to 1, and the decomposition represents the percent of the explained treatment effect that is explained by changes in the channels. Negative contributions are omitted.

Figure 15. Decomposition of the explained treatment effects on Social Capital Index



Each bar represents the explained treatment effect, normalized to 1, and the decomposition represents the percent of the explained treatment effect that is explained by changes in the channels. Simple I is the specification that does not allow any cross-channel effects and combines all school variables into one channel. Simple II is the specification that does not allow any cross-channel effects, and distinguishes between school performance (class assignment, grade, and dropout) and Secondary Diploma. Group Memb indicates that the outcome used is the Group Membership (at age 21 or 26) variable. Hyperactivity and Delinquent Behavior are not significantly related to treatment in the sample for which social capital data are available and are excluded from the estimation. Cross-channel specifications do not substantially change the results. Negative contributions are omitted.

Table 1. Outcome Variable Summary Statistics

	Non-disruptive population		Disruptive				
			Treatment			Control	
	mean	N	mean	sd	N	mean	Sd
Crime							
Has a criminal record at age 23	0.15	69	0.22	0.42	181	0.32	0.47
Record for Violent Crime	0.04	69	0.09	0.29	181	0.11	0.31
Record for Non-violent Crime	0.13	69	0.19	0.40	181	0.28	0.45
Number of crimes	0.68	69	1.13	2.95	181	2.15	5.79
Education							
Completed Secondary School	0.59	68	0.46	0.50	174	0.32	0.47
Economic outcomes							
Economic Index	0.12	43	0.18	0.30	116	0	0.48
% years 17-26 fulltime in work or study	0.80	39	0.81	0.23	113	0.73	0.27
% of years 17-26 working part or fulltime	0.71	39	0.78	0.21	113	0.69	0.28
% years 17-26 receiving Gov't transfer	0.07	39	0.12	0.19	113	0.14	0.24
Occupied by work or study 2001	0.96	40	0.90	0.30	108	0.90	0.30
Occupied by work or study fulltime 2001	0.91	40	0.90	0.30	108	0.85	0.36
Employed in 2006	0.90	25	0.96	0.20	77	0.86	0.35
Employed fulltime in 2006	0.85	25	0.96	0.20	77	0.82	0.39
Hourly wage in 2006	17.64	25	16.5	5.85	73	14.98	7.87
Social outcomes							
Social Index	0.12	43	0.22	0.62	116	0	0.54
Member of a group in 2006	0.45	25	0.52	0.51	77	0.31	0.47
Member of a group in 2001	0.28	40	0.33	0.47	108	0.19	0.40
Voted in 2001	0.55	40	0.48	0.51	107	0.49	0.50
Volunteering / Community service in 2001	0.30	40	0.45	0.50	108	0.38	0.49

Education data from the Ministry of Education. Crime data from the Ministry of Public Security of Quebec. Economic and Social outcomes from adult questionnaires in 2001 and 2006.

Table 2. Baseline Characteristics and Randomization Check

	Non-disruptive population	Disruptive							
		Control			Treatment			Difference	
	mean	mean	N	sd	mean	N	sd	C-T	p-value
Age	5.996	6.027	181	0.296	5.973	69	0.287	0.054	0.1997
Attended Pre-school	0.159	0.209	181	0.408	0.188	69	0.393	0.021	0.707
Birth order	1.661	1.558	181	0.819	1.5	68	0.723	0.058	0.6079
Number brothers	0.561	0.5	180	0.689	0.471	68	0.68	0.029	0.7637
Number sisters	0.528	0.333	180	0.518	0.574	68	0.676	-0.241	**0.0031**
Number half brothers	0.0419	0.0331	181	0.208	0	68	0	0.0331	0.1909
Number half sisters	0.0405	0.0387	181	0.323	0.0147	68	0.121	0.024	0.5514
Lives with both parents	0.592	0.525	181	0.501	0.464	69	0.502	0.061	0.3898
Age of mother	25.69	23.99	180	4.184	24.01	68	4.706	-0.020	0.9674
Age of father	28.67	26.9	161	5.336	28.28	56	5.332	-1.38	**0.0971**
Mother education	10.67	9.967	180	2.23	9.897	68	2.28	0.070	0.8276
Father education	10.81	9.694	160	2.452	9.933	60	2.428	-0.239	0.5182
Number of children in HH	1.140	0.972	181	0.903	1.074	68	0.798	-0.102	0.4176
Adversity index	0.299	0.429	181	0.237	0.43	68	0.268	-0.001	0.9623
Mother works	1.607	1.729	177	0.446	1.779	68	0.418	-0.050	0.4192
Mother fulltime	1.695	1.688	48	0.468	1.571	14	0.514	0.117	0.4277
Father Works	1.109	1.216	148	0.413	1.204	49	0.407	0.012	0.8582
Father fulltime	1.966	1.939	115	0.24	1.921	38	0.273	0.018	0.6982
Mother prestige	39.35	36.03	161	11.02	33.16	60	10.13	2.87	**0.0794**
Father prestige	40.74	35.19	156	9.581	35.22	53	9.825	-0.030	0.9851
Initial Aggression	4	14.51	181	4.779	14.62	69	4.576	-0.110	0.8637
Initial Anxiety	2.651	3.552	181	2.731	4.261	69	2.816	-0.709	**0.0703**
Initial Opposition	1.628	5.624	181	2.186	5.812	69	1.927	-0.188	0.5326
Initial Prosociality	8.205	6.519	181	4.793	6.986	69	4.51	-0.467	0.4855
Initial Combativeness	0.818	3.525	181	1.594	3.478	69	1.54	0.047	0.8349
Initial Inattention	2.227	4.193	181	2.353	4.188	69	2.178	0.005	0.9879
Initial Hyperactivity	0.978	2.794	180	1.208	2.956	68	1.19	-0.162	0.3469
Initial Antisociality	0.845	0.994	181	1.113	1.206	68	1.229	-0.212	0.1957

Data from baseline data collection, 1984.

Table 3. Treatment Impact on Main Outcomes

	(1)	(2)	(3)	(4)	(5)
	Economic Index (ML estimation)	Economic Index (z-score average)	Secondary Diploma	Criminal Record	Social Index (z-score average)
Treatment (0/1)	0.305**	0.211***	0.195**	-0.115*	0.222*
	(0.144)	(0.0764)	(0.0785)	(0.0650)	(0.113)
Observations	159	159	242	250	159
R-squared	0.170	0.130			0.127
Control Group Average	-0.07	0	0.32	0.33	0

Robust standard errors in parentheses

**** p<0.01, ** p<0.05, * p<0.1*

Columns 1,2 and 5 are OLS. Columns 3 and 4 report the marginal effects from logit regressions.
Controls for columns 1 and 4 are: age of father at subject's birth, prestige of mother's employment in 1984, number of sisters in 1984, father's job status in 1984, initial anxiety, pro-social and hyperactivity measures, dummies for imputation and, for column 1, dummies indicating whether data was missing at age 21. Controls for columns 2 and 3 are: age of father at subject's birth, prestige of mother's employment in 1984, initial anxiety measure, number of sisters in 1984, and dummies for imputation.
Economic Index composed of an employment dummy at age 26, a full-time employment dummy at age 26, hourly wage, the percent of years from 17-26 occupied full time in work and/or study, whether subject was in school or working at age 21, and the percent of years from 17-26 receiving government transfers. (All data in the economic index are self-reported.) Estimations for an ML-estimated variable and a simple average of the z-scores are both presented.

Secondary School is a dummy equal to 1 if the subject graduated from secondary school.(Administrative data)

Criminal record is a dummy equal to 1 if the subject ever had a criminal record. (Administrative data).
Social Capital Index is the unweighted average of the z-scores for whether the person voted in the last election, is the member of an organized group (for example, community groups or sport clubs),and volunteers (self-reported data at age 21 and 26).

Table 4. Description of Channels

Channel	Number of Components	Cronbach Alpha	N	One-sided p-value from t-test by treatment	p-value from KS test
	(1)	(2)	(3)	(4)	(5)
Social Trust	6	0.67	223	**0.02**	**0.04**
School Performance	6	0.78	250	**0.00**	**0.01**
Delinquent Behavior	37	0.94	250	**0.02**	**0.01**
Hyperactivity	6	0.88	248	**0.05**	0.11
Gregarious	14	0.82	241	0.24	0.64
Self Esteem	41	0.91	237	0.19	0.26
School Attitude	16	0.88	239	0.15	0.63
Altruism	10	0.94	248	0.19	0.18

Raw values are used in tests, and all significant p-values remain significant with inclusion of controls One-sided p-values from t-tests, two-sided p-values for Kolmorogorov-Smirnov tests, and one-sided p-values for OLS coefficients.

Table 5. Correlation between channels

	Social Trust	School Performance	Antisocial	Hyperactivity	Gregarious	Self Esteem	School Attitude	Altruism
Social Trust	1							
School Performance	0.1172	1						
Antisocial	0.0172	-0.4348	1					
Hyperactivity	0.0946	-0.3418	0.4782	1				
Gregarious	0.0765	-0.0685	0.1574	0.1121	1			
Self Esteem	-0.0691	-0.2483	0.1572	0.13	-0.194	1		
School Attitude	0.0076	-0.3775	0.413	0.2706	-0.1137	0.2266	1	
Altruism	-0.0547	-0.1053	0.0253	0.094	0.1213	0.1331	-0.0925	1

Table 6. Decomposition of Explained Treatment Effect on Criminal Behavior

	Simple I		Simple II		Cross-Channel I		Cross-Channel II	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	% of explained treatment effect	one-sided p-value	% of explained treatment effect	one-sided p-value	% of explained treatment effect	one-sided p-value	% of explained treatment effect	one-sided p-value
Delinquent Behavior	28%	*0.03*	26%	0.03	38%	*0.02*	29%	*0.06*
Social Trust	14%	0.12	15%	0.11	18%	*0.08*	19%	0.10
Hyperactivity					3%	0.32	15%	*0.09*
School (Combined)	58%	*0.01*						
School Performance			39%		33%	*0.04*	33%	*0.04*
Secondary Diploma			20%		9%	*0.17*	9%	0.17

Odd-numbered columns give the percent of the treatment effect that is explained by the change in the channel. Even-numbered columns report the one-sided p-value of the indirect effect from bootstrapping with 1000 draws. Zero and negative values omitted.

Table 7. Decomposition of Explained Treatment Effect on Secondary Diploma

	Simple II		Cross-Channel I		Cross-Channel II	
	(1)	(2)	(3)	(4)	(5)	(6)
	% of explained treatment effect	one-sided p-value	% of explained treatment effect	one-sided p-value	% of explained treatment effect	one-sided p-value
Delinquent Behavior	7%	0.19	25%	*0.04*	17%	*0.09*
Social Trust						
Hyperactivity	14%	0.10	19%	*0.09*	27%	*0.08*
School Performance	80%	*0.00*	56%	*0.02*	56%	*0.02*

Odd-numbered columns give the percent of the treatment effect that is explained by the change in the channel. Even-numbered columns report the one-sided p-value of the indirect effect from bootstrapping with 1000 draws. Zero and negative values omitted.

Table 8. Decomposition of Explained Treatment Effect on Economic Index

	Simple I		Simple II	
	(1)	(2)	(3)	(4)
	% of explained treatment effect	one-sided p-value	% of explained treatment effect	one-sided p-value
Social Trust	52%	*0.086*	53%	*0.090*
School (Combined)	48%	*0.093*		
School Performance			30%	0.187
Secondary Diploma			17%	0.267

Odd-numbered columns give the percent of the treatment effect that is explained by the change in the channel. Even-numbered columns report the one-sided p-value of the indirect effect from bootstrapping with 1000 draws. Zero and negative values omitted.

Table 9. Decomposition of Explained Treatment Effect on Social Capital Index

	Simple I		Simple II		Simple I (Group Membership)		Simple II (Group Membership	
	(1)	(2)	(3)	(4)	(5)	(6)	(5)	(6)
	% of explained treatment effect	one-sided p-value	% of explained treatment effect	one-sided p-value	% of explained treatment effect	one-sided p-value	% of explained treatment effect	one-sided p-value
Social Trust	72%	0.220	49%	0.200	59%	0.100	59%	0.100
School (Combined)	26%	0.420			41%	0.211		
School Performance							11%	0.402
Secondary Diploma			51%	0.260			30%	0.254

Odd-numbered columns give the percent of the treatment effect that is explained by the change in the channel. Even-numbered columns report the one-sided p-value of the indirect effect from bootstrapping with 1000 draws. Zero and negative values omitted.

Table 10. Cost-Benefit Assumptions for Rate of Return Calculations

		Reported Data	Minimum Wage
Change in hourly income due to employment for those who have not completed secondary school	$I_E^{NS} - I_U^{NS}$	$ 12.92	$ 7.45
Change in probability of employment (percent of time from 17-27 reported working) due to program	$\Pi(E^T) - \Pi(E^C)$	10%	10%
Change in hourly income due to secondary school if employed	$I_E^S - I_E^{NS}$	$2.54	$2.54
Change in probability of completing secondary school due to program	$\Pi(S^T) - \Pi(S^C)$	19%	19%
Total change in hourly income due to program	Δ_I	$ 1.77	$ 1.23

Table 11. Regressions of treatment on components of Economic Index

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	
	Economic Index	Employed 2006	Employed Full Time 2006	Hourly Wage 2006	Occupied in Work or School 2006	Occupied Full Time in Work or School 2006	Percent of Years 17-26 Receiving Government Transfer	Percent of Years 17-26 Occupied in Work or School Full Time	Percent of Years 17-26 Working full or part time
Treatment (0/1)	0.221***	0.392**	0.501***	0.306	0.0345	0.171	-0.121	0.404*	0.367**
	(0.0705)	(0.172)	(0.175)	(0.213)	(0.209)	(0.195)	(0.171)	(0.225)	(0.176)
Constant	0.427*	0.369	0.206	0.508	1.135	0.669	-0.305	1.038*	0.524
	(0.236)	(0.669)	(0.689)	(0.768)	(0.754)	(0.681)	(0.603)	(0.614)	(0.567)
Observations	159	102	102	98	148	148	152	152	152
R-squared	0.209	0.323	0.269	0.199	0.084	0.099	0.209	0.294	0.208

Robust standard errors in parentheses

**** p<0.01, ** p<0.05, * p<0.1*

Economic Index is the unweighted average of the z-scores of the other dependent variables. Data are self-reported in 2001 and 2006.

Table 12. Regressions of treatment on components of Social Capital Index

	(1)	(2)	(3)	(4)	(5)
	Social Capital Index	Group membership 2006	Group membership 2001	Voted in last election	Engaged in Community Service
Treatment (0/1)	0.222*	0.392*	0.376	0.0695	0.145
	(0.113)	(0.228)	(0.245)	(0.205)	(0.226)
Constant	-0.302	-1.599*	-0.562	0.558	-0.340
	(0.361)	(0.929)	(0.690)	(0.699)	(0.706)
Observations	159	102	148	147	148
R-squared	0.126	0.267	0.080	0.109	0.070

Robust standard errors in parentheses

**** p<0.01, ** p<0.05, * p<0.1*

Social capital index is the unweighted average of the z-scores of the other variables. Data are self-reported from 2001 and 2006.

Table 13. Measurement of Social Trust Channel

Components	Source	Year	Instrument	N	Coefficient with latent variable	p-value of Coefficient with latent variable	Impact of treatment (p-value)
Assumes accident is intentional	Subject	1991-1995	Annual Quest	216	0.071	0.108	0.116
Gets advice from friend	Subject	1995	Annual Quest	173	-0.818	**0.000**	**0.048**
Discusses personal issues with friend	Subject	1995	Annual Quest	174	-0.840	**0.000**	**0.018**
Trusts friend	Subject	1995	Annual Quest	174	-0.806	**0.000**	**0.087**
Trusts others	Subject	1992	Jesness	191	-0.275	**0.000**	**0.032**
Trusts the Police	Subject	1990-1995	Annual Quest	222	-0.126	0.126	**0.016**

Contribution to latent variable from standardized maximum likelihood solution. P-values for contribution to latent variable are two-sided, all others are one-sided.

Table 14. Measurement of School Performance Channel

Components	Source	Year	Instrument	N	Coefficient with latent variable	p-value of Coefficient with latent variable	Impact of treatment (p-value)
Math grades	School	1992-1995	Annual grades	214	0.400	**0.000**	**0.099**
French grades	School	1992-1995	Annual grades	213	0.466	**0.000**	**0.021**
Special Education class	School	1992-1995	Administrative	249	-0.799	**0.000**	**0.022**
Held back a grade	School	1992-1995	Administrative	249	-0.648	**0.000**	**0.058**
Dropped out	Subject	1992-1995	Administrative	250	-0.588	**0.000**	**0.038**
Secondary Diploma	Ministry of Education	2003	Administrative	242	0.539	**0.000**	**0.026**

Contribution to latent variable from standardized maximum likelihood solution. P-values for contribution to latent variable are two-sided, all others are one-sided.

Table 15. Measurement of Antisocial Channel

Components	Source	Year	Instrument	N	Coefficient with latent variable	p-value of Coefficient with latent variable	Impact of treatment (p-value)
Stole bike	Subject	1988-1993	Annual Quest	239	0.75	**0.000**	**0.00**
False alarm	Subject	1988-1993	Annual Quest	239	0.74	**0.000**	**0.03**
Stole between 10 and 100	Subject	1988-1993	Annual Quest	230	0.76	**0.000**	**0.02**
Accused other of starting fight	Subject	1988-1993	Annual Quest	223	0.62	**0.000**	**0.02**
Snuck in without paying	Subject	1988-1993	Annual Quest	239	0.78	**0.000**	0.14
Vandalize school	Subject	1988-1993	Annual Quest	239	0.65	**0.000**	0.02
Fights	Subject	1988-1993	Annual Quest	230	0.68	**0.000**	**0.09**
Broke in and stole	Subject	1988-1993	Annual Quest	230	0.74	**0.000**	**0.00**
Threaten to beat	Subject	1988-1993	Annual Quest	230	0.68	**0.000**	**0.02**
Mistreat others to get wants	Subject	1988-1993	Annual Quest	223	0.70	**0.000**	**0.06**
Fistfight	Subject	1988-1993	Annual Quest	239	0.73	**0.000**	0.48
Used weapon	Subject	1988-1993	Annual Quest	239	0.74	**0.000**	**0.05**
Easy to anger	Subject	1988-1993	Annual Quest	223	0.76	**0.000**	0.11
Unprovoked hitting	Subject	1988-1993	Annual Quest	239	0.72	**0.000**	0.23
Destroyed antenna	Subject	1988-1993	Annual Quest	229	0.66	**0.000**	**0.05**
Damage other people's things	Subject	1988-1993	Annual Quest	230	0.67	**0.000**	0.31
Used force to dominate	Subject	1988-1993	Annual Quest	239	0.69	**0.000**	**0.07**
Threw stuff at people	Subject	1988-1993	Annual Quest	239	0.67	**0.000**	0.37
Criminal Record before age 17	Ministry of Justice	1995	Administrative Data	250	0.39	**0.000**	0.17
Fights with other children	Teacher	1988-1993	Annual Quest	248	0.51	**0.000**	0.11
Tells lies	Teacher	1988-1993	Annual Quest	248	0.45	**0.000**	**0.03**
Mistreats and intimidates others	Teacher	1988-1993	Annual Quest	248	0.53	**0.000**	**0.08**
Blames others	Teacher	1988-1993	Annual Quest	248	0.45	**0.000**	0.29
Hits/bites	Teacher	1988-1993	Annual Quest	248	0.47	**0.000**	0.20
Imposes on others	Teacher	1988-1993	Annual Quest	223	0.45	**0.000**	0.46
Reacts badly to bums	Teacher	1988-1993	Annual Quest	223	0.49	**0.000**	**0.09**
Encourages others to fight	Teacher	1988-1993	Annual Quest	223	0.52	**0.000**	0.15
Dominates with physical force	Teacher	1988-1993	Annual Quest	223	0.60	**0.000**	0.16
Accuses others	Teacher	1988-1993	Annual Quest	223	0.48	**0.000**	0.11
Mistreats and threatens others	Teacher	1988-1993	Annual Quest	223	0.54	**0.000**	0.24
Peer rating - reactive aggression	Peers	1990	Clasen	198	0.51	**0.000**	0.27
Peer rating - Proactive aggression	Peers	1990	Clasen	198	0.50	**0.000**	**0.01**
Peer rating - aggression	Peers	1988-1990	Clasen	239	0.50	**0.000**	**0.02**
Peer rating - physical aggression	Peers	1988-1990	Clasen	239	0.49	**0.000**	**0.05**
Peer rating - fighting	Peers	1988-1990	Clasen	239	0.50	**0.000**	**0.08**
Peer rating - lies	Peers	1988-1990	Clasen	239	0.46	**0.000**	**0.04**
Peer rating - Antisocial	Peers	1988-1990	Clasen	239	0.51	**0.000**	**0.07**

Contribution to latent variable from standardized maximum likelihood solution. P-values for contribution to latent variable are two-sided, all others are one-sided.

Table 16. Measurement of Self-Esteem Channel

Components	Source	Year	Instrument	N	Coefficient with latent variable	p-value of Coefficient with latent variable	Impact of treatment (p-value)
Feel like I don't belong	Subject	1988-1990, 1992	Jesness	235	0.55	**0.00**	0.31
Strange ideas in my head	Subject	1988-1990, 1992	Jesness	235	0.56	**0.00**	0.39
Others do things more easily	Subject	1988-1990, 1992	Jesness	235	0.48	**0.00**	0.30
Strange things happen to me	Subject	1988-1990, 1992	Jesness	234	0.57	**0.00**	0.39
I feel dizzy for no reason	Subject	1988-1990, 1992	Jesness	235	0.49	**0.00**	0.24
I would like to be elsewhere	Subject	1988-1990, 1992	Jesness	235	0.48	**0.00**	0.27
I don't find anything to do	Subject	1988-1990, 1992	Jesness	235	0.57	**0.00**	0.15
I often feel alone and sad	Subject	1988-1990, 1992	Jesness	235	0.59	**0.00**	0.35
I'm in trouble	Subject	1988-1990, 1992	Jesness	235	0.61	**0.00**	0.34
Others are happier than me	Subject	1988-1990, 1992	Jesness	235	0.61	**0.00**	0.34
Don't understand what I feel	Subject	1988-1990, 1992	Jesness	235	0.62	**0.00**	0.29
I'm worried	Subject	1988-1990, 1992	Jesness	235	0.59	**0.00**	0.45
It's too hard for me to decide things	Subject	1988-1990, 1992	Jesness	235	0.50	**0.00**	0.20
I feel alone even when I'm with people	Subject	1988-1990, 1992	Jesness	235	0.56	**0.00**	0.07
Hard to catch my breath	Subject	1988-1990, 1992	Jesness	235	0.48	**0.00**	0.13
I don't get a real chance	Subject	1988-1990, 1992	Jesness	235	0.56	**0.00**	0.24
Things don't seem real	Subject	1988-1990, 1992	Jesness	235	0.51	**0.00**	0.40
I feel like something doesn't work	Subject	1988-1990, 1992	Jesness	235	0.58	**0.00**	0.28
I feel like I'm worthless	Subject	1988, 1990, 1995	Rosenberg	226	0.52	**0.00**	0.38
I don't have much to be proud of	Subject	1988, 1990, 1995	Rosenberg	226	0.58	**0.00**	0.36
I feel useless	Subject	1988, 1990, 1995	Rosenberg	226	0.41	**0.00**	0.26
I feel like a failure	Subject	1988, 1990, 1995	Rosenberg	226	0.54	**0.00**	0.46
I'm not good at anything	Subject	1994	Marsh	177	-0.40	**0.00**	0.27
I can't do anything right	Subject	1994	Marsh	178	-0.36	**0.00**	0.10
I don't have much to be proud of	Subject	1994	Marsh	178	-0.44	**0.00**	0.11
I'm sad	Subject	1993-1995	Kovacs	200	0.22	**0.00**	0.34
Things work for me	Subject	1993-1995	Kovacs	203	-0.48	**0.00**	0.34
I fail at everything I try	Subject	1993-1995	Kovacs	203	0.45	**0.00**	0.32
Things don't amuse me	Subject	1993-1995	Kovacs	203	0.40	**0.00**	0.38
I'm an unpleasant person	Subject	1993-1995	Kovacs	200	-0.33	**0.00**	0.38
Unpleasant things happen to me	Subject	1993-1995	Kovacs	199	0.36	**0.00**	0.25
I like myself	Subject	1993-1995	Kovacs	203	-0.48	**0.00**	0.16
I want to cry	Subject	1993-1995	Kovacs	201	-0.29	**0.00**	0.43
Something bothers me	Subject	1993-1995	Kovacs	203	-0.41	**0.00**	0.27
I want to be with others	Subject	1993-1995	Kovacs	203	0.25	**0.00**	0.11
Trouble sleeping	Subject	1993-1995	Kovacs	203	-0.41	**0.00**	0.13
Tired	Subject	1993-1995	Kovacs	203	0.35	**0.00**	**0.06**
I feel alone	Subject	1993-1995	Kovacs	203	0.48	**0.00**	0.13
I have a lot of friends	Subject	1993-1995	Kovacs	203	0.47	**0.00**	0.11
Someone loves me	Subject	1993-1995	Kovacs	202	-0.48	**0.00**	0.11

Contribution to latent variable from standardized maximum likelihood solution. P-values for contribution to latent variable are two-sided, all others are one-sided.

Table 17. Measurement of Hyperactivity Channel

Components	Source	Year	Instrument	N	Coefficient with latent variable	p-value of Coefficient with latent variable	Impact of treatment (p-value)
Very agitated	Teacher	1988-1993	Social Behavior Questionnaire	248	0.87	**0.00**	0.11
Fidgets constantly	Teacher	1988-1993	Social Behavior Questionnaire	248	0.86	**0.00**	**0.10**
Disobedient	Teacher	1988-1993	Social Behavior Questionnaire	248	0.71	**0.00**	0.12
Easily distracted	Teacher	1988-1993	Social Behavior Questionnaire	248	0.54	**0.00**	0.24
Jumps from one activity to another	Teacher	1988-1993	Social Behavior Questionnaire	223	0.72	**0.00**	**0.03**
Does not listen to directions	Teacher	1988-1993	Social Behavior Questionnaire	223	0.72	**0.00**	0.13

Contribution to latent variable from standardized maximum likelihood solution. P-values for contribution to latent variable are two-sided, all others are one-sided.

Table 18. Measurement of School Attitude Channel

Components	Source	Year	Instrument	N	Coefficient with latent variable	p-value of Coefficient with latent variable	Impact of treatment (p-value)
Impolite to teacher	Subject	1988-1995	Annual Quest	239	0.50	**0.00**	0.13
Disturb Class on purpose	Subject	1988-1995	Annual Quest	239	0.38	**0.00**	0.11
Likes school	Subject	1988-1995	Annual Quest	239	0.54	**0.00**	0.12
Capability	Subject	1988-1995	Annual Quest	205	0.54	**0.00**	0.32
Comparative success	Subject	1988-1995	Annual Quest	239	0.48	**0.00**	0.40
Hours spent on homework	Subject	1988-1995	Annual Quest	239	0.31	**0.00**	0.49
Can talk to teacher	Subject	1988-1995	Annual Quest	206	0.28	**0.00**	0.44
Grades are important	Subject	1988-1995	Annual Quest	239	-0.35	**0.00**	0.15
Try hard at school	Subject	1988-1995	Annual Quest	239	-0.49	**0.00**	0.12
Bad grades	Subject	1994	Marsh	177	-0.84	**0.00**	0.15
Learn quickly	Subject	1994	Marsh	177	-0.82	**0.00**	**0.08**
Bad at school work	Subject	1994	Marsh	177	-0.78	**0.00**	**0.03**
Does well on exams	Subject	1994	Marsh	174	-0.89	**0.00**	0.35
Has difficulty at school	Subject	1994	Marsh	175	-0.81	**0.00**	0.12
Good at school	Subject	1994	Marsh	175	-0.90	**0.00**	0.21
School is too hard	Subject	1994	Marsh	175	-0.60	**0.00**	0.42

Table 19. Measurement of Gregarious Behavior Channel

Components	Source	Year	Instr.	N	Coeffi. with latent variable	p-value of Coeff. with latent variable	Impact of treat (pvalue)
Likes sports	Subject	1994	Marsh	177	0.15	**0.06**	0.47
Difficulty of having friends (boys)	Subject	1994	Marsh	177	0.49	**0.00**	0.30
Ease of having friends (girls)	Subject	1994	Marsh	177	0.67	**0.00**	0.25
Good at sports	Subject	1994	Marsh	178	0.25	**0.00**	0.37
Clumsy at sports	Subject	1994	Marsh	178	0.29	**0.00**	0.33
Popular with boys	Subject	1994	Marsh	178	0.43	**0.00**	**0.06**
Popular with girls	Subject	1994	Marsh	178	0.80	**0.00**	0.44
Good friends (girls)	Subject	1994	Marsh	178	0.82	**0.00**	0.29
Girls try to avoid	Subject	1994	Marsh	175	0.54	**0.00**	0.28
Ease of having friends (boys)	Subject	1994	Marsh	175	0.60	**0.00**	0.19
Lots of attention from girls	Subject	1994	Marsh	176	0.81	**0.00**	0.22
I like everyone I know	Subject	1988-1990, 1992	Jesness	235	0.30	**0.00**	**0.07**
Everyone I know likes me	Subject	1988-1990, 1992	Jesness	235	0.24	**0.00**	0.21
Time spent in sports	Subject	1988-1995	Annual Quest	239	0.22	**0.00**	0.27

Contribution to latent variable from standardized maximum likelihood solution. P-values for contribution to latent variable are two-sided, all others are one-sided.

Table 20. Measurement of Altruism Channel

Components	Source	Year	Instrument	N	Coefficient with latent variable	p-value of Coefficient with latent variable	Impact of treatment (p-value)
Stops fights	Teacher	1988-1993	Social Behavior Questionnaire	248	0.72	**0.00**	0.39
Invites others to join	Teacher	1988-1993	Social Behavior Questionnaire	247	0.82	**0.00**	0.11
Helps injured child	Teacher	1988-1993	Social Behavior Questionnaire	248	0.85	**0.00**	0.38
Spontaneously cleans up	Teacher	1988-1993	Social Behavior Questionnaire	248	0.81	**0.00**	0.34
Congratulates others	Teacher	1988-1993	Social Behavior Questionnaire	248	0.70	**0.00**	0.36
Shows sympathy	Teacher	1988-1993	Social Behavior Questionnaire	248	0.74	**0.00**	0.26
Offers to help others	Teacher	1988-1993	Social Behavior Questionnaire	248	0.77	**0.00**	0.16
Helps ill child	Teacher	1988-1993	Social Behavior Questionnaire	248	0.85	**0.00**	**0.09**
Comforts crying child	Teacher	1988-1993	Social Behavior Questionnaire	248	0.89	**0.00**	0.30
Volunteers to clean up	Teacher	1988-1993	Social Behavior Questionnaire	248	0.65	**0.00**	0.11

Contribution to latent variable from standardized maximum likelihood solution. P-values for contribution to latent variable are two-sided, all others are one-sided.

Table 21. Coefficients from Maximum Likelihood estimation, Simple I, Criminal Record

	(1)		(2)	
	Treatment -> Channel		Channel -> Criminal Record	
	Coef.	p-value	Coef.	p-value
Delinquent Behavior	-0.122	0.050	0.252	0.000
Social Trust	0.175	0.006	-0.090	0.133
Hyperactivity	-0.089	0.156	0.000	0.994
School (Combined)	0.205	0.001	-0.313	0.000
Treatment			0.008	0.894
Number of Obs		250		

Estimates from maximum likelihood estimation. Variables used are residuals from regressions with control variables. Including control variables directly in the estimation does not substantially change the result but increases computation time. P-values are two-sided.

Column (1) gives the individual impact of treatment on each variable listed on the left hand side column. Column (2) gives the impact of each channel and the treatment dummy, holding the others constant, on having a criminal record.

Table 22. Coefficients from Maximum Likelihood estimation, Simple II, Criminal Record

	(1)		(2)	
	Treatment -> Channel		Channel -> Criminal Record	
	Coef.	p-value	Coef.	p-value
Delinquent Behavior	-0.122	0.050	0.252	0.000
Social Trust	0.176	0.006	-0.097	0.107
Hyperactivity	-0.088	-1.400	-0.018	0.782
School Performance	0.198	0.001	-0.230	0.001
Secondary Diploma	0.165	0.007	-0.140	0.040
Treatment			0.012	0.837
Number of Obs		250		

Estimates from maximum likelihood estimation. Variables used are residuals from regressions with control variables. Including control variables directly in the estimation does not substantially change the result but increases computation time. P-values are two-sided.

Column (1) gives the individual impact of treatment on each variable listed on the left hand side column. Column (2) gives the impact of each channel and the treatment dummy, holding the others constant, on having a criminal record.

Table 23. Coefficients from Maximum Likelihood estimation, Complex I, Criminal Record

	(1) Treatment -> Channel		(2) Channel -> School Performance		(3) Channel -> Secondary Diploma		(4) Channel -> Criminal Record	
	Coef.	p-value	Coef.	p-value	Coef.	p-value	Coef.	p-value
Delinquent Behavior	-0.12	0.05	-0.33	0.00	-0.06	0.34	0.25	0
Social Trust	0.18	0.01	0.11	0.07	-0.08	0.14	-0.10	0.107
Hyperactivity	-0.09	-1.40	-0.12	0.07	-0.17	0.01	-0.02	0.782
School Performance					0.44	0.00	-0.23	0.001
Secondary Diploma							-0.14	0.04
Treatment			0.13	0.03	0.07	0.20	0.01	0.837
Number of Obs				250				

Estimates from maximum likelihood estimation. Variables used are residuals from regressions with control variables. Including control variables directly in the estimation does not substantially change the result but increases computation time. P-values are two-sided. P-values are two-sided.
Column (1) gives the individual impact of treatment on each variable listed on the left hand side column. Column (2) gives the impact of each behavioral channel and the treatment dummy on school performance, holding the others constant. Column (3) gives the impact of each behavioral channel, the treatment dummy, and school performance on obtaining a secondary diploma, holding the others constant. Column (4) gives the impact of each channel and the treatment dummy, holding the others constant, on having a criminal record.

Table 24. Coefficients from Maximum Likelihood estimation, Complex II, Criminal Record

	(1) Treatment -> Channel		(2) Channel -> Delinquent		(3) Channel -> School Performance		(4) Channel -> Secondary Diploma		(5) Channel -> Criminal Record	
	Coef.	p-value	Coef.	p-value	Coef.	p-value	Coef.	p-value	Coef.	p-value
Delinquent Behavior					-0.33	0.00	-0.06	0.34	0.25	0.00
Social Trust	0.18	0.01	0.01	0.87	0.11	0.07	-0.08	0.14	-0.10	0.11
Hyperactivity	-0.09	0.16	0.45	0.00	-0.12	0.08	-0.17	0.01	-0.02	0.79
School Performance							0.44	0.00	-0.23	0.001
Secondary Diploma									-0.14	0.04
Treatment			-0.08	0.14	0.13	0.03	0.07	0.20	0.01	0.837
Number of obs					250					

Estimates from maximum likelihood estimation. Variables used are residuals from regressions with control variables. Including control variables directly in the estimation does not substantially change the result but increases computation time.
Column (1) gives the individual impact of treatment on each variable listed on the left hand side column. Column (2) gives the impact of Hyperactivity and Social Trust on Delinquent Behavior, holding the other, and the treatment dummy, constant. Column (3) gives the impact of each behavioral channel and the treatment dummy on school performance, holding the others constant. Column (4) gives the impact of each behavioral channel, the treatment dummy, and school performance on obtaining a secondary diploma, holding the others constant. Column (5) gives the impact of each channel and the treatment dummy, holding the others constant, on having a criminal record.

Table 25. Coefficients from Maximum Likelihood estimation, Simple II, Secondary Diploma

	(1) Treatment -> Channel		(2) Channel ->Secondary Diploma	
	Coef.	p-value	Coef.	p-value
Delinquent Behavior	-0.122	0.050	-0.059	0.349
Social Trust	0.178	0.005	-0.087	0.136
Hyperactivity	-0.088	0.161	-0.171	0.005
School Performance	0.198	0.001	0.442	0.000
Treatment			0.071	0.193
Number of Obs		250		

Estimates from maximum likelihood estimation. Variables used are residuals from regressions with control variables. Including control variables directly in the estimation does not substantially change the result but increases computation time. P-values are two-sided.

Column (1) gives the individual impact of treatment on each variable listed on the left hand side column. Column (2) gives the impact of each channel and the treatment dummy, holding the others constant, on obtaining a Secondary Diploma.

Table 26. Coefficients from Maximum Likelihood estimation, Simple II, Secondary Diploma

	(1) Treatment -> Channel		(2) Channel ->Secondary Diploma	
	Coef.	p-value	Coef.	p-value
Delinquent Behavior	-0.122	0.050	-0.059	0.349
Social Trust	0.178	0.005	-0.087	0.136
Hyperactivity	-0.088	0.161	-0.171	0.005
School Performance	0.198	0.001	0.442	0.000
Treatment			0.071	0.193
Number of Obs		250		

Estimates from maximum likelihood estimation. Variables used are residuals from regressions with control variables. Including control variables directly in the estimation does not substantially change the result but increases computation time. P-values are two-sided.

Column (1) gives the individual impact of treatment on each variable listed on the left hand side column. Column (2) gives the impact of each channel and the treatment dummy, holding the others constant, on obtaining a Secondary Diploma.

Table 27. Coefficients from Maximum Likelihood estimation, Complex I, Secondary Diploma

	(1) Treatment -> Channel		(2) Channel -> School Performance		(3) Channel ->Secondary Diploma	
	Coef.	p-value	Coef.	p-value	Coef.	p-value
Delinquent Behavior	-0.122	0.050	-0.330	0.000	-0.059	0.356
Social Trust	0.178	0.005	0.110	0.071	-0.087	0.136
Hyperactivity	-0.088	0.161	-0.115	0.073	-0.171	0.005
School Performance					0.442	0.000
Treatment			0.128	0.027	0.071	0.193
Number of Obs			250			

Estimates from maximum likelihood estimation. Variables used are residuals from regressions with control variables. Including control variables directly in the estimation does not substantially change the result but increases computation time. P-values are two-sided. P-values are two-sided.
Column (1) gives the individual impact of treatment on each variable listed on the left hand side column. Column (2) gives the impact of each behavioral channel and the treatment dummy on school performance, holding the others constant. Column (3) gives the impact of each behavioral channel, the treatment dummy, and school performance on obtaining a secondary diploma, holding the others constant.

Table 28. Coefficients from Maximum Likelihood estimation, Complex I, Secondary Diploma

	(1) Treatment -> Channel		(2) Channel -> Delinquent		(3) Channel -> School Performance		(4) Channel -> Secondary Diploma	
	Coef.	p-value	Coef.	p-value	Coef.	p-value	Coef.	p-value
Delinquent Behavior					-0.330	0.000	-0.059	0.356
Social Trust	0.178	0.005	0.008	0.008	0.110	0.071	-0.087	0.136
Hyperactivity	-0.088	0.161	0.446	0.000	-0.115	0.077	-0.171	0.005
School Performance							0.442	0.000
Treatment			-0.084	0.145	0.128	0.027	0.071	0.193
Number of Obs				250				

Estimates from maximum likelihood estimation. Variables used are residuals from regressions with control variables. Including control variables directly in the estimation does not substantially change the result but increases computation time.
Column (1) gives the individual impact of treatment on each variable listed on the left hand side column. Column (2) gives the impact of Hyperactivity and Social Trust on Delinquent Behavior, holding the other, and the treatment dummy, constant. Column (3) gives the impact of each behavioral channel and the treatment dummy on school performance, holding the others constant. Column (4) gives the impact of each behavioral channel, the treatment dummy, and school performance on obtaining a secondary diploma, holding the others constant.

Table 29. Coefficients from Maximum Likelihood estimation, Simple I, Economic Index

	(1)		(2)	
	Treatment -> Channel		Channel -> Econ Index	
	Coef.	p-value	Coef.	p-value
Social Trust	0.132	0.092	0.208	0.006
School (Combined)	0.137	0.076	0.187	0.013
Treatment			0.069	0.369
Number of Obs		159		

Estimates from maximum likelihood estimation. Variables used are residuals from regressions with control variables. Including control variables directly in the estimation does not substantially change the result but increases computation time. P-values are two-sided.

Column (1) gives the individual impact of treatment on each variable listed on the left hand side column. Column (2) gives the impact of each channel and the treatment dummy, holding the others constant, on the Economic Index

Table 30. Coefficients from Maximum Likelihood estimation, Simple II, Economic Index

	(1)		(2)	
	Treatment -> Channel		Channel -> Econ Index	
	Coef.	p-value	Coef.	p-value
Social Trust	0.134	0.087	0.218	0.004
School Performance	0.132	0.090	0.124	0.145
Secondary Diploma	0.088	0.265	0.107	0.213
Treatment			0.067	0.381
Number of Obs		159		

Estimates from maximum likelihood estimation. Variables used are residuals from regressions with control variables. Including control variables directly in the estimation does not substantially change the result but increases computation time. P-values are two-sided.

Column (1) gives the individual impact of treatment on each variable listed on the left hand side column. Column (2) gives the impact of each channel and the treatment dummy, holding the others constant, on the Economic Index

Table 31. Coefficients from Maximum Likelihood estimation, Simple I, Social Capital Index

	(1)		(2)	
	Treatment -> Channel		Channel -> Social Capital Index	
	Coef.	p-value	Coef.	p-value
Social Trust	0.129	0.100	0.057	0.474
School (Combined)	0.137	0.076	0.018	0.816
Treatment			0.107	0.176
Number of Obs		159		

Estimates from maximum likelihood estimation. Variables used are residuals from regressions with control variables. Including control variables directly in the estimation does not substantially change the result but increases computation time. P-values are two-sided.

Column (1) gives the individual impact of treatment on each variable listed on the left hand side column. Column (2) gives the impact of each channel and the treatment dummy, holding the others constant, on the Social Capital Index

Table 32. Coefficients from Maximum Likelihood estimation, Simple II, Social Capital Index

	(1)		(2)	
	Treatment -> Channel		Channel -> Social Capital Index	
	Coef.	p-value	Coef.	p-value
Social Trust	0.131	0.095	0.069	0.389
School Performance	0.132	0.090	-0.042	0.638
Secondary Diploma	0.088	0.263	0.109	0.216
Treatment			0.104	0.189
Number of Obs		159		

Estimates from maximum likelihood estimation. Variables used are residuals from regressions with control variables. Including control variables directly in the estimation does not substantially change the result but increases computation time. P-values are two-sided.

Column (1) gives the individual impact of treatment on each variable listed on the left hand side column. Column (2) gives the impact of each channel and the treatment dummy, holding the others constant, on the Social Capital Index

Table 33. Coefficients from Maximum Likelihood estimation, Simple I, Group Membership

	(1) Treatment -> Channel		(2) Channel -> Group Membership	
	Coef.	p-value	Coef.	p-value
Social Trust	0.128	0.104	0.125	0.112
School (Combined)	0.137	0.076	0.080	0.311
Treatment			0.077	0.330
Number of Obs		159		

Estimates from maximum likelihood estimation. Variables used are residuals from regressions with control variables. Including control variables directly in the estimation does not substantially change the result but increases computation time. P-values are two-sided.
Column (1) gives the individual impact of treatment on each variable listed on the left hand side column. Column (2) gives the impact of each channel and the treatment dummy, holding the others constant, on Group Membership.

Table 34. Coefficients from Maximum Likelihood estimation, Simple II, Group Membership

	(1) Treatment -> Channel		(2) Channel -> Group Membership	
	Coef.	p-value	Coef.	p-value
Social Trust	0.130	0.098	0.135	0.091
School Performance	0.132	0.090	0.290	0.773
Secondary Diploma	0.088	0.263	1.130	0.258
Treatment			0.074	0.346
Number of Obs		159		

Estimates from maximum likelihood estimation. Variables used are residuals from regressions with control variables. Including control variables directly in the estimation does not substantially change the result but increases computation time. P-values are two-sided.
Column (1) gives the individual impact of treatment on each variable listed on the left hand side column. Column (2) gives the impact of each channel and the treatment dummy, holding the others constant, on Group Membership.

Table 35. Attrition for Employment data

	Missing employment data			
	Age 21		Age 26	
	Number	%	Number	%
Control (N=181): C	76	42	104	57
Treatment (N=69): T	34	49	44	64
Total (N=250)	101	44	148	59
p-value of difference: C-T	0.30		0.37	

Table 36. Test for Attrition Bias

	Dependent Variable: Secondary School Completion				
	True Sample	Age 21		Age 26	
		Attrited Sample	Attrition Bias	Attrited Sample	Attrition Bias
	(1)	(2)	(3)	(4)	(5)
Treatment (0/1)	0.164**	0.171		0.133	
	(0.0697)	(0.105)		(0.128)	
			-0.007		-0.031
Constant	0.379	0.402		0.827	
	(0.288)	(0.432)		(0.564)	
Observations	242	139		101	
R-squared	0.156	0.119		0.155	

*Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1*
Column (1) gives the treatment impact on secondary school graduation estimated for the true (entire) sample. Column (2) gives the treatment impact estimated for a sample where attrition matches the attrition on employment data for age 21. Column (3) gives the treatment impact estimated for a sample where attrition matches the attrition on employment data for age 26. Columns (3) and (5) give the difference in the treatment impact estimates for (1) and (2) and (1) and (4), respectively.

APPENDIX B: SUPPLEMENTARY INFORMATION ON THE INTERVENTION PROGRAM

The activity program is covered in more detail in Bertrand and Reclus-Prince (1988), from which this section draws heavily.[6] The training sessions for the children lasted around 45 minutes and were conducted as follows.

1. The previous session's activity would be reviewed, the group would talk about whether they had used the previous session's behavior, and any homework would be discussed. Children would be praised for having tried to use the skill, and the group would try to identify the benefits of that behavior.
2. The activity for the session would be introduced with a brief explanation from the facilitator, who would ask children to talk about situation in life that are relevant to the behavior. The facilitator would highlight the results of the behavior. The behavior would be modeled, generally through role playing, where the children and/or the teacher would take different parts. Several role plays might be carried out to demonstrate the behavior.
3. The facilitator would guide a practice session, where children would take turns acting in different role plays that demonstrate the behavior, and would alternate roles. During and after the role plays, the facilitator would give feedback and articulate how the behavior was being demonstrated.
4. The facilitator would generalize the behavior by talking about its benefits and how it can make them (the children) feel better and make other people feel better. The facilitator might assign homework (for example, making a drawing of the behavior being discussed). The facilitator would try to work with children to anticipate situations that would call for the behavior, and plan to respond appropriately in those situations. The facilitator would also try to communicate the session's topic to the parents and teachers, who could reinforce that topic at home and in the classroom.
5. The facilitator would also work to reinforce the learned behaviors during the sessions, for example by using a scoring system such as the one below reported in Figure B.1.

[6] Bertrand and Reclus-Prince (1988) also provide references used in development of the program: Bessell and Palomares (1972), Cartledge and Milburn (1980), Schneider and Byrne (1984), Michelson et al (1983) and Shure and Spivack (1984).

Figure B.1: Example of activity program

	Duration of the activity			
Behaviour	First 10-min. period	Second 10-min. period	Third 10-min. period	Fourth 10-min. period
Looking at the person who is talking (1 point)				
Staying seated (1 point)				
Waiting your turn to talk (1 point)				
TOTAL				
Criteria: - Individual performance: 10 points - Choice of rewards: sticker, eraser, felt marker, etc. - Group performance (5 children): 50 points - Choice of rewards: group game, free time, etc.				

B.1) First Year: Social Skills

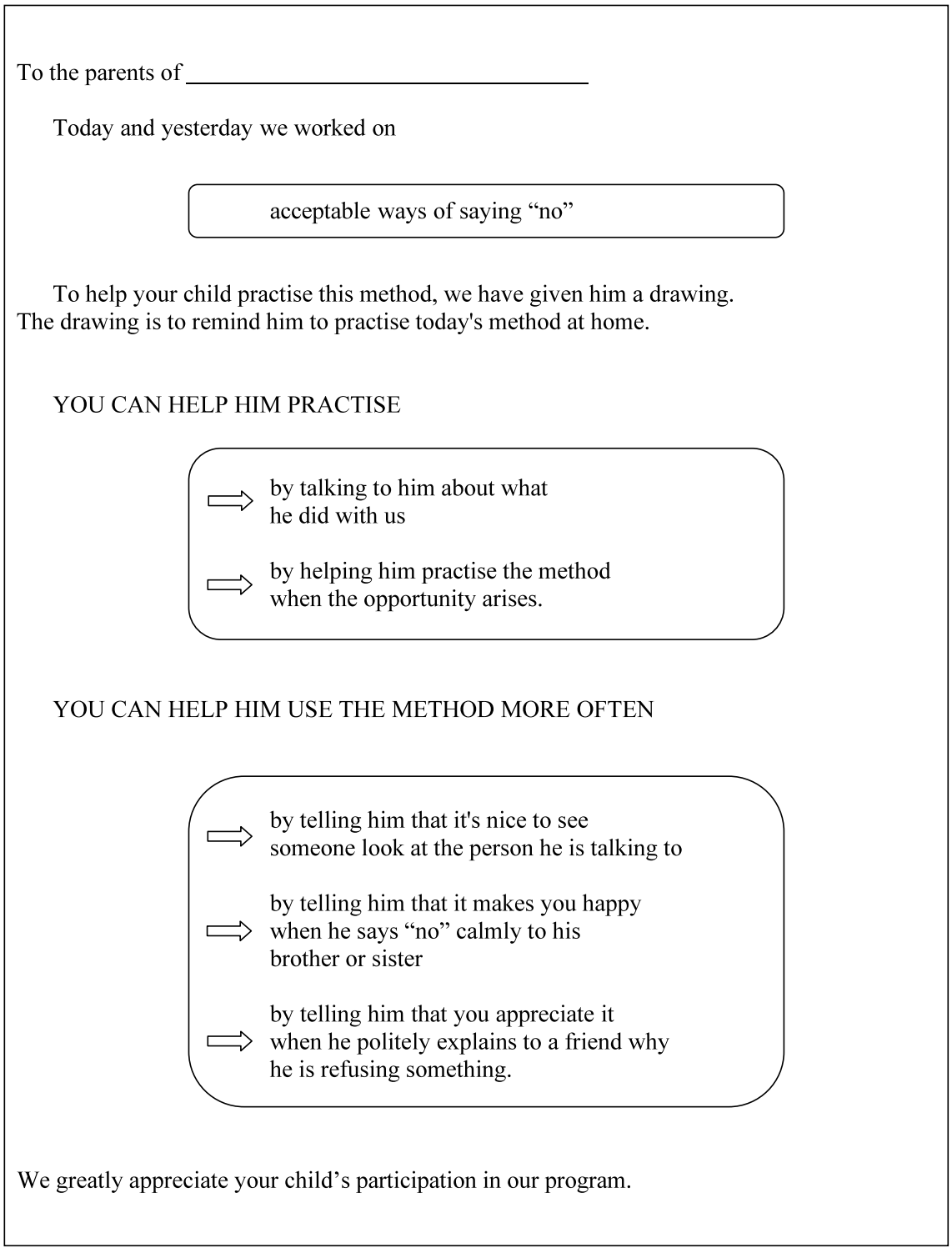
The trainings during first year focused on social skills, and in the second year self-control (described below). Nine behaviors were covered in the social skills training sessions:

- Making Contact (eye contact, smiling, approaching another person)
- Speaking Nicely (friendly body language, tone of voice, explaining what we like)
- Gentle Physical Contact (body language, touching with friendly intent, briefly)
- Helping (approaching someone who needs help, offering to help, waiting for the answer)
- Including and Inviting (body language, saying the person's name, inviting)
- Doing things together (body language, proposing how to do things, seeing whether the other person agrees)
- Saying No (body language, tone of voice, without anger, giving a reason)
- Asking Why (body language, letting the other person finish speaking, asking the question)
- Saying "You're bothering me" (body language, saying the person's name, saying what's wrong, asking the person to stop)

For example, in the "Speaking Nicely" session, the facilitator would begin by asking children if they tried the previous session's behavior (Making Contact), and the results. The facilitator would then explain what "Speaking Nicely" means and why it is useful, and the children would try to give examples of speaking nicely. The role play for modeling would involve a friend who arrives at school wearing a new sweater, and the way to speak nicely in this situation is to go up to the friend and make eye contact, smile, use a pleasant tone of voice, and explain that you like his sweater. Other role plays would include congratulating a friend who got a good grade, encouraging a partner who made a mistake in a game, or thanking parents for doing something nice for. After several role plays, the facilitator would review how important words and body language are to being understood, and assign homework and give out letters for parents and teachers.

In the "Saying No" session, the facilitator would ask if the children had practiced the previous session's behavior (Doing Things Together), and discuss how it went. Then the facilitator would introduce the topic of "Saying No", and explain that we are allowed to say no if we don't want to do something, and that we can do so clearly and without getting angry. The children would then discuss times that they had said no, or times that they wanted to say no but did not feel comfortable doing so. The first role play is about a friend who wants to borrow a bicycle. To say no, the children are told to look at their friend, speak firmly and without anger, and to give a reason for saying no. Other role plays are when a child is not interested in playing but wants to be alone, when a sibling wants to share candy or wants to play. The facilitator reviews that it is good to say honestly what we want or don't want, and clearly explain our reasons, so that the other person can better understand why we are saying no, and accept our answer. Then homework is assigned and letters to parents and teachers are distributed.

Figure B.2: Example of Take-home sheet for parents from Social Skills trainings



To the parents of _____

Today and yesterday we worked on

acceptable ways of saying “no”

To help your child practise this method, we have given him a drawing. The drawing is to remind him to practise today's method at home.

YOU CAN HELP HIM PRACTISE

⇨ by talking to him about what he did with us

⇨ by helping him practise the method when the opportunity arises.

YOU CAN HELP HIM USE THE METHOD MORE OFTEN

⇨ by telling him that it's nice to see someone look at the person he is talking to

⇨ by telling him that it makes you happy when he says “no” calmly to his brother or sister

⇨ by telling him that you appreciate it when he politely explains to a friend why he is refusing something.

We greatly appreciate your child’s participation in our program.

B.2) *Second Year: Self-Control*

In the second year, the sessions focused on the theme of self-control. The principal was to provide a roadmap for children to decide on an appropriate response in a given situation. Four steps are reinforced over the ten sessions:

1. Identify what is happening and feelings associated with it
2. Think about possibilities to say and do
3. Choose an appropriate solution
4. Do the solution, and feel good about it

These steps are reinforced in each session, where the child would say the steps aloud and add the activity-specific details. The specific topics of the ten sessions were

- Paying attention and listening
- Obtaining information before acting
- Following rules and instructions
- Controlling myself
- What to do when angry
- What to do when feeling left out
- What to do when feeling like hitting
- What to do when teased
- Asking forgiveness
- Expressing appreciation

For example, in the unit "I feel like hitting… what do I do?" the facilitator would begin by asking the children if they had use the skills used in the previous session (Controlling Myself). Then the facilitator would introduce the topic of hitting, talking about all the times when one feels like hitting or pushing, and why. The facilitator would explain the appropriate behavior, that is when one wants to hit, one must stop, pause, tell oneself "I won't hit, calm down" and choose another way to act that will not hurt the other person. Children would be asked to give examples from their own life. The first role played, by the facilitator, would involve someone who is bumped into and hurt by another student who was rushing along. Following the steps above, the children are told to

1. Identify what is happening and feelings associated with it: He bumped into me. Was it on purpose? I feel angry, and I want to hit him.

2. Think about possibilities to say and do: I could hit him, I could fall down on the floor and cry, I could go tell my teacher, I could yell at him. I think that usually that classmate is not mean to me. I could do or think about something else. I could move away from him.
3. Choose an appropriate solution: I will stop, tell myself that I won't hit him, and choose another way of dealing with the situation. I will tell him clearly but without being angry how I feel.
4. Do the solution, and feel good about it: I tell him how I feel, that he hurt me, and that I want him to be more careful next time. I tell myself, good job for not getting into a fight! I am proud of myself.

Additional role plays are performed with the children involve, all demonstrating the four steps (identify, think, choose, do), and the principle of stop, pause, say "I won't hit" and choose another action. Other role plays include someone breaking a toy, aggression from others, teasing, or having a foot stepped on accidentally. The facilitator then gives feedback, reinforcing the four steps and the principal of stop. Then the facilitator helps each child prepare their homework, which involves identifying opportunities to practice the behavior of not hitting when you want to.

Figure B.3: Example of take home sheet for self-control training

Activity No. 3

TO FOLLOW RULES AND INSTRUCTIONS

- **I listen**
- **I check whether I have understood**
- **I do what is asked**

I practise the activity by

- **Looking at the person who is speaking**
- **Speaking when it is my turn to speak**
- **Staying seated and remaining calm**
- **Eating without making a mess**

I will practise again tonight at suppertime
My signature _____

Parent's signature _____

Facilitator's comment: