The Impact of Childhood Social Skills and Self-Control Training on Economic and non-Economic Outcomes: evidence from a randomized experiment using administrative data

Yann Algan, Elizabeth Beasley, Sylvana Côté, Frank Vitaro, Richard E. Tremblay and Jungwee Park

ONLINE APPENDIX

Sections in this Appendix

Figures included in this Appendix

Tables included in this Appendix

A.   The Montreal Longitudinal Experimental Study

The Montreal Longitudinal Experimental Study (MLES) began in 1984 and piloted a prevention program for disruptive children with two main components: social skills training in a small group format involving a majority of pro-social peers and training in parenting skills during family visits (see Tremblay et al. 1995 for a detailed description of the experiment).

1.   SELECTION

Kindergarten teachers at 53 schools in Montreal, Canada, were asked to rate the behavior of their male students at the end of the 1984 school year with the Social Behavior Questionnaire (Tremblay et al. 1987). These schools were 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 participants 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.

The disruptiveness scale of the Social Behavior Questionnaire was used to select the boys for the intervention. The scale measures the frequency of physical aggressions, oppositional behavior and hyperactive behavior (Cronbach α = 0.93). Boys with a score above the 70th percentile (N = 250) on this disruptiveness scale were considered to be at high risk of later antisocial behavior.

These 250 participants were randomly assigned to a treatment (69 boys) and a control group (181 boys) by drawing names from a box. The control group was initially divided into two groups, a control group with contact only for yearly assessments and a control group with intensive assessments (at home, at school and in the laboratory). This second control group was used for detailed studies of family interactions, peer interactions and bio-psycho-social development. Another purpose of the second control group was to control for the intensive attention the intervention group were receiving. There have been suggestions that the effects of some interventions are simply due to the attention the children and parents receive. Giving intensive attention to a control group without any intention to change their behavior created an attention-control group. The two control groups 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. These participants are included in the analysis as belonging to their initially assigned treatment groups (intention-to-treat analysis) to avoid any bias due to differential

attrition. We use intention-to-treat analysis and use initial group assignment as the indicator for treatment.

2. INTERVENTION PROGRAM

The intervention program was implemented over a 2-year period, from ages 7 to 9 years (Grades 2 and 3). As already mentioned, the prevention program included a child training component and a parent training component to improve children's non-cognitive skills, particularly social skills and self-control. An additional teacher training, originally planned, was not implemented. The timeline for the program and data collection is shown in Figure 1 of the main text.

Child non-cognitive skills training

The first aspect of the intervention consisted of direct training of non-cognitive skills to children. The experiment drew on randomized and non-randomized studies of children on emotional regulation and impulse control, social-information processing and how to interpret other's intent (Cartledge and Milburn 1980, Kettlewell and Kaush 1983, Michelson et al. 1983, Schneider and Byrne 1987, Weiss et al. 1992, Dodge and Godwin 2013, Dodge 2003).

The training sessions were conducted at school (outside the classroom), in groups of four to seven children, of which one or two would be the treatment participants, and the rest would be boys identified by their teachers as highly pro-social. This arrangement was intended to provide positive role models for the treatment participants and also avoid stigmatizing the treatment participants. 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 on 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 with you. 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 other party, suggested different action plans to solve the problem, anticipated their consequences, selected one action plan and, finally, reinforced themselves for their 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: he has been playing tag, and he just got tagged and is now out. He's upset because he is the first person to be tagged out, and he's angry and disappointed. The facilitator demonstrated how children can respond in this situation: he should notice clues in his body that he is going to lose control (clenching fists, feeling hot), he should think about what happened to make him feel this way (he got tagged first, is worried other kids will laugh at him), he should choose a way to avoid making an angry outburst (count to ten, move away, say to himself "calm down", breathe), and then he should act and praise himself. The facilitator then invited 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 would make observations about what the actors are doing, how they are following the steps, and give feedback. At the end of the session, the facilitator would fill out a workbook with the children to explain how they can practice self-control until the next session ("homework").

Parent 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, Patterson et al. 1975). 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 parent 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, 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 improvement 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 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 was originally targeted to teachers. 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 target boys. However, compliance with this aspect of the intervention was very weak.

The training and support activities were carried out by two university-trained child-care workers, one a psychologist and one a social worker, all working full time. 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.

3.   DETAILED CONTENTS OF TRAINING

The activity program is covered in more detail in other articles (Bertrand and Reclus-Prince 2010) from which this section draws heavily, and which explicitly lay out the links with the child development literature used in the development of the program. 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 S 1.

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.

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 each 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 used 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 S 1. Example of activity program

Behaviour	Duration of the activity			
	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.				

Figure S 2. Example of Take-home sheet for parents

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.

Figure S 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:

B. Compliance

Compliance was not complete. 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 participants are included in the analysis as belonging to their initially assigned treatment groups (intention-to-treat analysis). Of those assigned to the treatment group, 67% agreed to participate. Table S 1 shows the difference between compliers and non-compliers on baseline variables. Differences are significant at the 10% level or higher for 4 out of 21 variables tested: prestige of mother's job, age of mother, initial aggression, and initial fighting. We should not expect compliers and non-compliers to be similar – since the treatment required parent cooperation and some investment of parent time, it is not surprising that some parent characteristics are different between the two groups. In part because there are differences on observables, our preferred specification uses Intention to Treat (ITT) estimation, so that all subjects assigned to the treatment group are considered treated for the analysis, and we provide estimates adjusted for non-compliance in for completeness and as a credibility check on the size of the treatment estimates.

Table S 1. Characteristics of compliers and non-compliers

	N	Non-Complier Mean	Non-Complier SD	Complier Mean	Complier SD	Difference	Equality of means p-value
Father Works	69	3.30	3.48	3.41	3.47	0.11	0.90
Mother Works	69	2.00	1.60	1.83	0.38	-0.17	0.48
Prestige of Mother	60	37.03	11.25	31.36	9.17	-5.66	0.04
Prestige of Father	53	35.26	8.13	35.20	10.58	-0.06	0.98
Family Adversity	68	0.45	0.29	0.42	0.26	-0.03	0.71
Number of Kids in Fam	68	1.09	0.81	1.07	0.80	-0.03	0.90
Father Education	60	9.84	1.77	9.98	2.70	0.13	0.84
Mother Education	68	10.14	2.42	9.78	2.23	-0.35	0.55
Age of Mother	68	22.32	4.55	24.82	4.61	2.50	0.04
Age of Father	56	28.48	5.36	28.18	5.39	-0.30	0.85
Lives with both Parents	69	0.48	0.51	0.46	0.50	-0.02	0.87
Age	69	6.00	0.30	5.96	0.28	-0.05	0.54
Attended Preschool	69	0.13	0.34	0.22	0.42	0.09	0.39
Initial Aggression	69	13.30	4.68	15.28	4.43	1.98	0.09
Initial Anxiety	69	4.65	2.95	4.07	2.76	-0.59	0.42
Initial Opposition	69	5.65	1.77	5.89	2.01	0.24	0.63
Initial Prosociality	69	6.00	4.80	7.48	4.33	1.48	0.20
Initial Fighting	69	2.91	1.65	3.76	1.42	0.85	0.03
Initial Antisocial	68	1.52	1.31	1.04	1.17	-0.48	0.13
Initial Hyperactivity	68	2.70	1.18	3.09	1.18	0.39	0.20
Initial Inattention	69	4.78	1.88	3.89	2.27	-0.89	0.11

Data from baseline data collection, 1984.

C.   Attrition

For the skills measured in adolescence, attrition is lower in early adolescence (less than 10% for most behavior outcomes) than in later adolescence (around 15%). Table S 2 presents rates of attrition for the different adolescent variables and the p-value of the difference between treatment and control. In no case are the attrition rates significantly different in treatment and control groups.

Table S 2. Attrition

	(1)	(2)	(3)	(4)	(5)
		% Missing		Difference	P-value of
	Non-disruptive	Control	Treatment	(T-C)	difference
Adult Social and Economic Outcomes					
Tax data (at least one year)	0%	0%	3%	3%	0.58
Number of Crimes	0%	0%	0%	0%	
Secondary Completion	0%	1%	0%	-1%	0.54
Young adult survey 2001	33%	40%	58%	2%	0.80
Young adult survey 2006	45%	58%	64%	6%	0.37
At least one young adult survey	25%	36%	38%	2%	0.80
Skills: Age 10-13					
Trust	2%	2%	4%	2%	0.36
Friendliness	1%	1%	1%	1%	0.48
Aggression Control	1%	1%	1%	1%	0.48
Attention Control	1%	1%	1%	1%	0.48
Self Esteem	4%	7%	9%	2%	0.57
Altruism	1%	1%	1%	1%	0.48
Grades	7%	13%	10%	-3%	0.58
Held Back	0%	1%	0%	-1%	0.54
Special Education	1%	0%	0%	0%	
Skills: Age 14-17					
Trust	10%	16%	13%	-2%	0.63
Friendliness	10%	16%	13%	-2%	0.63
Aggression Control	10%	16%	13%	-2%	0.63
Attention Control	11%	17%	15%	-2%	0.69
Self Esteem	13%	20%	16%	-4%	0.42
Altruism	13%	21%	19%	-2%	0.71
Grades	10%	14%	13%	-1%	0.79
Held Back	1%	1%	0%	-1%	0.54
Special Education	1%	1%	0%	-1%	0.38

Table shows the percent of subjects from each group missing data for each of the items. Column 5 provides the p-value of the difference between the treatment and control groups. The non-disruptive group is composed of those children who scored below the 70th percentile of anti-social behavior on the initial questionnaire in 1984. This non-disruptive group did not participate in the randomized evaluation and serve as a reference group. Those who scored above the 70th percentile were randomized into either the treatment or control groups.

However, attrition in the young adult survey data is high. This data was reported using self-administered paper questionnaires that were sent by mail to the subjects, and many subjects either did not fill out the questionnaire or did not fill out all parts of the questionnaire. In 2000, 49% of the treatment group are missing data and 46% of the control group, and in 2005, 63% of the intervention subjects are missing, and 57% of the control

subjects are missing (in the normative population, 45% are missing responses). The six percentage points difference between the treatment and control groups has a p-value of 0.3662, so the difference between groups alone is not statistically significant.

We provide two pieces of evidence that the results from the young adult survey data should still be taken into account. First, the estimate of the labor market impact from the young adult survey data is quite similar to the estimate of the labor market impact from the tax data. Second, we carry out an attrition falsification exercise: we construct an artificial secondary school graduation variable which mimics the attrition for the 2000 and 2005 questionnaires, and we compare the coefficient from a regression using this sample to a regression using the true administrative data, which is nearly complete. That is, we delete information in the secondary school graduation variable for observations that are missing data for the 2000 and 2005 questionnaires, to try to ascertain the possible size and direction of the attrition bias. We find a small downward bias in our treatment coefficient, which suggests that the attrition is likely to reduce our treatment coefficient.

Table S 3 shows that from the original 250 subjects, roughly 44% are missing data for 2000, and 59% are missing data for 2005. 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, but the difference is not statistically significant.

Table S 3. Attrition for self-reported employment data

	Missing employment data			
	2000		2005	
	Number	%	Number	%
Control (181)	76	42	104	57
Treatment (69)	34	49	44	64
Total (250)	101	44	148	59
p-value of difference		0.30		0.37

This level of attrition raises concern that our estimate of the impact on treatment 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 GRIP 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.

1.   COMPARISON OF LABOR MARKET OUTCOMES IN SELF-REPORTED RESULTS AND TAX DATA RESULTS

It is difficult to make a direct comparison of the survey data and the tax data, since the format of the data and questions are different in each case. Using the survey data, we construct a variable equal to the percent of time that a person reports being occupied full time, either in work or study. As shown in Table 5, using the adult survey data, the impact of the treatment is estimated to be between 6 percentage points and 10 percentage points, from a control group average of 77% (representing an increase of around 10%). The tax data does not indicate whether employment is full time or part time. Using the tax data, as shown in Table 6, over the 19-year period, the control group had at least some employment income for 13 years (65% of the time) and the treatment group had at least some employment income for 15 years (75% of the time). This difference of ten percentage points is an increase of around 15% - that is, less than is reported in the survey data. These estimates are of similar direction and order of magnitude, and suggest that, if anything, the attrition in the adult survey data is likely to bias the estimate of impact towards zero.

2.   USING ADMINISTRATIVE DATA ON SCHOOLING TO ESTIMATE BIAS

This finding is reinforced in an attrition falsification exercise. 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. Administrative data on secondary school completion is available for nearly the entire sample (242 out of 250 subjects, or 97%).

First, we create a false secondary school variable that takes the value of missing if the subject is missing data for employment in 2000 (or 2005), and the value of the true secondary school variable if they have data for employment in 2000 (or 2005). We then estimate the impact of the program on secondary school completion using our preferred paper specification. We compare this to an estimate of the impact of the program which uses data from the entire sample. We control for all variables related to treatment assignment and that

are related to attrition across treatment groups.

The first three columns of Table S 4 show this test for employment in 2000 (note that in any case we find no significant impact on employment in 2000). The second three columns show this test for employment in 2005. The regression using the constructed (attrited to match attrition in the 2005 employment variable) secondary school variable has a smaller and non-significant coefficient on treatment (column 4) than the true coefficient (column 5). This suggests that the coefficient of treatment on employment (column 6) is likely to be an underestimation of the impact on employment.

Table S 4. Falsification test of attrition bias

	(1)	(2)	(3)	(4)	(5)	(6)
	Test for attrition on employed 2000			Test for attrition on employed 2005		
	Secondary School		Employed 2000	Secondary School		Employed 2005
	00 sample	Full Sample		04 sample	Full Sample	
Treatment (0/1)	0.171	0.164	0.0793	0.133	0.164	0.108
	(0.105)	(0.0697)	(0.0556)	(0.128)	(0.0697)	(0.0547)
Constant	0.402	0.379	1.268	0.827	0.379	1.205
	(0.432)	(0.288)	(0.295)	(0.564)	(0.288)	(0.267)
Observations	139	242	140	101	242	102
R-squared	0.119	0.156	0.077	0.155	0.156	0.139

Robust standard errors in parentheses

D.   Measurement of Skills

For adolescent outcomes, we break the data into two periods and analyze data that are available from both periods: early adolescence (ages 10 to 13, or 1988-1991), and late adolescence (ages 14 to 17, or 1992-1995). The year 1992 was chosen as the break year because it is in this year that the treatment and control groups begin to diverge in whether they have repeated a year in school.

Due to space considerations, exact translations and wording for the questions is not provided here but is available upon request as well as in supporting materials such as Tremblay et al 1987 and 1992. We use individual questions from the different measuring instruments rather than the original scales themselves. This approach allows for the possibility that individual questions might cluster together effectively and allows subject-reported and teacher-reported data to be used together when possible. Note that there may be attenuation bias due to measurement error with respect to the estimation of the latent variable outcomes as well as the real-valued outcomes. If this is the case, it would bias the coefficients of the measurement variables on the skill towards zero, and in turn the coefficients of the treatment on the skill and the skill on the other outcomes towards zero. Any potential bias should be against finding a significant result.

Our identification of skills is based on exploratory factor analysis. Combining all data available and averaging over the years available for the entire ELEM sample of 1037 subjects, we use factor analysis to examine how the factors combine into groups potentially measuring the same latent variable. Groupings are the same when the 69 treatment subjects are excluded. We include the original questions that were used to identify the disruptive sub-sample, and questions from several well-known psychological inventories (Jesness and Wedge 1983, Kovacs 1983, Marsh 1990, Rosenberg 1965, Tremblay et al. 1992).[1]

---

[1] The original measure of disruptiveness included teacher ratings of the following behavior: restlessness, fidgeting, destroying things, fighting, disliked by others, irritable, disobedient, dishonest, bullying, refusing to share, blaming others, inconsiderate, and kicking, biting, or hitting. We combine these teacher variables with student reported data and produce two new measures related to disruptiveness, Aggression Control and Attention Control.

Note that our interest is in comparing levels of the measured construct between the treatment and the control group within this population, rather than estimating a baseline level of a particular construct in this population and comparing it to another population. We provide the Cronbach alphas as confirmation that the variables group together well. Note that the Cronbach alphas are the same whether or not the treatment group is included in the sample, to the 0.001 place, except for Trust in late adolescence which is 0.6802 in the full sample and 0.6778 when the treatment subjects are excluded.

To measure the skills, we use both principal component analysis and a simple average of z-scores for the non-missing variables available for a given observation.[2] The estimated skills are quite similar, and the results do not differ substantially but there are fewer observations available for the principal component analysis. We use the z-score averages in the analysis that follows, so changes are in terms of standard deviations. Correlation between the different adolescent outcomes is given in Table S 12.

Aggression Control (Table S 5) includes variables on fighting, damaging property, extreme acting out (self-reported), and fighting and bullying (teacher reported).

Attention Control (Table S 6) includes variables on difficulty in applying himself in school, finding it hard to concentrate, finding it hard to find things to do (self-reported), and weak concentration, easily distracted, and jumps from one thing to another (teacher reported).

Note that Aggression Control can be understood to roughly correspond the construct of Aggressiveness-Opposition according to the DSM, and Attention Control can be understood to roughly correspond to the construct of Impulsiveness-Hyperactivity-Inattention (American Psychiatric Association 2013).

Trust (Table S 7) measures generalized trust and the ability to interpret the intentions of others. It includes variables on whether the subject trusts others, strangers, the police and teachers or whether she thinks one is better off not to trust anyone (self-reported). Perspective taking includes variables on whether the subject assumes that a

---

[2] We retain an observation only if there are non-missing values for at least two of the component variables.

bump is intentional and gets angry (self and teacher reported) and is inconsiderate of others (teacher-reported).

Friendship (

Table S 8) measures close relationships with family and friends. It includes variables on whether the subject gets advice from his best friend, shares his problems and personal information with his best friend, lies, cares about whether other people like him, spends time with his friend and talking to his friend (self-reported) and whether he tends to work alone and is liked by other children (teacher-reported).

Self-esteem (Table S 9) measures psychological well-being and self-esteem. It includes the components of the Rosenberg Self-Esteem scale, as well whether the subject thinks other people do things more easily than he does, people talk about him behind his back, he has trouble making decisions, he feels alone even when he's with someone, and that when something doesn't work, he is to blame (self-reported) as well as whether he gives up easily (teacher reported).

Altruism (Table S 7) measures active altruism. It is composed only of teacher-reported prosocial behavior and includes whether the subject tries to stop others from fighting, invites a child who is left out to play, helps injured children, volunteers to put things away, congratulates others, shows sympathy, helps children having trouble, helps ill children, or volunteers to clean up a spill.

School Performance (Table S 11) include yearly grades in Math and French, as well as the status of the subject's class each year: whether the subject had repeated a grade or was in special education class.

IQ Verbal IQ was tested in 1991, when the subjects were around 13 years old using the Sentence Completion Test (Tremblay 2010).

Table S 5. Components of Aggression Control

	Source	Non-disruptive N	Non-disruptive mean	Control N	Control mean	Treatment N	Treatment Mean	Control - Treatment Diff	p-value
					Ages 10-13: Cronbach alpha = 0.83				
Fights	Subject	769	0.35	173	0	66	-0.04	-0.04	0.77
Mistreats others	Subject	712	0.12	157	0	63	0.22	0.22	0.14
Steals	Subject	769	0.18	173	0	66	0.28	0.28	0.02
Force others to do things	Subject	764	0.23	171	0	66	0.17	0.17	0.23
Damage other's property	Subject	727	0.27	165	0	64	-0.02	-0.02	0.90
Impolite to Teacher	Subject	764	0.45	171	0	66	0.14	0.14	0.31
Disturbs Class	Subject	764	0.57	171	0	66	0.14	0.14	0.31
Vandalizes School	Subject	768	0.23	173	0	66	0.24	0.24	0.06
Make people angry	Subject	754	0.30	169	0	63	0.28	0.28	0.05
React strongly to insult	Subject	754	0.40	169	0	63	0.35	0.35	0.02
Crazy if provoked	Subject	754	0.30	169	0	63	0.20	0.20	0.17
Fights	Teacher	778	0.73	180	0	68	0.20	0.20	0.16
Bullies	Teacher	778	0.74	180	0	68	0.21	0.21	0.14
Hits/bites/kicks	Teacher	778	0.61	180	0	68	0.14	0.14	0.31
					Ages 14-17: Cronbach alpha = 0.86				
Fights	Subject	700	0.32	150	0	59	0.11	0.11	0.46
Mistreats others	Subject	700	0.24	150	0	59	0.20	0.20	0.20
Steals	Subject	700	0.22	149	0	59	0.23	0.23	0.10
Force others to do things	Subject	700	0.27	150	0	59	0.12	0.12	0.43
Damage other's property	Subject	698	0.25	147	0	59	0.17	0.17	0.23
Impolite to Teacher	Subject	698	0.42	147	0	59	0.20	0.20	0.18
Disturbs Class	Subject	698	0.26	147	0	58	0.19	0.19	0.22
Vandalizes School	Subject	698	0.22	147	0	59	0.25	0.25	0.07
Make people angry	Subject	671	0.19	138	0	55	0.29	0.29	0.07
React strongly to insult	Subject	673	-0.01	138	0	55	0.06	0.06	0.70
Crazy if provoked	Subject	670	0.25	138	0	54	0.15	0.15	0.36
Fights	Teacher	687	0.38	142	0	56	-0.17	-0.17	0.32
Bullies	Teacher	687	0.37	143	0	56	0.03	0.03	0.87
Hits/bites/kicks	Teacher	686	0.30	142	0	56	-0.10	-0.10	0.56

Variables are standardized based on the mean and standard deviation of the control group. Variables are oriented so that more positive values indicate better outcomes (more self-control).

Figure S 4. Distribution of Aggression Control
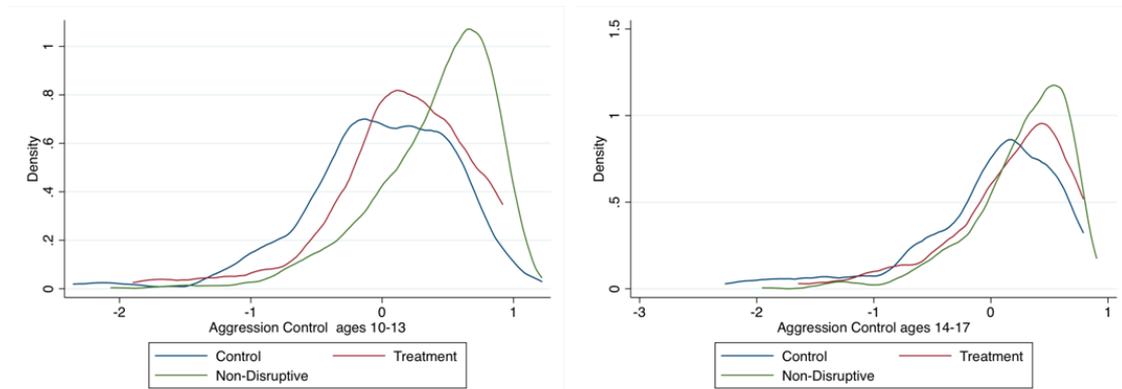
Kolmogorov-Smirnov test for equality of Treatment and Control distributions gives a p-value of 0.04 for ages 10-13 and 0.01 for ages 14-17.

Table S 6. Components of Attention Control

Variable	Source	Non-disruptive		Control		Treatment		Control - Treatment	
		N	mean	N	mean	N	Mean	Diff	p-value
				Ages 10-13: Cronbach alpha = 0.73					
Difficult to apply self at school	Subject	764	0.10	171	0	66	-0.13	-0.13	0.38
Tries hard at school	Subject	764	0.28	171	0	66	0.16	0.16	0.26
Hard to concentrate	Subject	754	0.21	169	0	63	0.19	0.19	0.18
Daydream more than I should	Subject	754	0.17	169	0	63	0.24	0.24	0.09
Weak concentration	Teacher	778	0.65	180	0	68	0.17	0.17	0.21
Easily distracted	Teacher	778	0.62	180	0	68	0.15	0.15	0.29
Jumps from one thing to another	Teacher	709	0.51	160	0	63	0.31	0.31	0.03
				Ages 14-17: Cronbach alpha = 0.67					
Difficult to apply self at school	Subject	698	0.05	147	0	57	0.10	0.10	0.50
Tries hard at school	Subject	698	0.22	147	0	58	0.17	0.17	0.27
Hard to concentrate	Subject	672	0.15	138	0	55	0.22	0.22	0.16
Daydream more than I should	Subject	669	-0.06	138	0	53	-0.05	-0.05	0.75
Weak concentration	Teacher	687	0.48	142	0	56	-0.10	-0.10	0.51
Easily distracted	Teacher	687	0.54	143	0	56	-0.06	-0.06	0.71
Jumps from one thing to another	Teacher	687	0.37	142	0	56	0.05	0.05	0.77

Variables are standardized based on the mean and standard deviation of the control group. Variables are oriented so that more positive values indicate better outcomes (more self-control).

Figure S 5. Distribution of Attention Control

Kolmogorov-Smirnov test for equality of Treatment and Control distributions gives a p-value of 0.02 for ages 10-13 and 0.99 for ages 14-17.

Table S 7. Components of Trust

Variable	Source	Non-disruptive N	Non-disruptive mean	Control N	Control mean	Treatment N	Treatment Mean	Treatment - Control Diff	Treatment - Control p-value
				Ages 10-13: Cronbach alpha = 0.61					
Trust Police	Subject	764	0.24	171	0	66	0.33	0.33	0.02
Thinks a bump is on purpose	Subject	712	0.24	157	0	63	0.11	0.11	0.47
Cares what teacher thinks	Subject	764	0.06	170	0	66	0.10	0.10	0.49
Angry at a bump	Teacher	708	0.56	160	0	63	0.24	0.24	0.10
Inconsiderate	Teacher	777	0.68	180	0	68	0.20	0.20	0.15
If police don't like you…	Subject	754	0.26	169	0	63	0.03	0.03	0.83
Better to trust no one	Subject	754	0.19	169	0	63	-0.18	-0.18	0.22
People like to play the boss	Subject	754	0.32	169	0	63	0.19	0.19	0.22
People don't do as they say	Subject	754	0.32	169	0	63	0.25	0.25	0.09
OK to steal from dishonest store	Subject	754	0.24	169	0	63	0.10	0.10	0.49
Tempted to lie when in trouble	Subject	754	0.29	169	0	63	0.32	0.32	0.03
OK to steal from someone rich	Subject	753	0.13	169	0	63	0.13	0.13	0.37
				Ages 14-17: Cronbach alpha = 0.68					
Trust Police	Subject	700	0.31	149	0	59	0.35	0.35	0.02
Thinks a bump is on purpose	Subject	700	0.24	150	0	59	0.09	0.09	0.56
Cares what teacher thinks	Subject	698	0.10	147	0	59	0.07	0.07	0.68
Angry at a bump	Teacher	685	0.48	143	0	56	0.00	0.00	0.98
Inconsiderate	Teacher	686	0.36	143	0	56	-0.05	-0.05	0.76
If police don't like you…	Subject	672	0.29	137	0	55	0.38	0.38	0.01
Better to trust no one	Subject	672	0.30	136	0	55	0.29	0.29	0.06
People like to play the boss	Subject	673	0.05	138	0	54	0.09	0.09	0.56
People don't do as they say	Subject	672	0.24	136	0	54	0.18	0.18	0.26
OK to steal from dishonest store	Subject	672	0.06	136	0	54	0.02	0.02	0.91
Tempted to lie when in trouble	Subject	673	0.11	138	0	55	0.36	0.36	0.02
OK to steal from someone rich	Subject	673	0.14	138	0	55	0.01	0.01	0.96

Variables are standardized based on the mean and standard deviation of the control group. Variables are oriented so that more positive values indicate better outcomes (more Trust).
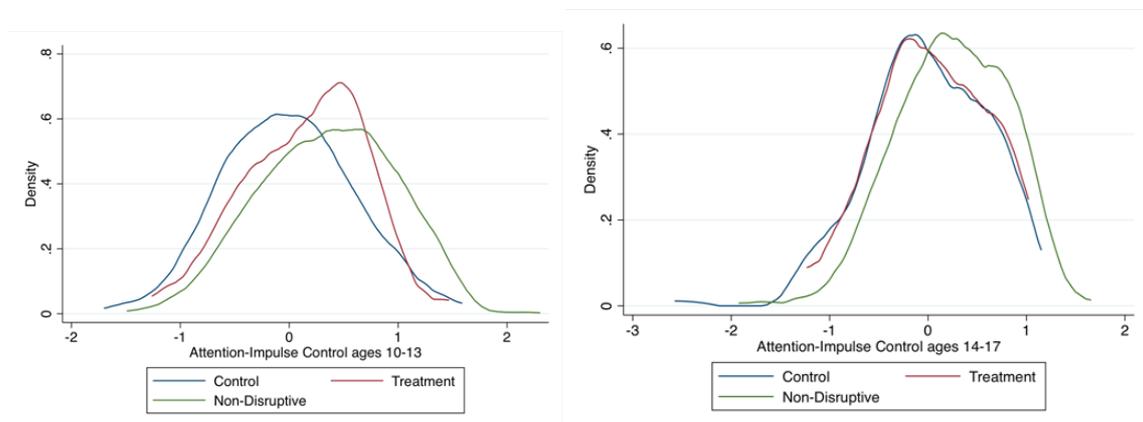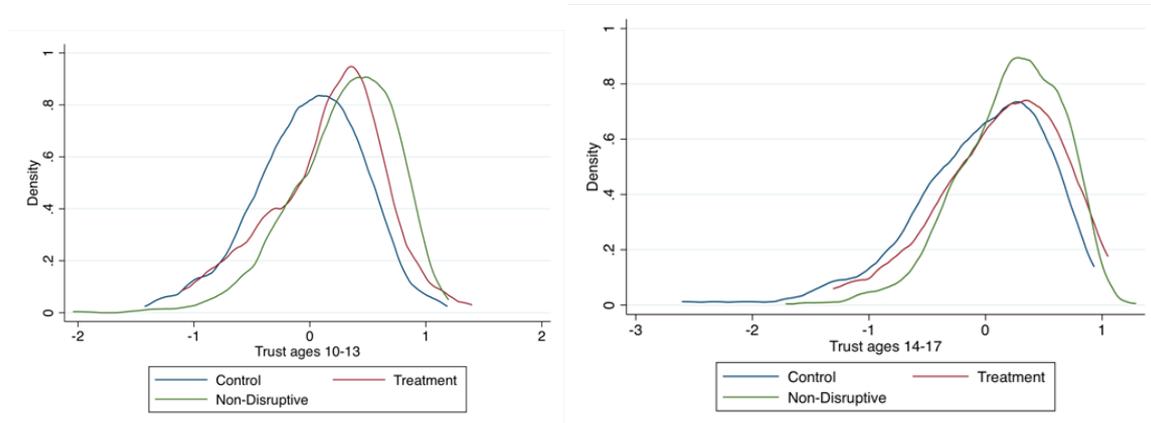
Figure S 6. Distribution of Trust

Kolmogorov-Smirnov test for equality of Treatment and Control distributions gives a p-value of 0.00 for ages 10-13 and 0.22 for ages 14-17.

Table S 8. Components of Friendliness

Variable	Source	Non-disruptive N	Non-disruptive mean	Control N	Control mean	Treatment N	Treatment Mean	Control - Treatment Diff	Control - Treatment p-value
		Ages 10-13: Cronbach alpha = 0.64							
Trusts best friend	Subject	718	0.19	162	0	64	-0.15	-0.15	0.33
Gets advice best friend	Subject	718	0.04	162	0	64	-0.17	-0.17	0.27
Talks pers best friend	Subject	718	-0.09	162	0	64	-0.06	-0.06	0.68
Talks probs best friend	Subject	718	-0.1	162	0	64	-0.18	-0.18	0.23
Care if other people like me	Subject	754	0.06	169	0	63	-0.03	-0.03	0.86
I never lie	Subject	754	-0.04	169	0	63	-0.15	-0.15	0.33
Better not to talk to anyone	Subject	754	0.25	169	0	63	-0.02	-0.02	0.91
Time spent with friend	Subject	764	0.12	171	0	66	0.05	0.05	0.71
Time talking to friend	Subject	763	0	171	0	66	-0.13	-0.13	0.37
Tends to work alone	Teacher	778	0.17	180	0	68	0.06	0.06	0.67
Not liked by other children	Teacher	778	0.73	180	0	68	0.1	0.1	0.5
		Ages 14-17: Cronbach alpha = 0.65							
Trusts best friend	Subject	700	0.19	150	0	59	0.04	0.04	0.81
Gets advice best friend	Subject	700	0.05	150	0	59	0.14	0.14	0.34
Talks pers best friend	Subject	700	0.1	150	0	59	0.1	0.1	0.53
Talks probs best friend	Subject	700	0.05	150	0	59	-0.08	-0.08	0.62
Care if other people like me	Subject	686	0.15	142	0	56	-0.09	-0.09	0.59
I never lie	Subject	671	0.12	138	0	55	0.33	0.33	0.03
Better not to talk to anyone	Subject	671	-0.05	137	0	54	-0.11	-0.11	0.51
Time spent with friend	Subject	672	0.22	138	0	55	-0.18	-0.18	0.28
Time talking to friend	Subject	686	0.47	143	0	56	0.01	0.01	0.96
Tends to work alone	Teacher	675	0.03	146	0	57	0.11	0.11	0.46
Not liked by other children	Teacher	700	-0.03	150	0	59	-0.09	-0.09	0.55

Variables are standardized based on the mean and standard deviation of the control group. Variables are oriented so that more positive values indicate better outcomes (more Friends).

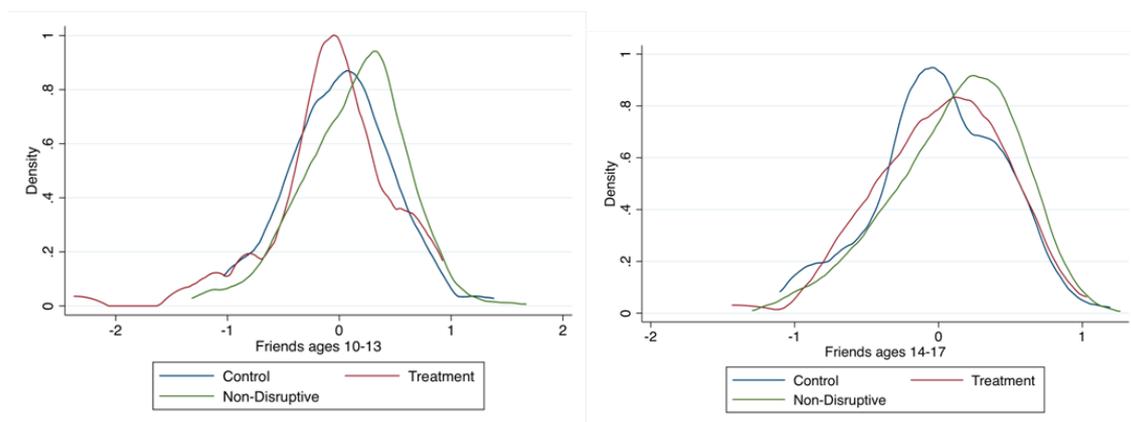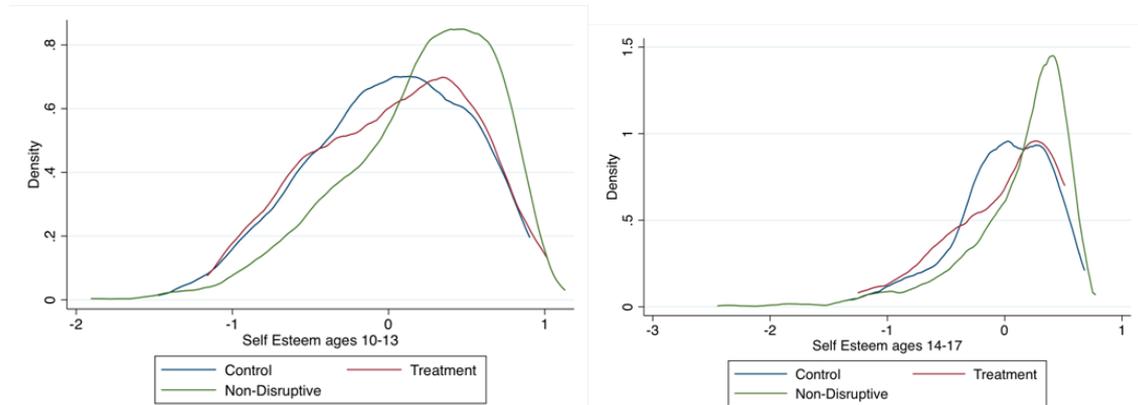## Figure S 7. Distribution of Friendliness

Kolmogorov-Smirnov test for equality of Treatment and Control distributions gives a p-value of 0.55 for ages 10-13 and 0.96 for ages 14-17.

Table S 9. Components of Self-Esteem

Variable	Source	Non-disruptive N	Non-disruptive mean	Control N	Control mean	Treatment N	Treatment Mean	Control - Treatment Diff	Control - Treatment p-value
				Ages 10-13: Cronbach alpha = 0.82					
Others more capable than I	Subject	754	0.13	169	0	63	0.06	0.06	0.67
Talk behind my back	Subject	754	0.22	169	0	63	0.02	0.02	0.87
No one understands what I feel	Subject	754	0.27	169	0	63	0.16	0.16	0.27
I feel worried	Subject	754	0.12	169	0	63	0.03	0.03	0.86
Have trouble deciding	Subject	754	0.08	169	0	63	0.24	0.24	0.11
Feel alone even when I'm not	Subject	754	0.19	169	0	63	-0.29	-0.29	0.06
Something doesn't work	Subject	754	0.29	169	0	63	0.09	0.09	0.53
Satisfied with myself	Subject	719	0.1	161	0	62	-0.22	-0.22	0.17
I'm not worth anything	Subject	719	0.07	161	0	62	-0.04	-0.04	0.78
I have good qualities	Subject	719	0.2	161	0	62	0.01	0.01	0.94
I do things well	Subject	719	0.06	160	0	62	-0.15	-0.15	0.35
Not much to be proud of	Subject	719	0.25	161	0	62	0.09	0.09	0.53
I feel useless	Subject	719	0.14	161	0	62	0.14	0.14	0.34
I'm a valuable person	Subject	719	0.14	161	0	62	-0.09	-0.09	0.55
I should have more self-respect	Subject	719	0.1	160	0	62	0.05	0.05	0.72
I feel like a failure	Subject	718	0.22	161	0	62	-0.03	-0.03	0.86
Positive attitude towards myself	Subject	718	0.28	160	0	62	0.04	0.04	0.79
Gives up easily	Teacher	778	0.7	180	0	68	0.17	0.17	0.23
				Ages 14-17: Cronbach alpha = 0.75					
Others more capable than I	Subject	671	0.11	138	0	55	0.07	0.07	0.68
Talk behind my back	Subject	670	0.15	137	0	55	-0.14	-0.14	0.38
No one understands what I feel	Subject	669	0.12	138	0	54	-0.23	-0.23	0.16
I feel worried	Subject	673	0.1	138	0	55	-0.14	-0.14	0.38
Have trouble deciding	Subject	670	0.04	137	0	54	-0.17	-0.17	0.3
Feel alone even when I'm not	Subject	671	0.24	138	0	54	0.16	0.16	0.3
Something doesn't work	Subject	672	0.29	138	0	55	0.07	0.07	0.65
Satisfied with myself	Subject	615	0.02	124	0	50	0.1	0.1	0.53
I'm not worth anything	Subject	613	0.08	124	0	50	0.05	0.05	0.76
I have good qualities	Subject	615	0.08	123	0	50	0.03	0.03	0.86
I do things well	Subject	615	-0.06	124	0	50	0.09	0.09	0.53
Not much to be proud of	Subject	615	-0.09	123	0	50	-0.25	-0.25	0.18
I feel useless	Subject	615	-0.07	124	0	50	-0.16	-0.16	0.35
I'm a valuable person	Subject	616	0.22	124	0	50	0.21	0.21	0.17
I should have more self-respect	Subject	618	0.16	124	0	50	-0.16	-0.16	0.34
I feel like a failure	Subject	618	0.05	124	0	50	-0.07	-0.07	0.68
Positive attitude towards myself	Subject	617	0	124	0	50	0.21	0.21	0.17
Gives up easily	Teacher	687	0.56	143	0	56	-0.11	-0.11	0.49

Variables are standardized based on the mean and standard deviation of the control group. Variables are oriented so that more positive values indicate higher Self-Esteem.

Figure S 8. Distribution of Self-Esteem

Kolmogorov-Smirnov test for equality of Treatment and Control distributions gives a p-value of 0.82 for ages 10-13 and 0.39 for ages 14-17.

Table S 10. Components of Altruism

Variable	Source	Non-disruptive N	Non-disruptive mean	Control N	Control mean	Treatment N	Treatment Mean	Control - Treatment Diff	Control - Treatment p-value
					Ages 10-13: Cronbach alpha = 0.92				
Tries to stop a fight	Teacher	778	0.17	180	0	68	-0.10	-0.10	0.48
Invites left-out child to play	Teacher	778	0.11	179	0	68	-0.18	-0.18	0.21
Helps an injured child	Teacher	778	-0.01	180	0	68	-0.05	-0.05	0.74
Volunteers to put things away	Teacher	778	0.12	180	0	68	0.02	0.02	0.87
Congratulates others	Teacher	778	0.26	180	0	68	-0.07	-0.07	0.62
Shows sympathy	Teacher	778	0.20	180	0	68	-0.18	-0.18	0.20
Helps a child having trouble	Teacher	778	0.14	180	0	68	-0.11	-0.11	0.44
Helps an ill child	Teacher	778	0.05	180	0	68	-0.17	-0.17	0.22
Helps a crying child	Teacher	778	0.11	180	0	68	-0.09	-0.09	0.52
Volunteers to clean a spill	Teacher	777	-0.08	180	0	68	-0.17	-0.17	0.22
					Ages 14-17: Cronbach alpha = 0.90				
Tries to stop a fight	Teacher	687	-0.03	142	0	56	0.13	0.13	0.43
Invites left-out child to play	Teacher	685	-0.10	143	0	56	-0.09	-0.09	0.56
Helps an injured child	Teacher	684	-0.01	142	0	56	0.03	0.03	0.84
Volunteers to put things away	Teacher	686	-0.01	142	0	56	-0.15	-0.15	0.37
Congratulates others	Teacher	686	0.07	142	0	56	0.00	0.00	0.98
Shows sympathy	Teacher	686	0.15	142	0	56	0.13	0.13	0.42
Helps a child having trouble	Teacher	687	-0.02	143	0	56	-0.24	-0.24	0.12
Helps an ill child	Teacher	684	-0.10	142	0	56	-0.15	-0.15	0.33
Helps a crying child	Teacher	685	0.02	142	0	56	-0.05	-0.05	0.78
Volunteers to clean a spill	Teacher	686	-0.12	143	0	56	0.02	0.02	0.90

Variables are standardized based on the mean and standard deviation of the control group. Variables are oriented so that more positive values indicate more Altruism.
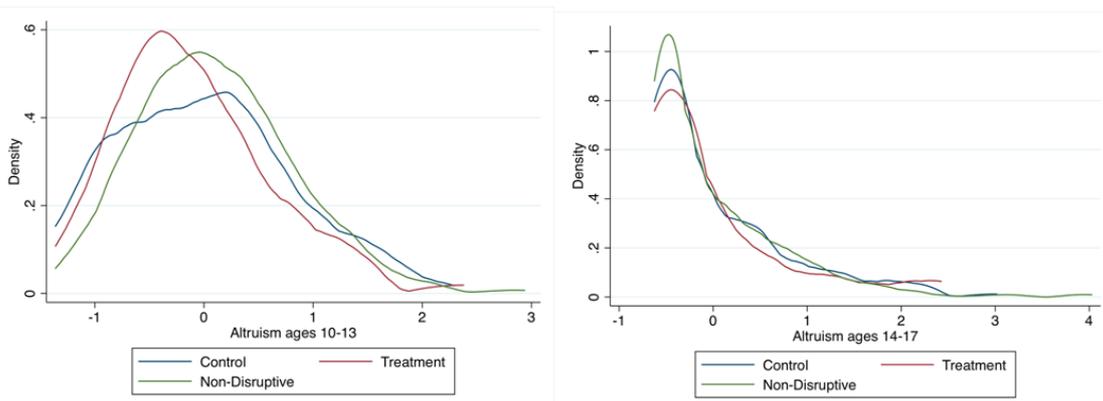
Figure S 9. Distribution of Altruism

Kolmogorov-Smirnov test for equality of Treatment and Control distributions gives a p-value of 0.27 for ages 10-13 and 0.89 for ages 14-17.

Table S 11. School Performance.

Variable	Source	Non-disruptive		Control		Treatment		Control - Treatment	
		N	mean	N	mean	N	Mean	Diff	p-value
						Ages 10-13			
Math grades	School Admin	732	0.39	158	0.00	62	0.17	0.17	0.25
French grades	School Admin	732	0.39	156	0.00	62	0.04	0.04	0.78
All grades	School Admin	734	0.38	158	-0.01	62	0.11	0.11	0.42
Special Education	School Admin	783	0.08	181	0.21	69	0.20	0.00	0.95
Grade Repetition	School Admin	783	0.11	181	0.26	69	0.26	0.00	0.96
						Ages 14-17			
Math grades	School Admin	700	0.48	154	0.00	60	0.18	0.18	0.25
French grades	School Admin	703	0.42	152	0.00	60	0.25	0.25	0.09
All grades	School Admin	705	0.44	155	-0.01	60	0.21	0.22	0.10
Special Education	School Admin	779	0.22	179	0.46	69	0.36	-0.10	0.11
Grade Repetition	School Admin	781	0.34	180	0.60	69	0.50	-0.10	0.12

Grades are standardized based on the mean and standard deviation of the control group. Grade Repetition and Special Education are real values, reflecting the proportion of each group that either repeated a grade or was in special education classes.

Figure S 10. Distribution of Grades
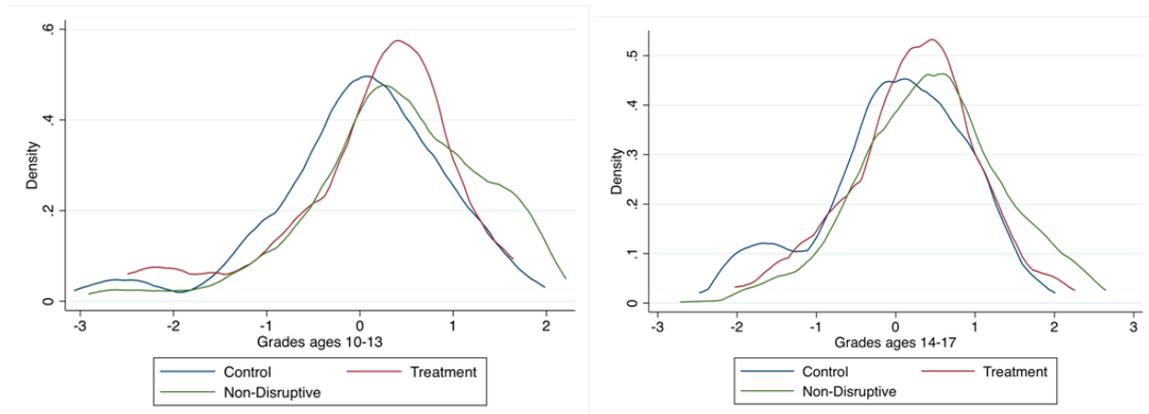
Kolmogorov-Smirnov test for equality of Treatment and Control distributions gives a p-value of 0.02 for ages 10-13 and 0.06 for ages 14-17. Note that the treatment impact on grades from age 10-13 is not significant under any other specification.

Figure S 11. Class assignment over time
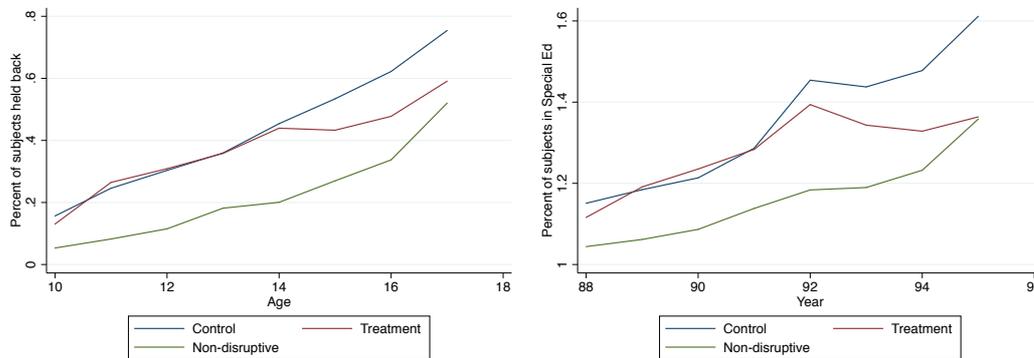
P-value from unconditional test for equality of Treatment and Control group averages for ages 10-13: 0.96 (Grade Repetition) and 0.95 (Special Education). P-value from test for equality of Treatment and Control group averages for ages 14-19: 0.12 (Grade Repetition) and 0.11 (Special Education)

Figure S 12. Distribution of Verbal IQ
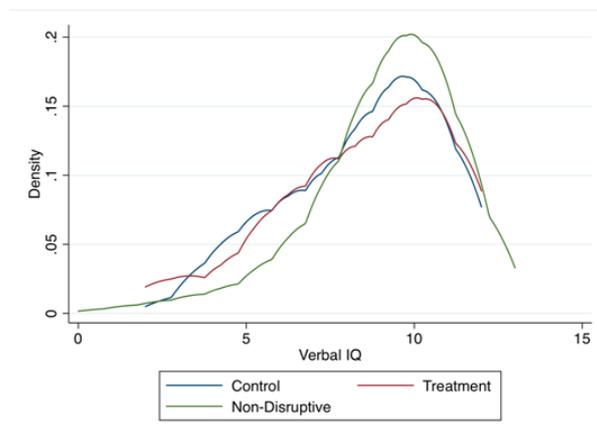
Kolmogorov-Smirnov test for equality of Treatment and Control distributions (age 13) gives a p-value of 1.00.

Table S 12. Correlation between Adolescent Outcomes

	Early Adolescent Outcomes (10-13)									
	Agg. Control	Att. Control	Trust	Friends	Self-Esteem	Altruism	Grades	Grade Repetition	Special Education	IQ
Agg. Control	1.00									
Att. Control	0.44	1.00								
Trust	0.64	0.42	1.00							
Friendliness	0.04	0.26	0.23	1.00						
Self-Esteem	0.33	0.64	0.43	0.32	1.00					
Altruism	0.14	0.11	0.19	0.13	0.03	1.00				
Grades	0.25	0.60	0.31	0.25	0.46	0.14	1.00			
Grade Repetition	-0.24	-0.31	-0.27	-0.28	-0.35	0.05	-0.29	1.00		
Special Education	-0.25	-0.28	-0.25	-0.27	-0.29	0.04	-0.28	0.89	1.00	
IQ	0.18	0.34	0.27	0.35	0.40	-0.06	0.40	-0.53	-0.48	1.00

	Late Adolescent Outcomes (14-17)									
	Agg. Control	Att. Control	Trust	Friends	Self-Esteem	Altruism	Grades	Grade Repetition	Special Education	Group Member
Agg. Control	1.00									
Att. Control	0.42	1.00								
Trust	0.67	0.41	1.00							
Friendliness	-0.05	0.08	0.06	1.00						
Self-Esteem	0.24	0.45	0.31	0.22	1.00					
Altruism	0.08	0.00	0.11	-0.05	-0.08	1.00				
Grades	0.27	0.47	0.29	0.14	0.23	0.04	1.00			
Grade Repetition	-0.27	-0.37	-0.28	-0.29	-0.33	0.20	-0.49	1.00		
Special Education	-0.28	-0.30	-0.29	-0.29	-0.33	0.22	-0.37	0.85	1.00	
Group Member	-0.01	0.03	0.00	0.00	-0.02	0.04	-0.03	0.04	0.03	1.00

Table presents the correlation for each set of early and late adolescent outcomes. Aggression Control, Attention Control, Trust, Friends, Self-Esteem, Altruism, and Grades are the z-score averages of the component variables presented in Tables S 2 – S 8, except for Grade Repetition and Special Education which are the percent of years where the subject repeated a grade or was in special education.

E.   Details on administrative tax data

Most people in Canada are required to file a tax return, even if they have zero income (and having zero income is rare, as it would require not only receiving no employment income but also no social transfers of any kind or unemployment insurance benefits). 98% of the treatment and control groups matched at least once during the 17-year period under consideration, which began in 1998, when the individuals were, in general, around 20 years old. However, for some years, subjects were not matched to tax data, potentially because they did not file a return, or potentially for other technical reasons (see Findlay et al 2018). The average match rate was around 80% by year and increasing over time (Table S 13). We cannot know why subjects might be missing tax data for a given year, and so it is not reasonable to impute zero employment income to those years. Instead, we take the average over available tax years, and do not include missing years in that average. (Note that there are individuals who report tax returns with zero employment income, since they must also report unemployment benefits, social transfers, and so on). Even though other populations may have yearly match rates over 90%, recall that this sample was originally selected from low SES areas in Montreal, and the treatment and control groups come from the most disruptive children of the original sample. Given the characteristics of this group (for example, a high school graduation rate of less than 50%) lower match rates than the general population are unsurprising.

The overall match rate between the administrative tax data from Statistics Canada and the survey data from MLES was nearly 100%. As shown in Table S 13, while this rate varied by year, in only one year was the difference in the match rate between treatment and control significant (1999), and the overall match rate is not statistically significantly different in the treatment and the control groups.

Table S 13. Match rate by group and year

	(1) Non-disruptive	(2) Control	(3) Treatment	(4)
	% of matches	% of matches	% of matches	p-value Treatment vs Control
1998	86%	81%	83%	0.72
1999	86%	78%	88%	0.06
2000	87%	74%	80%	0.35
2001	87%	80%	83%	0.65
2002	84%	76%	84%	0.15
2003	86%	77%	84%	0.21
2004	85%	84%	81%	0.59
2005	85%	82%	78%	0.46
2006	85%	82%	81%	0.83
2007	86%	75%	80%	0.45
2008	86%	78%	80%	0.76
2009	83%	80%	81%	0.78
2010	85%	77%	81%	0.46
2011	85%	79%	81%	0.71
2012	86%	80%	84%	0.42
2013	87%	80%	80%	0.98
2014	86%	83%	81%	0.67
2015	87%	83%	88%	0.33
2016	87%	83%	88%	0.33
2017	87%	80%	87%	0.21
Matched at least once	98%	98%	97%	0.75

Table shows the percent of each group that was successfully matched to Statcan administrative records each year.

To give some context of variations over time, Table S14 below presents the yearly averages for selected outcomes for the control and treatment groups, as well as the difference and the p-value of the difference.

Table S 14. Selected outcomes by year

	Employment income				
	(1) Control group mean	(2) Treatment group mean	(3) Diff. T-C (ITT)	(4) Diff T-C (%)	(5) P-value of raw diff in means
1999	19141	22713	3572	19%	0.15
2000	22169	24463	2295	10%	0.45
2001	24836	29084	4248	17%	0.20
2002	26673	31540	4867	18%	0.17
2003	25775	31636	5861	23%	0.12
2004	28569	33969	5399	19%	0.17
2005	30857	40150	9293	30%	0.04
2006	32394	41166	8772	27%	0.06
2007	35863	38499	2636	7%	0.60
2008	37419	40910	3492	9%	0.52
2009	34284	42999	8715	25%	0.09
2010	36130	43667	7537	21%	0.15
2011	38453	45759	7307	19%	0.14
2012	38406	43082	4676	12%	0.36
2013	36646	44080	7434	20%	0.15
2014	36512	46757	10246	28%	0.06
2015	34865	43970	9105	26%	0.08
2016	37508	44831	7324	20%	0.17
2017	38456	45210	6753	18%	0.22

Table shows the yearly levels of outcome for each group, in 2020 CAD. Employment income includes all individual income from employment, including self-employment.

	(1) Control group mean	(2) Treatment group mean	(3) Diff. T-C (ITT)	(4) Diff T-C (%)	(5) P-value of raw diff in means
			Social Benefits		
1999	1966	1817	-149	-8%	0.81
2000	2083	1283	-800	-38%	0.19
2001	2066	1595	-470	-23%	0.49
2002	2600	1938	-663	-25%	0.36
2003	2890	1765	-1126	-39%	0.14
2004	2400	1309	-1090	-45%	0.10
2005	1950	1614	-336	-17%	0.62
2006	2395	1395	-1000	-42%	0.15
2007	2452	1821	-631	-26%	0.42
2008	1810	1397	-413	-23%	0.50
2009	2160	765	-1395	-65%	0.03
2010	2024	939	-1085	-54%	0.10
2011	2047	301	-1746	-85%	0.00
2012	2006	1032	-974	-49%	0.12
2013	2005	828	-1177	-59%	0.06
2014	2104	1043	-1061	-50%	0.07
2015	2163	1482	-681	-31%	0.26
2016	2102	1060	-1042	-50%	0.07
2017	2055	873	-1182	-58%	0.04

Table shows the yearly levels of outcome for each group, in 2020 CAD. Social transfers are transfers such as disability benefits. Unemployment insurance is not included in social transfers.

	Contribution to professional organization				
	(1) Control group mean	(2) Treatment group mean	(3) Diff. T-C (ITT)	(4) Diff T-C (%)	(5) P-value of raw diff in means
1999	115	184	69	60%	0.07
2000	161	230	69	43%	0.19
2001	181	276	96	53%	0.10
2002	224	278	54	24%	0.39
2003	227	381	154	68%	0.03
2004	260	392	132	51%	0.08
2005	286	438	152	53%	0.05
2006	316	417	101	32%	0.24
2007	296	400	104	35%	0.21
2008	327	386	59	18%	0.50
2009	296	362	66	22%	0.44
2010	276	386	110	40%	0.19
2011	333	448	114	34%	0.22
2012	317	351	35	11%	0.70
2013	302	404	102	34%	0.24
2014	276	470	194	70%	0.03
2015	262	437	174	66%	0.04
2016	275	446	171	62%	0.05
2017	282	403	122	43%	0.16

Table shows the yearly levels of outcome for each group, in 2020 CAD. Contributions to professional organizations are annual union, professional, or similar dues. This includes malpractice insurance.

F.   Robustness checks and supplemental regressions

1.   SKILLS DATA

Since randomization was carried out at the individual level (within schools), and control and treatment participants are present in each school, neither fixed effects nor clustering are necessary, and they do not substantially change the treatment estimate or the standard errors. There were 53 schools in 1984. The average number of children in the disruptive group in each school is three. In 29 schools, there is only one child per school. The fixed effect specification presented as a robustness check includes the entire ELEM sample so that there are sufficient observations to estimate fixed effects without effectively dropping observations. We do this by including a dummy variable representing disruptive or non-disruptive, and a dummy variable for treatment (the coefficient of which would be identical to the coefficient on treatment in a specification including only the experimental group, but in practice varies slightly due to different coefficients estimated on the covariates). We present a specification that includes clustered standard errors at the level of the school in 1984 and fixed effects at the level of the school in 1984. We do not find any substantial differences in the specification with clustering and school fixed effects.

Table S 15. Robustness checks on adolescent skills data

Early adolescence	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Trust		Friendliness		Aggression		Attention		Self-esteem		Altruism	
	No school effects	School fixed effects and clusters	No school effects	School fixed effects and clusters	No school effects	School fixed effects and clusters	No school effects	School fixed effects and clusters	No school effects	School fixed effects and clusters	No school effects	School fixed effects and clusters
Treatment	0.17	0.175	-0.038	-0.05	0.143	0.143	0.187	0.176	0.0507	0.0651	-0.101	-0.0948
	(0.07)	(0.079)	(0.073)	(0.076)	(0.078)	(0.078)	(0.082)	(0.063)	(0.075)	(0.069)	(0.105)	(0.097)
Disruptive sample	-0.28	-0.3	-0.0778	-0.0937	-0.415	-0.431	-0.333	-0.37	-0.171	-0.196	-0.0805	-0.116
	(0.039)	(0.036)	(0.038)	(0.044)	(0.045)	(0.045)	(0.05)	(0.046)	(0.044)	(0.054)	(0.066)	(0.071)
Constant	0.00984	0.182	0.0259	-0.134	0.0638	-0.28	-0.14	-0.562	-0.036	-0.335	-0.00994	0.393
	(0.091)	(0.064)	(0.095)	(0.054)	(0.097)	(0.083)	(0.131)	(0.108)	(0.109)	(0.112)	(0.139)	(0.113)
Observations	1,018	1,018	1,028	1,028	1,028	1,028	1,028	1,028	986	986	1,026	1,026
R-squared	0.081	0.143	0.08	0.149	0.127	0.178	0.099	0.177	0.059	0.132	0.012	0.067

OLS, robust standard errors in parentheses. All regressions include controls for treatment group imbalances. "Disruptive sample" is an indicator variable for the sample of boys who scored above the 70th percentile in terms of disruptiveness and were randomized into treatment or control groups. "Treatment" is an indicator variable for being assigned to the treatment group. Odd numbered columns include no school effects, even numbered columns include 1984 school fixed effects and standard errors are clustered at the school level.

Table S 15 Continued

Late adolescence	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Trust		Friendliness		Aggression		Attention		Self-esteem		Altruism	
	No school effects	School fixed effects and clusters	No school effects	School fixed effects and clusters	No school effects	School fixed effects and clusters	No school effects	School fixed effects and clusters	No school effects	School fixed effects and clusters	No school effects	School fixed effects and clusters
Treatment	0.173	0.197	0.055	0.0658	0.173	0.198	0.0416	0.0622	0.00984	0.0403	-0.0754	-0.102
	(0.081)	(0.1)	(0.067)	(0.069)	(0.082)	(0.107)	(0.089)	(0.073)	(0.068)	(0.086)	(0.124)	(0.134)
Disruptive sample	-0.26	-0.278	-0.0829	-0.0795	-0.292	-0.318	-0.238	-0.26	-0.118	-0.126	0.0178	0.0251
	(0.051)	(0.054)	(0.04)	(0.044)	(0.052)	(0.062)	(0.054)	(0.048)	(0.04)	(0.043)	(0.071)	(0.078)
Constant	0.0315	-0.129	0.154	0.175	0.0474	0.0235	-0.0813	0.00308	0.105	0.301	-0.243	0.194
	(0.096)	(0.11)	(0.1)	(0.11)	(0.098)	(0.114)	(0.124)	(0.154)	(0.104)	(0.085)	(0.194)	(0.174)
Observations	919	919	919	919	919	919	914	914	888	888	886	886
R-squared	0.06	0.111	0.059	0.144	0.081	0.138	0.051	0.1	0.036	0.101	0.01	0.093

OLS, robust standard errors in parentheses. All regressions include controls for treatment group imbalances. "Disruptive sample" is an indicator variable for the sample of boys who scored above the 70th percentile in terms of disruptiveness and were randomized into treatment or control groups. "Treatment" is an indicator variable for being assigned to the treatment group. Odd numbered columns include no school effects, even numbered columns include 1984 school fixed effects and standard errors are clustered at the school level.

Table S 15 continued

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
	Grades (early)		Held Back (early)		Special Ed (early)		Verbal IQ		Grades (late)		Held Back (late)		Special Ed (late)	
	No school effects	School fixed effects and clusters	No school effects	School fixed effects and clusters	No school effects	School fixed effects and clusters	No school effects	School fixed effects and clusters	No school effects	School fixed effects and clusters	No school effects	School fixed effects and clusters	No school effects	School fixed effects and clusters
Treatment	0.184	0.173	-0.0312	-0.0342	-0.0219	-0.0172	0.145	0.255	0.286	0.307	-0.141	-0.149	-0.138	-0.152
	(0.14)	(0.131	(0.052)	(0.053)	(0.048)	(0.051)	(0.381)	(0.38)	(0.126)	(0.126)	(0.061)	(0.069)	(0.061)	(0.056)
Disruptive sample	-0.34	-0.37	0.139	0.152	0.11	0.122	-0.455	-0.605	-0.395	-0.444	0.223	0.245	0.214	0.236
	(0.082)	(0.0795	(0.029)	(0.023)	(0.025)	(0.020)	(0.2)	(0.208)	(0.080)	(0.089)	(0.033)	(0.030)	(0.034)	(0.030)
Constant	-0.292	-0.427	0.0916	-0.0383	0.105	-0.0043	8.882	7.942	-0.218	-1.481	0.575	0.44	0.335	0.325
	(0.197)	(0.16	(0.063)	(0.054)	(0.057)	(0.050)	(0.48)	(0.351)	(0.184)	(0.182)	(0.080)	(0.064)	(0.078)	(0.087)
Observations	954	954	1,033	1,033	1,033	1,033	885	885	920	920	1,030	1,030	1,027	1,027
R-squared	0.101	0.166	0.114	0.176	0.076	0.147	0.058	0.142	0.11	0.175	0.16	0.222	0.128	0.188

OLS, robust standard errors in parentheses. All regressions include controls for treatment group imbalances. "Disruptive sample" is an indicator variable for the sample of boys who scored above the 70th percentile in terms of disruptiveness and were randomized into treatment or control groups. "Treatment" is an indicator variable for being assigned to the treatment group. Odd numbered columns include no school effects, even numbered columns include 1984 school fixed effects and standard errors are clustered at the school level. Early and Late refer to the early or late adolescent periods.

2.   TAX DATA

Income data presents concerns since extreme values, especially in a small sample such as this, may have a potentially large influence on the estimate of impact, leading to results which are driven by a handful of observations. The alternate specifications tested below give some comfort that this is not the case. In addition, our simplest measure of impact is based on the average over time, ignoring missing years. However, there are other strategies for dealing with missing data, and it is important to make sure that our results do not depend on how missing data is treated. In Table S 16, we present the results from several different specifications for the variables measuring money: log transformation, upper and lower Lee bounds, imputation of group means for missing years, Windsorized income data, inclusion of fixed effects for the missing years, a panel specification, and a median regression. We also note that when the data are transformed into an even simpler measure, a binary variable equal to one if any income is reported or zero if not, the results still hold as presented in Table 6.

Table S 16. Robustness checks on tax data

	(1) School level effects	(2) Missing year effects	(3) Log	(4) Windsorized data	(5) Lower Lee Bound	(6) Upper Lee Bound	(7) Group mean imputed for missing values	(8) Median regression	(9) Panel
Employment income 1998-2017	7421	8387	1.3	7683	7504	8137	8091	6340	8526
	(3303)	(3731)	(0.4)	(3223)	(3287)	(3408)	(3414)	(5109)	(3528)
Pension contribution	80	87	0.1	117	-9	53	48	105	40
	(178)	(206)	(0.3)	(149)	(178)	(194)	(193)	(73)	(206)
Amount of social benefits 1998-2017	-1212	-1458	-1.1	-1091	-1437	-1322	-1318	-218	-1286
	(459)	(484)	(0.4)	(343)	(401)	(425)	(423)	(62)	(435)
Amount contributed to professional org	106	130	0.8	114	108	130	129	120	134
	(51)	(64)	(0.3)	(50)	(53)	(57)	(57)	(46)	(62)
Contributions to unemployment insurance	84	101	0.8	101	99	102	102	105	104
	(45)	(48)	(0.3)	(43)	(43)	(44)	(43)	(73)	(45)
Charity	36	33	0.0	0	2	31	32	0	35
	(25)	(24)	(0.1)	(1)	(4)	(22)	(22)	(1)	(25)
Tuition	10	-2	0.1	6	-58	6	7	0	9
	(53)	(52)	(0.1)	(12)	(26)	(42)	(42)	(4)	(45)

Tax data from Statcan, 1998-2017. Each row is an outcome variable, and within each row each column gives the coefficient on treatment for a different specification, with standard errors in parentheses. Except for the panel specification, the outcome variables are the averages over the time period. Column (1) uses 1984 school fixed effects and clustered standard errors. Column (2) includes fixed effects for any missing years. Column (3) transforms the outcome variable using a log. Column (4) uses Winsorized data, where the data above the 95th or below the 5th percentile are replaced with the 95th or 5th percentile value, respectively. Columns (5) and (6) provide the lower and upper Lee bounds, where group (treatment or control) with lower attrition is trimmed either from the top or bottom of the distribution to generate equivalent attrition in the two groups. Column (7) imputes the group (treatment or control) mean in a given year for any missing values. Column (8) uses a median regression. Columns (5) through (8) have 245 observations. Column (9) uses the dataset in a panel form, with year fixed effects (4022 observations). All specifications includ controls for baseline imbalences.

G. Disentangling impact

1. PARENT PROGRAM COMPONENTS

In addition to information on adolescent behavior and attitudes, we use participant-reported variables on parent behavior. As with the adolescent outcomes, we use exploratory factor analysis to group variables into four types of behavior, the component variables of which are shown in Table S 17-Table S 20, and summary statistics are reported in Table S 21. Figures for the distributions of the parent behaviors are available on request.

Relationship includes variables on the quality of the relationship between the parents and the adolescent: the quality of communication, whether the parent acts warmly towards the adolescent, and whether they go on family outings.

Punitive Behavior includes variables on punitive and even abusive behavior: the adolescent feels rejected; the parents punish the adolescent by hitting, mocking, calling names or forbidding things; and the parents' rules are considered unjust by the adolescent.

Rules includes variables on in the rules of the household: curfew; hours of TV watching; homework; having friends over; eating dinner together; and going out with friends.

Supervision includes variables on how closely the parents supervise the adolescent: whether there is generally a parent at home and whether the parents know where and with whom the adolescent is.

Table S 17. Components of Parent Punitive Index.

Variable	Source	Non-disruptive N	Non-disruptive mean	Control N	Control mean	Treatment N	Treatment Mean	Control - Treatment Diff	Control - Treatment p-value
					Ages 10-13				
Rejected by Parent	Subject	764	-0.27	171	0	66	0.03	0.03	0.81
Parent hits	Subject	769	-0.20	173	0	66	-0.1	-0.1	0.47
Parent forbids things	Subject	764	-0.08	171	0	66	0.03	0.03	0.86
Parent mocks	Subject	764	-0.13	171	0	66	-0.07	-0.07	0.63
Parent calls names	Subject	764	-0.32	171	0	66	-0.23	-0.23	0.12
Parent is unjust	Subject	764	-0.08	171	0	66	0.07	0.07	0.63
					Ages 14-17				
Rejected by Parent	Subject	700	-0.35	150	0	59	-0.09	-0.09	0.60
Parent hits	Subject	700	-0.18	150	0	59	-0.22	-0.22	0.13
Parent forbids things	Subject	700	-0.09	150	0	59	0.01	0.01	0.92
Parent mocks	Subject	700	-0.10	150	0	59	0.13	0.13	0.39
Parent calls names	Subject	700	-0.26	150	0	59	-0.12	-0.12	0.41
Parent is unjust	Subject	700	-0.10	149	0	59	0.06	0.06	0.68

Variables are standardized based on the mean and standard deviation of the control group.
Variables are oriented so that more positive values indicate better outcomes (less punitive).

Table S 18. Components of Parent Rules Index.

Variable	Source	Non-disruptive N	Non-disruptive mean	Control N	Control mean	Treatment N	Treatment Mean	Control - Treatment Diff	Control - Treatment p-value
					Ages 10-13				
Curfew	Subject	764	0.08	171	0	66	-0.01	-0.01	0.94
Rule about TV	Subject	764	-0.1	171	0	66	-0.02	-0.02	0.91
Rule about Homework	Subject	763	-0.1	171	0	66	0.11	0.11	0.45
Rule about Friends	Subject	764	-0.2	171	0	66	-0.20	-0.20	0.16
Rule about Dinner	Subject	764	-0.1	171	0	66	0.10	0.10	0.50
Rule about Going Out	Subject	764	-0.18	171	0	66	-0.06	-0.06	0.69
					Ages 14-17				
Curfew	Subject	699	0.11	150	0	59	0.20	0.20	0.19
Rule about TV	Subject	699	-0.05	150	0	59	-0.02	-0.02	0.88
Rule about Homework	Subject	699	-0.14	149	0	59	-0.06	-0.06	0.68
Rule about Friends	Subject	699	-0.08	150	0	59	0.03	0.03	0.83
Rule about Dinner	Subject	699	-0.04	150	0	59	-0.03	-0.03	0.83
Rule about Going Out	Subject	699	-0.09	150	0	58	0	0	0.98

Variables are standardized based on the mean and standard deviation of the control group.
Variables are oriented so that more positive values indicate more rules.

Table S 19. Components of Parent Supervision Index.

Variable	Source	Non-disruptive N	mean	Control N	mean	Treatment N	Mean	Control - Treatment Diff	p-value
					Ages 10-13				
Parent at home at night	Subject	764	0.07	171	0	66	0.15	0.15	0.29
Parent knows where I am	Subject	769	0.28	173	0	66	0.02	0.02	0.89
Parent knows who I'm with	Subject	769	0.21	173	0	66	-0.07	-0.07	0.61
					Ages 14-17				
Parent at home at night	Subject	700	-0.17	149	0	58	-0.08	-0.08	0.59
Parent knows where I am	Subject	700	0.24	149	0	58	0.22	0.22	0.15
Parent knows who I'm with	Subject	700	0.24	150	0	58	0.18	0.18	0.23

Variables are standardized based on the mean and standard deviation of the control group.
Variables are oriented so that more positive values indicate more supervision.

Table S 20. Components of Parent Relationship.

Variable	Source	Non-disruptive N	mean	Control N	mean	Treatment N	Mean	Control - Treatment Diff	p-value
					Ages 10-13				
Family outings	Subject	764	0.01	171	0	66	-0.18	-0.18	0.22
Talk future with parent	Subject	764	0.07	171	0	66	-0.11	-0.11	0.44
Parent explains rules	Subject	763	0.33	171	0	66	0.02	0.02	0.89
Talk thoughts with parent	Subject	764	0.13	171	0	66	-0.14	-0.14	0.32
Parent explains requests	Subject	763	0.17	171	0	66	-0.18	-0.18	0.23
Talk feelings parent	Subject	764	0.06	171	0	66	-0.23	-0.23	0.09
Parent encourages	Subject	764	0.24	171	0	66	-0.09	-0.09	0.52
					Ages 14-17				
Family outings	Subject	696	0.06	148	0	59	0.02	0.02	0.87
Talk future with parent	Subject	699	0.08	150	0	59	-0.08	-0.08	0.6
Parent explains rules	Subject	700	0.20	150	0	59	0.15	0.15	0.29
Talk thoughts with parent	Subject	700	0.10	150	0	59	-0.03	-0.03	0.85
Parent explains requests	Subject	699	0.23	149	0	59	-0.02	-0.02	0.89
Talk feelings parent	Subject	700	0.26	149	0	59	0.16	0.16	0.3
Parent encourages	Subject	700	0.15	150	0	59	-0.08	-0.08	0.61

Variables are standardized based on the mean and standard deviation of the control group.
Variables are oriented so that more positive values indicate better outcomes (better family relationships).

Table S 21. Parent Behavior Summary Statistics.

	Num. Var	Alpha	Non-Disruptive mean	N	Disruptive: Control mean	N	Disruptive: Treatment mean	N
				Ages 10-13				
Parent Relationship	7	0.76	0.14	764	0	171	-0.13	66
Parent Punitive	6	0.71	-0.18	764	0	171	0.05	66
Parent Rules	6	0.69	-0.10	764	0	171	-0.01	66
Parent Supervision	3	0.29	0.19	769	0	173	0.03	66
				Ages 14-17				
Parent Relationship	7	0.82	0.15	700	0	150	0.02	59
Parent Punitive	6	0.75	-0.18	700	0	150	-0.04	59
Parent Rules	6	0.58	-0.05	699	0	150	0.02	59
Parent Supervision	3	0.64	0.10	700	0	150	0.11	58

Behaviors are the z-score averages of the component variables, except for Held Back and Special Ed, which are the percent of years where the subject was held back or in special education. All p-values are two sided and robust to inclusion of control variables.

We find no impact of treatment on any of the parent behavior in either the early or late adolescent period (Table S 22). If anything, there is a small negative impact on parent relationships, but this is significant in only one specification and the direction of the effect is not consistent over time. We cannot exclude the possibility that the program had an impact on some aspect of the parent relationship not mentioned here but given the lack of impact on the behaviors we do have data on it seems unlikely that changes in parent behavior are driving the impact of the program.

Table S 22. Treatment Impact on Parent Behavior.

| | Early Adolescence (10-13) | | | | | | | | |
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	p-value of raw difference in means (t-test, Disruptive Sample)	OLS Treatment Effect: Disruptive Sample	Conditional Treatment Effect: Disruptive Sample	Treatment Effect: Full Sample	Conditional Treatment Effect: Full Sample	Conditional Treatment effect with full sample, Clustered Standard Errors	Conditional Treatment effect with full sample, Fixed Effects	Conditional Treatment effect with full sample, IPW	P-value of raw difference in means (permutation test, Disruptive Sample)
Relationship	0.16	-0.131	-0.147*	-0.131	-0.123	-0.123	-0.138	-0.129	0.13
		(0.0894)	(0.0867)	(0.0905)	(0.0892)	(0.0872)	(0.0942)	(0.0907)	
Punitive	0.60	-0.0474	-0.0220	-0.0474	-0.0405	-0.0405	-0.0439	-0.0407	0.575
		(0.0862)	(0.103)	(0.0852)	(0.0952)	(0.0925)	(0.0949)	(0.0920)	
Rules	0.88	-0.0125	-0.0717	-0.0125	-0.0595	-0.0595	-0.0462	-0.0610	0.900
		(0.0947)	(0.0945)	(0.0986)	(0.0891)	(0.103)	(0.0943)	(0.0922)	
Supervision	0.76	0.0273	-0.00237	0.0273	-0.00370	-0.00370	-0.00270	-0.0105	0.762
		(0.0824)	(0.0852)	(0.0911)	(0.0880)	(0.104)	(0.0892)	(0.0879)	
	Late Adolescence (14-17)								
Relationship	0.85	0.0192	0.0261	0.0192	0.0267	0.0267	0.0307	0.00773	0.842
		(0.101)	(0.102)	(0.106)	(0.111)	(0.0997)	(0.107)	(0.104)	
Punitive	0.72	-0.0355	-0.0220	-0.0355	-0.0194	-0.0194	-0.0202	-0.0116	0.681
		(0.0942)	(0.0960)	(0.0901)	(0.0957)	(0.0863)	(0.0915)	(0.0954)	
Rules	0.81	0.0206	-0.00537	0.0206	0.000113	0.000113	-0.00657	0.0134	0.806
		(0.0880)	(0.0915)	(0.0958)	(0.0944)	(0.0827)	(0.0894)	(0.0876)	
Supervision	0.36	0.110	0.0350	0.110	0.0450	0.0450	0.0348	0.0503	0.300
		(0.114)	(0.119)	(0.117)	(0.120)	(0.0933)	(0.129)	(0.117)	

Each cell of column (1) provides the p-value for the raw difference between the treatment and the control group for each of the outcomes (in rows). Each cell of columns (2)-(8) gives the regression coefficient of the treatment dummy variable on each of the outcomes (in rows). Columns (2)-(8) include robust standard errors in parentheses. *** p<0.01, ** p<0.05, *p<0.10. Columns (1)-(3) and (9) use data from the disruptive sub-sample only. Columns (4)-(8) use data from the entire sample. Columns (1), (2), (4), and (9) include no controls. Columns (3) and (5)-8) include controls for imbalances between the treatment and control groups: age of father, prestige of mother's work, and initial anxiety. Values for these variables are imputed for a handful of observations based on the predicted mean from all complete baseline variables, and the regressions include a dummy variable when variables are imputed. Column (8) uses inverse probability weighting to adjust for attrition under certain assumptions. The p-values in column (9) are obtained from a permutation test of the difference of the means (where the permuted value is treatment group) with 2000 repetitions. Clustered standard errors and fixed effects are at the school level in 1984.

2.   TEACHER PROGRAM COMPONENTS

A third component of the intervention was two meetings with the classroom teachers. 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 target boys. However, compliance with this aspect of the intervention was very weak, as only half of the teachers agreed to participate. Since the intervention was very light (two meetings) and compliance was very low, it seems unlikely that the teacher training was the part of the intervention that was driving the impact.

3.   SUGGESTIVE EVIDENCE ON MECHANISMS OF IMPACT

While we cannot fully disentangle different channels of impact due to correlation between the measured skills, we follow a simple procedure to provide suggestive evidence on this question: we observe treatment impact on the potential channel, the correlation between the potential channel and the outcome and the extent to which the estimated treatment effect is reduced when a potential channel is included as a covariate.[3] That is, if a skill 1) is changed by the treatment, 2) is correlated to the outcome and 3) accounts for some of the variation in the outcome induced by the treatment, we take this as suggestive evidence that that the skill is functioning as a channel of impact.

Table S 23 shows the correlation of cognitive and non-cognitive skills (described in Section 2.3.2) used in this paper[4] to the principal outcomes, including criminality and secondary school completion as detailed in Boisjoli et al. (2007), and Vitaro et al. (1999, 2012). Higher non-cognitive and cognitive skills are positively related to better social and economic outcomes. Columns (1) and (2) show that two of the pro-social variables – Trust and Friendliness – are associated with stronger educational outcomes (less grade repetition and special education assignment, and higher graduation rates), positively related to employment and earnings, and negatively related to social transfers. The correlation with Trust is substantially higher than with Friendliness for criminal records and years with employment income. Column (3) shows that

---

[3] Heckman et al. (2013) carry out a formal decomposition of impact of the Perry Preschool Program, finding that changes in non-cognitive skills explain the bulk of the impact of the program. In their strategy, the latent skills were orthogonal to one another. In this paper, we are examining sub-sets of non-cognitive skills, which are by their nature correlated (for example, the correlation between Trust and Attention Control is around 0.4) and so decomposing the impact is impossible.

[4] Vergunst et al. (2019) estimate the correlation between behaviors measured by the initial 1984 screening evaluation filled out by teachers. The skills used in this paper are different.

Altruism is not related to any of the outcomes. This underscores that "pro-social" skills should be differentiated into separate behaviors. Columns (4) and (5) show that for self-control, higher Attention Control and Aggression Control are both related to better schooling outcomes, less crime, more employment, and less reliance on social transfers. The correlation of Attention Control with the education outcomes is higher (corresponding to findings in Vitaro et al. 1999), and the correlation of Aggression Control with Criminal Records is higher. Self-esteem (column (6)) and IQ (column (7)) are also positively related to better outcomes.

Table S 23. Correlation between potential channels and selected outcomes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				Aggression	Attention	Self-	
	Trust	Friendliness	Altruism	Control	Control	esteem	IQ
Years of special education	-0.36	-0.29	0.00	-0.35	-0.48	-0.47	-0.50
Ever repeated a grade	-0.29	-0.21	-0.06	-0.29	-0.48	-0.39	-0.34
Years of grade repetition	-0.37	-0.31	-0.02	-0.35	-0.55	-0.51	-0.52
Grades	0.32	0.15	0.06	0.32	0.51	0.35	0.30
Criminal record	-0.22	-0.06	-0.03	-0.30	-0.21	-0.15	-0.12
Secondary completion	0.32	0.24	0.03	0.34	0.51	0.41	0.40
Group membership	0.08	0.10	0.06	0.04	0.15	0.14	0.06
Years with employment income	0.24	0.10	0.08	0.27	0.28	0.19	0.24
Ever married	0.04	0.03	0.00	0.04	0.10	0.08	0.04
Years with social transfers	-0.19	-0.24	-0.08	-0.16	-0.25	-0.25	-0.32
Years with unemployment benefits	0.02	0.01	0.03	0.05	-0.02	0.02	0.03
Years contributing to prof. org.	0.05	0.02	0.07	0.14	0.13	0.09	0.11
Average earnings from employment	0.25	0.19	0.07	0.26	0.36	0.27	0.28
Average social transfers	-0.22	-0.20	-0.08	-0.20	-0.26	-0.25	-0.33
Average unemployment benefits	-0.01	0.01	0.03	0.01	-0.04	0.01	0.07
Average contributions to employment insurance	0.23	0.14	0.07	0.25	0.30	0.22	0.26

Each cell shows the correlation between the outcome (rows) and the potential channels (columns). The sample is all subjects including non-disruptive.

We provide suggestive evidence on channels of impact using a simple test where we compare the size of the coefficient on treatment with and without the skill as a covariate.[5] A reduction in the size of the treatment coefficient implies that some of the impact of treatment is

---

[5] The regression without the covariate is restricted to the sample that has full data for that covariate to avoid conflating changes in the coefficient due to changes in the sample with changes in the coefficient due to inclusion of the covariate.

explained by the treatment's impact on that skill. Table S 24 provides the reduction in the treatment coefficient when the potential channel is included as a covariate. Treatment impacts on Attention Control, which had the highest correlation to academic outcomes including secondary school completion, explain a larger percentage of the treatment impact on those outcomes than the other skills (around 34% in each case). Aggression Control, on the other hand, has the highest correlation with the criminal record, and treatment impacts on this skill explain a larger percent of the treatment impact on that outcome than the other skills (34%). Finally, treatment impacts on Attention Control and Trust explain the highest percentof the treatment impact on earnings and group membership (11% and 18%, respectively).

Table S 24. Treatment impact explained by potential channels

	(1)	(2)	(3)
	% explained by attention control	% explained by aggression control	% explained by trust
Years of special education	34%	17%	27%
Ever repeated a grade	28%	15%	21%
Years of grade repetition	35%	16%	26%
Secondary school completion	34%	16%	25%
Criminal record	18%	34%	15%
Group membership	2%	2%	11%
Ever married	4%	5%	6%
Average union dues	6%	3%	-4%
Average social transfer	9%	4%	10%
Average earnings	15%	6%	18%

Columns (1), (2), and (3) show the reduction, in percent, of the coefficient on treatment when the potential channel is included as a covariate. For example, the coefficient of treatment on special education is reduced by 34% when Attention Control is included as a covariate.

This pattern fits well with our theoretical understanding of which types of behavior are related to which adult outcomes: attention and concentration in the classroom is related to school outcomes, trust and attention are related to labor market and social outcomes, and delinquency and aggression are related to criminal behavior.

H.   Cost-benefit analysis

1. SOURCES FOR COST AND BENEFIT ESTIMATES

In order to provide information about the adult impact of investment in childhood behavioral interventions, we compare the cost of the intervention to the impact of the intervention under a set of conservative assumptions and provide a sensitivity analysis. We estimate the cost of the program based on known staff costs, as the principal cost of the program was the salary of the trainers and implementers, and no other inputs were used. The implementation team was composed of one full time social worker, two full time childcare specialists (BA level), one psychologist, and one half-time program administrator over the course of the program (two years). We do not include the cost of evaluation or questionnaires. We use median reference hourly wages for these professions in Quebec in 2011 (Statistics Canada, 2011), adjusted for inflation to 1985 (half a year), 1986 (full year), and 1987 (half a year). We assume 40 hours per week paid for 52 weeks. We assume that other costs amount to 30% of salaries (photocopies, transport, training, and so on). Under these assumptions, the total program cost per offer was around $ 10,855 in 2020 CAD.[6] In terms of overall program costs and potential target populations, recall that the impact estimates are based on the group that was initially targeted, that is, the most disruptive 30% of boys from low SES schools, so the cost of a program based on this evaluation and its target group would be of a smaller magnitude than one that targeted the entire population or a larger group.

We monetize the benefits of reduced grade repetition and special education, based on the cost avoided (additional years of schooling for repeaters and additional costs of the services required for special education). The cost of education per student per year (representing the cost of repeating a grade) was $ 7,507 in 2008 CAD (Ministère de l'Éducation, du Loisir et du Sport 2014).[7] Information on the cost of special education during this period is not available; we assume that special education assignment cost an additional 50% per student, or $ 3,754 in 2008

---

[6] Our estimate of the total cost per person using 2011 data on salaries was $9,327 in 2011 CAD. To provide a figure closer to current policy budgets, we convert to 2020 CAD using the average inflation rate over this period (1.7%) and do not account for the time value of money. This yields $10,855 in 2020 CAD, which is an estimate of what it would cost in 2019 to implement such a program. This is in contrast to the comparison of costs and benefits, which must account for the time value of money (using a 3% discount rate), and so the cumulative cost of the program varies over time. This is because, for the cost-benefit analysis, we take the cost of foregoing other possible uses of money into account. We also use the discount rate of 3% on benefits, both monetary and non-monetary, to reflect this time preference. The sensitivity analysis in Table 9 provides estimates under 2% and 5% discount rates.

[7] We assume that the additional cost of repetition is incurred in 1996, the year after the subjects would have graduated had they never repeated (and so represents remaining in school for one more year), and the price in 1996 CAD was approximately $ 5,769. We assume one additional year of schooling per repeater.

CAD.[8] Since no data on school outcomes was collected after age 17, we assume benefits in terms of reduced spending stop after age 17.

For benefits due to reduced crime, we use the estimated reduction in the number of crimes from age 18-25 (about one fewer crime per person in the treatment group over that period). For simplicity, we assume that the impact was spread evenly over the seven years of 18-25, or a reduction of about 0.16 crimes per person per year. We assume that the treatment impact fades out over time (see Sampson and Laub 2003) at a rate of 10% per year after age 25 (so that at age 26, the treatment group commits 0.14 fewer crimes, at age 27, 0.13 fewer crimes, and so on) and that by age 35 there is no difference between the treatment and the control group. We then compare the cumulative discounted number of crimes avoided to the discounted cost at age 35.

To provide a lower bound on the monetary benefits due to reduced crime we use the administrative costs of crime (cost of arrest, holding, court time, and administration). This is because the treatment impact seems to be driven by reductions in non-violent crime, so benefit from reduced costs to victims would be small.[9] This estimate does not include the costs of policing, which would substantially increase the estimate of the monetary benefit. Based on information from the Canadian Department of Justice, these costs were $ 1,912 per crime in 2003 CAD (Canada Department of Justice 2008).[10]

For earnings and social transfers, the estimated treatment impact is an average over ages 20-49, and we assume that it is the same in each year from ages 20-39, or $ 5,708 2020 CAD per year for earnings and $ 929 2020 CAD per year for social transfers.

2. CALCULATION AND SENSITIVITY ANALYSIS

Time t begins in 1985, the first year of the program. In 1985, 1986, and 1987 (but not after) there are only costs of the program, and measured benefits begin in $t=7$, or 1992 when the subjects were 14 years old. For each time t, we calculate the value of the accumulated benefit due to reduced crime from that period and the previous periods as

---

[8] Approximately $ 2,800 in 1995 CAD (the relevant period was 1992-1995).

[9] Monetary benefits due to reduced criminal behavior are difficult to estimate, as they include not only the cost of enforcement and, if applicable, incarceration, but also the cost to society of the crime committed. Different strategies to monetize the cost of crime are using jury awards as an estimate of the monetary damage due to different types of crimes (Heller et al. 2017), or the value of a statistical life (for murders) and victim assault costs (Heckman et al. 2010). Our data do not include the type of crime committed, so we cannot do this type of analysis.

[10] Adjusting for inflation to the year that the crime reductions were first documented, 1996, gives a figure of $ 1,659.

$$AB_t^{crime} = (\gamma_t^{crime} \times p_t^{crime}) + (AB_{t-1}^{crime}(1+\delta)) \qquad (1)$$

where γ_t^{crime} is the number of crimes avoided in time t due to the treatment, p_t^{crime} is the value of avoiding a crime in time t, AB_{t-1}^{crime} is the total accumulated benefit in the previous time t-1, and δ is the discount rate of 0.03. Equivalent calculations are carried out for the benefits of reduced repetition and special education assignment. For employment income, the accumulated benefit in time t is

$$AB_t^{income} = (\gamma_t^{income}) + (AB_{t-1}^{income}(1+\delta)) \qquad (2)$$

where γ_t^{income} is the additional income in time t due to the treatment. An equivalent calculation is carried out for the accumulated benefit in each period due to the reduction in social transfers. Income and social transfer benefits begin in 1998, or t=13. The accumulated benefit in each year is then

$$AB_t = \left(AB_t^{repetition} + AB_t^{specialed} + AB_t^{crime} + AB_t^{income} + AB_t^{transfers} \right) \qquad (3)$$

This sum of accumulated benefits in each year is compared to the accumulated cost of the program in that year t, which is equal to the time-discounted value of the program expenditures in 1985, 1986 and 1987 (or, t=0 through t=2).[11] We provide results until age 39 (the last year for which we have tax data available).

By calculating accumulated benefits and accumulated costs at each time period, we can examine returns under different assumptions. Table S 25 provides results under different interest rates.

Table S 25. Sensitivity analysis

	Decline in economic returns	Discount rate	Dollars in benefits per dollar in costs
Overall benefits (base case)	0	3%	$11
Overall benefits	0	2%	$15
Overall benefits	0	5%	$6
Taxpayer benefits only (base case)	0	3%	$2
Taxpayer benefits only	0	2%	$3
Taxpayer benefits only	0	5%	$1

---

[11] The program was two years long, beginning in the fall of 1985 and ending in the spring of 1987 (in line with the school calendar). Half a year of cost is allocated to 1985 and 1987, and a full year of cost is allocated to 1986.

Table shows comparison of cumulative costs and benefits under different discount rate assumptions. Overall benefits includes increased earnings, reduced social transfers, reduced schooling costs (repetition and special education) and reduced criminality. Taxpayer benefits exclude increased earnings as a benefit. Discount rate is applied to both costs and benefits.

I. References for Online Appendix

American Psychiatric Association, *The Diagnostic and Statistical Manual of Mental Disorders: DSM 5* (2013).

L. Bertrand, D. Reclus-Prince, Social Skills Training. The Montreal Longitudinal and Experimental Study: The Intervention Program and its Long Term Effects. In R. Tremblay (Ed.), *The Montreal Longitudinal Experimental Study: The Intervention and its Long Term Effects*, 51-74. (available at *https://gripinfo.ca*).

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).

K. Dodge, Do social information-processing patterns mediate aggressive behavior? In B. B. Lahey, T. E. Moffitt, A. Caspi (Eds), *Causes of conduct disorder and juvenile delinquency*. New York, NY, US: Guilford Press (2003)

K. Dodge, J. Godwin, 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 (2013).

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

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) 249-277.

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

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).

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

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*. Castalia, Eugene, Oregon (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, Oregon (1975).

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).

Statistics Canada. Table 14-10-0307-01 Employee wages by occupation, annual, inactive. https://doi.org/10.25318/1410030701-eng

R. E. Tremblay. The Montreal longitudinal and experimental study. In R. Tremblay (Ed.), *The Montreal Longitudinal Experimental Study: The Intervention and its Long Term Effects*,. (available at *https://gripinfo.ca*) (2010).

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

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).

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).