The Right Timing Matters: Identifying Sensitive Periods in the Formation of Socio-Emotional Skills*

Laura Breitkopf[†], Shyamal Chowdhury[¶‡], Daniel A. Kamhöfer[†‡], Hannah Schildberg-Hörisch[†‡§], and Matthias Sutter[§#✠♠]

[†] *Heinrich Heine University Düsseldorf, Düsseldorf Institute for Competition Economics*

[¶] *The University of Sydney, School of Economics*

[§] *Max Planck Institute for Research on Collective Goods, Bonn*

[‡] *IZA Institute of Labor Economics, Bonn*

[#] *University of Cologne*

[✠] *University of Innsbruck*

[♠] *CESifo, Munich*

March 10, 2022

preliminary, please do not quote without permission

Abstract

In a randomized controlled trial (RCT) with about 3,200 children in grades 2 to 5 in Bangladesh we examine the existence of sensitive periods in the formation of socio-emotional skills. In sensitive periods, returns to investments into skills are particularly high. Using a novel design for identifying sensitive periods, we implement the same intervention in different school grades to study its age-specific treatment effects on children's self-control, patience, and prosociality. Our results identify sensitive periods in the formation of self-control and patience around the ages 7 and 8, while prosociality appears to be similarly malleable throughout ages 7 to 11.

Keywords: Sensitive periods, skill formation, randomized controlled trial, self-control, patience, prosociality, social and emotional learning program, experiments with children, Bangladesh

JEL Classifications: *C93, D01, D64, D90, J13, J24*

*Financial support from the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) through grant no. SCHI 1377/1 and under Germany's Excellence Strategy (EXC 2126/1-390838866) as well as from the Lions Clubs International Foundation is gratefully acknowledged. IRB approval (Heinrich Heine University Düsseldorf) was granted under study number 6212. The intervention was preregistered at the AEA RCT Registry (RCT ID AEARCTR-0003129). For useful comments and discussions, we are grateful to Sarah Cattan, Fabian Kosse, Petter Lundborg, Gerhard Riener, Joachim Winter, and Jonathan Zinman as well as seminar participants in Düsseldorf and at GIGA, Hamburg. In addition, the paper benefited from feedback of many conference participants at the IZA/ECONtribute Human Capital Workshop 2021, M-BEES 2021, IAREP/SABE 2021 Conference, EALE Conference 2021, and VfS 2021 Annual Conference. We thank the officials of the Bangladesh Ministry of Primary and Mass Education, especially the then Minister Adv. Mostafizur Rahman Fizar for his whole-hearted support in the implementation of the program in public primary schools, all LQ trainers from India, and translators and editors from Bangladesh and India, especially Israt Zerin, for her tireless efforts. We also like to thank Arifur Rahman, Rabbi Rahim, Arun Jyoti and Delowar Hossain for their assistance in the implementation of the LQ SfG program in the field.

1 Introduction

The model of skill formation by Cunha and Heckman (2007, 2008) is the seminal theoretical contribution to the development of children's cognitive and socio-emotional skills in economics. In this model, skills are the product of genetic and environmental initial conditions at conception, parental characteristics (e.g., IQ, genes, education, income), and parental or public investments in children. Skill formation is modeled as a dynamic, multistage process: children's skills change over time as the result of accumulating investments and exhibit both self-productivity and complementarity.[1]

A key assumption of the model is the existence of sensitive periods for the development of each skill. Sensitive periods are those maturational stages in which investments are especially productive.[2] While cognitive skills like IQ are most malleable in early childhood years with sensitive periods likely below age 3 (Shonkoff and Phillips, 2000; Knudsen et al., 2006; Heckman and Mosso, 2014), research has not yet isolated sensitive periods in the formation of socio-emotional skills in general and economic preferences in particular (J-PAL, 2013; Kautz et al., 2014). This paper aims at addressing that gap.

The empirical identification of sensitive periods is challenging for several reasons. First, in observational data, investments are often endogenous such that returns to investments cannot be interpreted in a causal manner. Second, the identification of sensitive periods requires comparing the returns to the *same investment into skills for children of different ages*. As an important prerequisite for clean inference, the investment needs to be implemented during the same time period for children of different ages to ensure that possibly interfering time trends do not overlay age-specific treatment effects. Moreover, investment intensity needs to be held constant across the different age groups. In contrast, studying an investment that is introduced at the same stage for all children and observing child outcomes in different grades in cross-sectional data compounds age heterogeneity in investment returns with differences in treatment length. This exacerbates the difficulties in learning about sensitive periods.[3]

In this paper, we propose and implement a novel and clean design to identify sensitive periods that can be applied more broadly in future work. To enable causal inference, we set up an RCT in which we assign a given investment to treatment children of different ages, holding treatment period and intensity constant across all age groups. We then measure heterogeneity in the treatment effect along the age distribution (proxied by school grade). Grade-specific treatment

[1] Skills produced at one stage do not only persist but also augment the skills attained at later stages. This so-called self-productivity embodies the idea that skills are reinforcing and cross-fertilizing, i.e., a higher stock of a given skill in one period raises the stock of the same or another skill in the next period. A second key feature of skill formation is complementarity: skills produced at one stage raise the productivity of investment in skills at subsequent stages. Together, complementarity and self-productivity produce multiplier effects such that skills together with investments are predicted to beget skills (Cunha and Heckman, 2007, 2008; Cunha et al., 2010).

[2] Moreover, there may exist critical periods in the formation of skills. If an individual does not receive a stimulus during a critical period it may be difficult or even impossible to develop a skill later in life.

[3] For example, providing free lunch at school from first grade onwards and comparing children in different grades or introducing such a program in different grades and observing outcomes at some point in time after school doesn't allow to disentangle age effects from the number of years children had access to the lunch (see, e.g., Hoynes et al., 2016, and Lundborg et al., 2021, as examples for quasi-experimental studies on the effects of food stamp or school lunch programs on economic, educational and health outcomes).

effects that are substantially larger than those for the same skill in other grades indicate the existence of a sensitive period.

The investment under consideration is a well-established social and emotional learning (SEL) program, the Lions Quest Skills for Growing program, which provides the same investment to children of different ages. The program aims at supporting young people confronted with the challenges of growing up: they learn how to manage their emotions, achieve their goals, care about and have wholesome relationships with others, and act responsibly. It has a long-standing history and follows a curriculum that is implemented by the children's teachers in the classroom environment. The curriculum comprises lessons on personal development, social development, as well as responsible decision-making with respect to health and prevention. Based on its detailed documentation such as its "Universal Program Guide" (LCIF, 2016) and the grade-specific "Facilitator's Resource Guides" for teachers, we hypothesize that program participation may increase three important socio-emotional skills: children's self-control, patience, and prosociality.

Self-control and patience are both integral to people's intertemporal decision-making as modeled by time preferences. Self-control captures an individual's present-bias that influences the extent to which individuals are able to resist temptations and suppress immediate impulses in order to achieve their long-term goals. Higher self-control is associated with higher educational attainment, better health, greater labor market success, more financial well-being, and greater overall life satisfaction (Tangney et al., 2004; Moffitt et al., 2011; Cobb-Clark et al., 2019). Patience reflects the long-run discount factor in inter-temporal utility and has been shown to predict education, labor market and health outcomes, and savings (DellaVigna and Paserman, 2005; Sutter et al., 2013; Golsteyn et al., 2014; Alan and Ertac, 2015), for example. Capturing altruistic behaviors, prosociality has been linked to both individual-level outcomes such as labor market success (Deming, 2017) and societal outcomes such as the provision of public goods and management of commons (Ostrom et al., 2002). Each of our skill measures combines children's revealed preferences measured in incentivized experiments and validated survey batteries administered to children or their mothers. This synthesis of survey and lab-in-the-field assessment of skills results in measures that reflect the multi-dimensional nature of the underlying skills and characterize individuals comprehensively (Falk et al., 2018; Kosse et al., 2020). Our approach also reduces measurement error and potential demand effects (Hertwig and Ortmann, 2001).

Our RCT comprises about 3,200 children. In particular, we compare self-control, patience, and prosociality of children in grades 2 to 5 (aged 7 to 11) in 69 treatment schools in rural Bangladesh, who participated in the Lions Quest Skills for Growing program for 28 weeks, to those of children in 66 control schools. By comparing grade-specific treatment effects of the same program, we are able to identify possible sensitive periods in the formation of self-control, patience, and prosociality between the ages 7 to 11.

Our main findings can be summarized as follows. Overall, participation in the Skills for Growing program enhances self-control and prosociality in elementary school children significantly. Averaging treatment effects across grades yields increases of 10.7 and 8.8 percent of a standard deviation, respectively. While positive as well, the overall treatment effect on patience (4.4

of a standard deviation) is not significantly different from zero. Comparing treatment effects across grades reveals substantial heterogeneity that points to sensitive periods in the formation of the analyzed skills. In particular, for self-control and patience as key dimensions of time preferences, treatment effects are substantially larger for younger children in grade 2—and in the case of self-control also grade 3—than for older children. While grade-specific treatment effects are close to zero and not significant for older children, they range between 15 and 21 percent of a standard deviation for second graders ($p < 0.05$) for both self-control and patience. Grade-specific treatment effects for prosociality are relatively large and significant throughout grades 2 to 5 (13 to 14 percent of a standard deviation, all $p < 0.05$ or < 0.1), with the exception of grade 3, suggesting that grades 2, 4, and 5—and, if the grade 3-specific treatment effect was an outlier, the whole age range between 7 and 11 years—can be considered as a sensitive period for the formation of prosociality.

The contribution of this study is threefold. First, as a conceptual contribution, we propose a novel design for identifying sensitive periods in the formation of children's skills (Cunha and Heckman, 2007) and provide a first proof of concept. While the results of this paper are only a first step towards addressing the lack of knowledge regarding the timing of sensitive periods in the formation of children's socio-emotional skills, our proposed design can also be applied in future research endeavors. Related research in developmental psychology and neuroscience defines sensitive periods as limited periods during which effects of experience on the brain are unusually strong (Knudsen, 2004; Zeanah et al., 2011; Hartley and Frankenhuis, 2020), which closely resembles the definition applied in this paper. Studies focus on the development of perceptual, cognitive, and emotional capacities such as vision, language proficiency, the formation of social relationships and stress management, as well as on underlying neural circuits (see Knudsen, 2004, and his references for cited examples). Identification of sensitive periods in these disciplines has proven difficult as well and mainly follows environmental deprivation paradigms originating in animal research. Human developmental research therefore studies contexts where deprivation naturally occurs such as for children raised in institutions who lack sensitive and responsive caregivers (Gabard-Durnam and McLaughlin, 2020). Enriching environments via childhood interventions is a new approach bypassing the obvious ethical problems that would arise in controlled deprivation studies.

Second, our findings have important policy implications. For example, our results on sensitive periods in the formation of self-control and patience align with "the earlier, the better" findings regarding the development of cognitive skills (see, e.g., Knudsen et al., 2006; Zeanah et al., 2011; Heckman and Mosso, 2014), extending evidence that earlier investments often have larger returns than later ones in the domain of time preferences. More generally, our findings underline that the same investment may be more effective at some ages than others. Identifying sensitive periods is thus crucial for an effective and efficient timing of parental or public investments, including interventions that aim at enhancing socio-emotional skills. Finally, our results show that even if returns to investments in the cognitive skills of disadvantaged children beyond age 3 are low, returns to investments in socio-emotional skills can still be comparably high, as hypothesized by Cunha and Heckman (2007) and Borghans et al. (2008).

Third, we evaluate the Lions Quest Skills for Growing program that formed the basis of the RCT with respect to its impact on self-control, patience, and prosociality. Although the effect sizes we document are slightly smaller than those occasionally found for intensive model programs (Heckman et al., 2010) that are especially designed to foster specific skills and target disadvantaged children only, our results indicate that available, large-scale programs can provide an effective tool for improving children's socio-emotional skills. We thereby add to the literature on interventions for elementary school children, in which large-scale evaluations that are based on RCTs such as ours are rare (which is documented in, e.g., Rodríguez-Planas, 2012, and Kautz et al., 2014); for studies in the school context focusing on self-control and patience, see Alan and Ertac (2018) and Sorrenti et al. (2020) who also provide a comprehensive review of recent intervention studies on socio-emotional skills; for results on prosociality, see John and Thomsen (2015), Alan and Ertac (2017), Rao (2019), Cappelen et al. (2020) and Kosse et al. (2020). In light of the frequent implementation of the Lions Quest programs (cf. section 2.1), rigorous evaluations, especially of the Skills for Growing program targeting elementary school children, are surprisingly scarce and suffer from methodological limitations and small sample sizes. Only two studies have evaluated the Lions Quest Skills for Growing program using quasi-experimental designs. Kidron et al. (2015) report on a two-year implementation in grades 3 to 5 in eight US elementary schools and find positive effects on students' self-reported interpersonal skills and perception of their school environment as safe and supportive, as well as on disruptive behavior at school. However, the authors document low implementation quality and in the end the program was even delivered by Lions Quest guidance counselors instead of teachers. Gol-Guven (2017) collected data in four schools in Turkey (two program and two control schools) and finds positive effects on school climate, students' behaviors, and conflict resolution skills, but not on students' perceptions of school.[4]

The remainder of this paper is organized as follows. In the next section, we discuss the design of our study, the intervention, its implementation, and hypotheses. Section 3 first provides details on sampling, data collection, and the randomization procedure. We then describe experimental and survey measures of self-control, patience, and prosociality and how we construct the respective outcome indices. Section 4 presents our results and robustness checks, before we conclude in section 5.

2 Study design and hypotheses

Our study design builds on the Lions Quest Skills for Growing (LQ SfG) program for two main reasons: First, as discussed in detail in section 2.3, we hypothesize that it may affect the formation of three important socio-emotional skills—self-control, patience, and prosociality—that are powerful predictors of individual decision-making and many life outcomes (Almlund

[4]More studies deal with the subsequent programs for adolescents but exhibit similar shortcomings in their evaluation-setups or investigate, for instance, the effects on teachers instead of student outcomes (see, e.g., Matischek-Jauk et al., 2018, and Maalouf et al., 2019, and the references therein). Meta-analyses on the impact of different universal social and emotional learning programs attest participants in general improved attitudes, behavior, academic performance, and indicators of well-being (Durlak et al., 2011; Taylor et al., 2017). However, by far not all reviewed studies used randomized designs and even though all programs underlying the reviewed studies targeted the development of social and emotional skills, most of them did not assess skills as an outcome (Durlak et al., 2011). Also, neither of these meta-studies discusses sensitive periods.

et al., 2011; Moffitt et al., 2011; Falk et al., 2018; Cobb-Clark et al., 2019). Second, the program provides well-established means to target the same skills with the same investments at different ages during childhood. It thus enables a design that allows investigating existence and timing of sensitive periods in the formation of these skills, see section 2.2. In section 2.1, we start by providing basic information on the intervention.

2.1 The intervention: The Lions Quest Skills for Growing program

Developed by the Lions Clubs International Foundation, a global non-profit organization, Lions Quest (LQ) programs have a longstanding history. Together with its sister programs, the LQ Skills for Adolescence for middle schoolers and the LQ Skills for Action for high schoolers, the LQ Skills for Growing (SfG) program for primary school children has been realized in more than 100 countries worldwide (Maalouf et al., 2019). LQ programs are classroom-based social and emotional learning (SEL)[5] programs that aim at helping young people to find their way by learning how to manage emotions, achieve their goals, have supportive relationships with others, and act in a responsible and caring manner.[6]

The SfG program comprises six units that each consist of several lessons: 1) building a positive learning community, 2) personal development, 3) social development, 4) health and prevention, 5) leadership and service, and 6) a reflection section on what has been learned. According to the program's "Scope and Sequence Sheet" (see appendix Figure A1), the program aims at promoting the "SEL competencies" self-awareness and self-management, social awareness and relationship skills, as well as responsible decision-making. The documentation of the program translates these SEL competencies, in turn, into the academic concepts of self-discipline, impulse control, goal-setting, working cooperatively, empathy, and self-confidence as underlying skills to be fostered by the program units (again, see Figure A1).

Throughout the implementation of the program, we collaborated with the Lions Clubs International Foundation to stick to standard procedures as closely as possible. Moreover, the Ministry of Primary and Mass Education of Bangladesh supported the program's rollout, directing treatment schools to teach the LQ SfG program. The program was implemented by the children's primary school teachers following a pre-existing curriculum. In cooperation with the Lions Clubs International Foundation, learning materials got translated and adopted to the local context (for instance, pictures of children in the materials were changed to depict Bangladeshi rather than U.S. American or Indian children). In the course of our program adoption, school teachers got trained as LQ teachers in intense three-day workshops by international, qualified LQ trainers and received a textbook with detailed instructions. Children got a student journal in which they could summarize the topics and do homework. Parents were invited to a single, mid-program meeting with the LQ teachers, implementation staff, and local education authorities in their children's school. The program ran over the course of one school year following a

[5]Social and emotional learning is a form of positive youth development (PYD, see `https://youth.gov/`). The term "social and emotional learning" was introduced by the Collaborative for Academic, Social and Emotional Learning (CASEL), a Chicago-based consortium of educators and educational scholars, see `https://casel.org/what-is-sel/`. Other major SEL programs include Promoting Alternative Thinking Strategies (PATHS), Life Skills Training (LST), and the Seattle Social Development Project (SSDP); see `https://pg.casel.org/review-programs/` for available SEL programs (all webpages were last accessed on August 11, 2021).

[6]For web content on the LQ SfG program, see `https://lions-quest.org` (last accessed on August 11, 2021).

28-week schedule between February and October 2019.[7] Lessons took place on a weekly basis and were held during classes that teachers could use flexibly (these classes are, e.g., often used for physical education). Total instruction time is thus constant across treatment and control schools.

Each lesson lasts for about 30 to 40 minutes and is divided into four parts. First, the teacher presents an every-day situation, like a short story or pictures of someone getting bullied, and identifies together with the students why this situation is problematic ("discovery phase"). Second, students are encouraged to share similar experiences and the class discusses reasons and solutions for the problem ("connecting phase"). When learning how to make good decisions, for example, children are taught to act according to the "Think, Predict, Choose Model." Being confronted with an emotional decision such as choosing which friend to invite to a festival, they are trained to follow a traffic light approach and step back, calm down, reflect on their options and the consequences and carefully work out what to do. They also discuss how to keep up their motivation for tedious tasks by not following immediate impulses but reminding themselves of why this goal is important to them or where they have been successful before. Third, students reenact the presented situations in role plays or solve tasks in pairs or small groups and employ the solutions and strategies they have talked about ("practicing phase"). Finally, teachers assign a homework related to the week's topic ("applying phase"). Students are, for instance, asked to discover and solve similar situations in their daily life and document their progress in the student journal. Appendix Figure A2 shows examples for English versions of the instruction materials, teachers' resources guides, and student journals.

2.2 Identification of sensitive periods

The key objective of our study design is to provide new insights into the timing of sensitive periods in the formation of socio-emotional skills. Following Cunha and Heckman (2007) and their seminal work on the development of skills, childhood has more than one stage during which skills develop. Stages differ in how easily various skills are acquired and how productive investments in given skills are, for example, due to changing brain plasticity as children grow. Stages that are more effective in producing a certain skill and in which returns to investments into skills are particularly high are called sensitive periods for the acquisition of this skill. While this theoretical framework is well-established, empirical evidence on sensitive periods in the formation of socio-emotional skills is lacking (Cunha et al., 2006; J-PAL, 2013; Heckman and Mosso, 2014; Kautz et al., 2014).

The SfG program provides a particularly well-suited tool for investigating sensitive periods in the formation of socio-emotional skills, as it is designed to deliver the same investment (same content and objectives) in all grades. Appendix Figure A1 displays the program's official Scope and Sequence Sheet that summarizes the SfG curriculum and the targeted SEL components from pre-kindergarten to grade 8. It underlines that program lessons in each unit have the same aims and target the same skills from pre-kindergarten up to grade 8. Only the specific SfG materials

[7] In order to fit the SfG program into one school year with one lesson per week, the leadership and service unit was not implemented, as Lions Club recommends for a shorter, 28-week schedule. Given the content of this unit, we do not suspect that including it would do much in terms of fostering prosociality, patience, and self-control. In case it does, the estimated effects of the 28-week schedule are lower bounds of the effects of the full program.

(e.g., example stories used) are partly adjusted to better reflect students' cognitive development and everyday environment in the respective school grades. This feature allows us to introduce the same investment in different grades and hence to identify possible sensitive periods in the formation of children's skills. We measure sensitive periods as heterogeneity in the treatment effect of the intervention along the school grade. Skill- and grade-specific treatment effects that substantially exceed those for the same skill in other grades point towards the existence of a sensitive period.

In order to allow for clean inference, our design has several further important features. First, the SfG intervention was implemented during the same time period for children of different ages, excluding possibly different period effects. Second, treatment intensity is the same for all children in the treatment group, i.e., all treated children were equally long exposed to the intervention.

2.3 Hypotheses

Based on program structure, content, and aims, we derive the following hypotheses:

HYPOTHESIS 1A. Participation in SfG increases self-control.

HYPOTHESIS 2A. Participation in SfG increases patience.

Self-control and patience are both integral to people's intertemporal decision-making. In psychological research, self-control is often conceptualized as impulse control, while a common way to formalize it in economic theory is time-inconsistency. For example, $\beta\delta$-preferences or quasi-hyperbolic discounting (Laibson, 1997; O'Donoghue and Rabin, 1999), one of the most commonly used models of inter-temporal choice in behavioral economics (Ericson and Laibson, 2019), assume the following utility function: $U^t(u_t, u_{t+1}, ..., u_T) = u_t + \beta \sum_{\tau=t+1}^{T} \delta^\tau u_\tau$, where $0 < \beta \leq 1$ and $0 < \delta \leq 1$. Total utility is given by U^t, u_t is flow utility in period t, and parameter δ represents long-run, time-consistent discounting ("patience"). β is a present-bias parameter that indicates whether and how much an individual favors the current period over later periods. The smaller is β, the stronger is the degree of present-bias, which can be thought of as larger self-control problems. If $\beta < 1$, $\beta\delta$-preferences represent time-inconsistent preferences; that is, individuals place more relative weight on the current period once it has arrived than in any previous period and are thus more likely to give in to temptations and impulses in the here and now. If $\beta = 1$, $\beta\delta$-preferences coincide with time-consistent, exponential discounting. Time-consistent individuals ($\beta = 1$) are often referred to as individuals who do not have any self-control problems and time-inconsistent individuals ($\beta < 1$) as having self-control problems.

The SfG unit on personal development as well as lessons on how to best restrain yourself and keep calm when working in groups, on managing stress and strong emotions, on recognizing the connection between thoughts, emotions, and actions, or on how to set long-term goals, motivate yourself, and build healthy habits are expected to foster impulse control, making responsible and forward-thinking decisions, and self-management. This is likely to be reflected in a higher

level of self-control. Although not directly mentioned in the SfG's Scope and Sequence Sheet, the lessons that relate to intertemporal decision-making may also affect patience, and be it indirectly. Children learn, for instance, to think things through and consider all choices before making a decision.[8]

HYPOTHESIS 3A. Participation in SfG increases prosociality.

Prosociality comprises altruistic or prosocial behavior in interpersonal situations which comes down to behaviors that benefit others. In economics, models of social preferences (e.g. Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000) are consistent with such behaviors. The SfG units on building a positive learning environment and on social development, as well as lessons on how to show empathy and appreciation for others, on working together, and on social engagement are expected to foster social awareness and relationship skills. This is likely to be reflected in a higher level of prosociality.

While there is very little empirical evidence on sensitive periods in the formation of socio-emotional skills, research on the formation of cognitive skills suggests that earlier investments are more effective in enhancing skills than the same investments at a later stage. Based on these empirical findings on the formation of cognitive skills (Knudsen et al., 2006; Heckman and Mosso, 2014) and the theoretical framework of skill formation (Cunha and Heckman, 2007, 2008; Cunha et al., 2010), we formulate the following hypotheses:

HYPOTHESIS Participation in the SfG program in earlier grades is more beneficial than in later grades for the formation of

1B. self-control,
2B. patience,
3B. prosociality.

3 Data

In this section, we first provide details on sampling, data collection, and the randomization procedure. We then describe experimental and survey measures of self-control, patience, and prosociality and how we construct the outcome indices.

3.1 Sampling procedure and data collection

Data were collected in the four rural districts Netrokona, Sunamganj, Chandpur and Gopalganj of Bangladesh.[9] These districts represent four of the eight administrative divisions of the country. In the course of a previous survey that was conducted in 2014 and 2016 (see Chowdhury

[8] We also expect the intervention to increase self-esteem. However, self-esteem is usually assessed by children and adolescents themselves (as opposed to their parents or teachers, for instance) through survey scales. As confirmed by pre-tests, most of the children in our intervention are too young to answer self-esteem survey items meaningfully. For example, the commonly used Rosenberg scale is generally applied from age 10 onwards. This prevents us from considering self-esteem as an outcome variable.

[9] Even though, with around 163 million inhabitants, Bangladesh is the world's eighth most populous country, about two thirds of the population are living in rural areas. Bangladesh's living conditions are evolving rapidly. Bangladesh's GDP at purchasing power parity per capita increased from US-$ 851 in 1990 to US-$ 5,083 in 2021, which places Bangladesh at rank 148 out of 199 countries (The World Bank, 2021).

et al., 2022), 11 subdistricts were chosen based on the availability of NGOs willing to collaborate. 150 villages were randomly drawn from the 11 subdistricts. In 2018, the 150 villages were visited and a public primary school suitable for sampling school children was chosen. Most villages had only one primary school. If a village had more than one school, the school with the majority of students from the village and situated at the village center was selected. As a one-to-one village–school matching was not always possible due to some schools serving multiple villages, the process resulted in a selection of 135 primary schools. Primary school in Bangladesh is compulsory and covers grades 1 to 5, starting at age 6.[10] Public primary education is free of charge.

The Lions Quest Skills for Growing program was taught at school level in 68 of the 135 selected public primary schools. To measure its effects, we sampled children and their families from these 135 primary schools. In general, from each school and each grade 2 to 5, five students were sampled randomly from class lists. If a school was serving two or three of the original 150 sample villages, two or three times as many students were chosen, respectively. This results in about 2,800 families of sampled students. Sampling started in grade 2 instead of grade 1 to give children time to accustom themselves to the new school environment before being exposed to the intervention and ended with grade 5 as the highest grade in primary school. Children in grades 2 to 5 are in general between the ages of 7 and 11 Some children are slightly older if they had to repeat classes. Interviews with the sampled student, both parents (if available), and one randomly selected sibling (again, if available) were conducted by a specialized survey firm (ECONS) at the families' homes. If siblings were still in primary school, they usually visited the same school and therefore have the same treatment status as their sampled siblings (when in grade 2 to 5). Due to delays in the start of program implementation, the intervention did not start in 2018 as planned, but only in 2019 and the originally sampled students were one grade higher than initially expected. That is, students sampled in grade 2 were in grade 3 at the time of the intervention and students sampled in grade 5 had left primary school. To maintain the originally intended sample composition, the wave of data collection comprised a refreshment sample of new second graders for whom we do not have (pre-treatment) baseline information on skills. Our final sample consists of 3,222 children, see Table 1.

As a result of this recruitment process, we have a large sample of families in which we measure both children's and parents' skills comprehensively.[11] The first part of the interviews (the "household survey") assessed survey information on socio-demographics, income, expenditures, employment, land ownership, credits and savings, assets, and health. It was answered by either the household head or his/her spouse (whoever was the most knowledgeable person for the respective part) using computer-assisted personal interviews (CAPI). The second part of the interviews (the "skill assessment") elicited economic preferences (time, risk, and social

[10]Nearly 60 percent of primary schools are managed and financed by the Ministry of Primary and Mass Education. The remaining primary schools are under the responsibility of other ministries and non-governmental organizations, also comprising (private and state-sponsored) religious schools (The World Bank, 2018).

[11]Since this is a multipurpose-built dataset (see German Research Foundation (DFG) project no. SCHI 1377/1: "Towards a better understanding of the development of non-cognitive skills in children: Malleability, sensitive periods, typical trajectories, and transmission within the family"), it also includes information on skills we do not expect to be affected by the intervention. Following the pre-registration, we do not analyze treatment effects on these skills.

TABLE 1: Sample overview

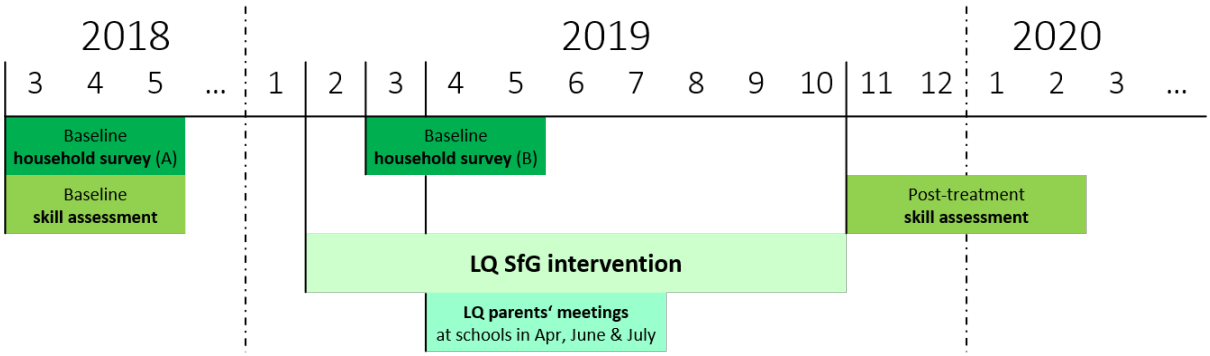
	(1) All	(2) Treatment group	(3) Control group
Number of schools	135	69	66
Number of students	3,222	1,637	1,585
– sampled students	2,809	1,434	1,375
– siblings	413	203	210
By grade			
– grade 2	890	455	435
– grade 3	789	386	403
– grade 4	763	393	370
– grade 5	780	403	377

NOTES: Own representation.

preferences), personality traits, and cognitive skills via paper-and-pencil interviewing (PAPI) for both the sampled children, up to one sibling, and their parents. We complement this comprehensive information on skills for the whole family with a questionnaire that mothers answered about their children. In the mother questionnaire, mothers assessed among other things their children's strengths and difficulties (including prosociality) as well as their self-control (for children up to age 13).

Up to now, we have collected two waves of data, see Figure 1: a baseline wave of interviews before the treatment in 2018 and early 2019 and a post-treatment wave of skill assessment interviews after the end of the program in 2019 and early 2020.

FIGURE 1: Timeline of data collection and intervention



NOTES: Own representation.

In our sample, children's mean age is 9.4 years and 51.4 percent of children are girls. On average, yearly household income is around 230,000 Taka (approximately 2,700 US-$). More than 90 percent of households have an electricity connection. Fathers' mean age is 43, mothers' mean age is 35. 58 percent of fathers and 72 percent of mothers can read and write. Almost all fathers and 88 percent of mothers are working. The latter, however, are usually looking after the family's live stocks or poultry instead of being formally employed. Table A4 in the appendix provides a more detailed sample description.

3.2 Stratified randomization

The treatment was randomized on school level following a stratified randomization procedure. The 135 participating schools were assigned to 35 strata, based on the 11 subdistricts, the villages' literacy rates, and their distance to the subdistrict capitals.[12] Distance to subdistrict capital is expected to possibly reflect lower school quality with more rural schools being less attractive for teachers and less effectively supervised by education officials. The village literacy rate is a proxy for the village population's educational level. Within each stratum, each school was randomly assigned to either treatment or control group using a random number generated in Stata.

69 schools were randomly drawn for the treatment group, the remaining 66 schools serve as our control group.[13] As the treatment was part of the curriculum in grades 2 to 5 of the treatment schools, children could not single-handedly drop out of the treatment. However, the teachers of one treatment school with 22 sampled students did not receive the LQ teacher training due to miscommunication and the school did therefore not implement the treatment. Given that 1,637 students were supposed to receive the treatment and only 22 of them did not, the compliance rate is 98.7 percent ($=1-{}^{22}/_{1,637}$). Since we will present intention-to-treat estimates throughout the paper, the students in this school still belong to the treatment group in all analyses.

3.3 Outcome variables

Our data include multiple measures to capture self-control, patience, and prosociality. In particular, our skill measurements include both children's revealed preferences from incentivized experiments and well-established survey scales answered by mothers about their children (or, in the case of patience, by children themselves). This combination of lab-in-the-field and survey assessment results in measures that reflect the multi-dimensional nature of the underlying skills and combines the advantages of incentive-compatible experiments and validated psychological survey questionnaires (Falk et al., 2018; Kosse et al., 2020). The remainder of this subsection describes the experiments and survey scales as well as their aggregation into final outcomes.

Experiments: Time and social preferences

Children participated in a sequence of experiments designed to measure two core dimensions of economic preferences: time and social preferences. To elicit preferences, we relied on well-established measurement tools that, in the case of time preferences, have been used in developing countries before. We still carefully pre-tested all items in our context and adapted payments to children's ages. We used standardized control questions to verify that participating children understood the instructions.[14]

[12] The strata were constructed by first splitting schools according to the 11 subdistricts they belonged to, then by a binary indicator whether the village is below or above the median distance to the subdistrict capital and lastly by a binary indicator whether the village is below or above the median literacy rate as measured by a preceding village survey in 2016. In some districts, all villages were either above or below the median literacy rate. Therefore, the number of strata is not 44 (which would equal $11\times2\times2$), but 35.

[13] Since not all strata contained an even number of schools, randomization within strata led to an unequal number of treatment and control group schools.

[14] Understanding is controlled by interviewers asking children in between to repeat their explanations (four times for the time preferences game and once for the social preferences game). Each time, the interviewer notes

The order of the experiments was randomly determined by rolling a die. Children were able to earn stars which were transformed into money after the experiments using age-specific exchange rates (proportional to pocket money: one star's value ranged between 10 and 30 Taka, depending on children's age, which equals approximately half of a child's weekly pocket money). Each child received one star as a show-up fee. All experiments took place in one-on-one settings in the families' homes and the interviewers ensured that members from the same household could not influence each other's decisions.

Time preferences: time-consistency and patience. In order to measure time preferences, we followed a simple choice list approach, used by, e.g., Bauer et al. (2012) in a similar form for adults in rural India. Each child had to make six choices which consisted of trade-offs between smaller, sooner and larger, later rewards (see Table 2). The six choices were grouped in three choice sets, each consisting of two choices with the same time delay. The early payment took place either on the next day (choice sets 1 and 2) or in a month (choice set 3), the later payment in three weeks (choice set 1), three months (choice set 2), or four months (choice set 3), respectively. The choice sets were ordered randomly.

Table 2: Time preferences experiment for children

Choice Set 1	2 stars tomorrow	vs.	3 stars in 3 weeks
	2 stars tomorrow	vs.	4 stars in 3 weeks
Choice Set 2	2 stars tomorrow	vs.	3 stars in 3 months
	2 stars tomorrow	vs.	4 stars in 3 months
Choice Set 3	2 stars in 1 month	vs.	3 stars in 4 months
	2 stars in 1 month	vs.	4 stars in 4 months

Notes: Own representation.

As an experimental measure of self-control, we observe children's *time-consistency* in choice sets which have the same three-months time delay but different starting points: starting either on the next day or in one month. Children are classified as time-consistent if they make identical choices in choice sets 2 and 3, implying that their current and future discount rates are equal. Additionally, to disentangle time-consistency from extreme impatience, we only classify children as time-consistent if they exhibit some degree of patience by choosing (and hence, waiting for) the larger, but later reward in at least one of the choice sets 2 and 3.

As an experimental measure of *patience* we simply count the number of patient choices in all six decisions, i.e., we count the number of larger, but later reward choices among all six decisions. This measure hence ranges from 0 to 6.

Social preferences: altruism. We followed an experimental protocol inspired by Fehr et al. (2008) which got extended by Bauer et al. (2014) to assess social preferences using dictator games. Children made four allocation choices dividing stars between themselves (x) and another child (y) of the same gender and roughly the same age, but unknown and unrelated to them

down whether the child understood the game after the first, second or third explanation or whether the child did not understand the game at this point. A child is indicated as having understood a game if answering each of the control questions correctly after at most three explanations of the interviewer.

(see Table 3). In each of the four choices (x,y), one option was the allocation (1,1), while the alternative allocation benefited one of the children (y>x in two cases and y<x in two cases).

TABLE 3: Social preferences experiments for children

Costly prosocial game	1 star for me 1 star for the other child (1,1)	**vs.**	2 stars for me 0 stars for the other child (2,0)
Costless prosocial game	1 star for me 1 star for the other child (1,1)	**vs.**	1 star for me 0 stars for the other child (1,0)
Costless envy game	1 star for me 1 star for the other child (1,1)	**vs.**	1 star for me 2 stars for the other child (1,2)
Costly envy game	1 star for me 1 star for the other child (1,1)	**vs.**	2 stars for me 3 stars for the other child (2,3)

NOTES: Own representation.

As our experimental measure of *altruism*, we count the number of stars a child allocated to herself and to the other child and calculate the overall share of stars a child has given to the other child across all four games. This share varies between 0.29 and 0.58.

In our preferred specifications, we set an experimental outcome to missing if a child answers the respective control question(s) incorrectly. In the time preference experiment, we also set answers to missing if children prefer the later, larger reward over the sooner, smaller one only for the lower interest rate (decisions 1, 3, and 5), but not the higher one (corresponding decisions 2, 4, and 6, respectively), as this likely reveals that children have not understood the experiment.[15] In robustness checks we also consider the answers of children excluded in our preferred specification.

Survey measures

SELF-CONTROL: REVERSED IMPULSIVITY SCALE FOR CHILDREN. We measure self-control by using the reversely coded Impulsivity Scale for Children (ISC) that was introduced and validated by Tsukayama et al. (2013). The authors conceptualize effortful and volitional control that is exerted in order to achieve long-term goals. Their scale is designed to capture both impulsive behavior in the social context and with respect to schoolwork. The ISC consists of eight items which record how often specific behaviors occur that are rated on a five-point Likert scale by the mother regarding her children aged 6 to 13. Appendix Table A1 lists the eight items. For example, mothers had to state whether their child interrupts other people "at least once per day," "about once per week," "about 2 to 3 times per month," "about once per month," or "almost never." Answers are combined with equal weighting into one scale.

PATIENCE. Children were asked to rate how well the statement "I am good at giving up something nice today (e.g., a reward) in order to get something even nicer in the future (e.g.,

[15] This affects 116 children, out of 3,222 children in total and 3,140 children with otherwise complete information.

a larger reward)." applies to them on a five-point Likert scale from 1 ("not at all right") to 5 ("absolutely right").

Prosocial behavior. We make use of the prosociality scale of the well-established and widely used Strengths and Difficulties Questionnaire (Goodman, 1997; Goodman et al., 2000) to measure the extent to which children behave prosocially, i.e., interact with others in a positive and cooperative way in their daily routine. Mothers rated five items related to their children's prosocial behavior on a three-point scale such as "Considerate of other people's feelings" or "Shares readily with other children (treats, toys, pencils, etc.)," see appendix Table A2 for all five items. Answers are combined with equal weighting into one scale.

Aggregation of measures

To obtain a comprehensive assessment of the socio-emotional skills under study, we combine information from experiments and surveys into aggregate measures that reflect the multi-dimensional nature of the underlying skills. We refer to the aggregate measures as self-control (index), patience (index), and prosociality (index).

The combined outcomes are calculated as follows: We first standardize each experiment and survey component to have a mean of 0 and a standard deviation of 1 across control group observations. We then calculate an individual's mean over the standardized components that enter the final outcome index. This index is again standardized to have a mean of 0 and a standard deviation of 1 in control group terms (z-score). If one of the measures is missing, we use only the remaining component for the child.[16]

Self-control. The self-control index combines experimentally elicited time-consistency with the reversed Impulsivity Scale for Children (ISC).

Patience. The patience index combines experimentally elicited patience with a survey question on patience.

Prosociality. The prosociality index combines experimentally elicited altruism with survey-assessed prosocial behavior, measured by the respective SDQ scale.

Figure A3 displays the post-treatment distribution of our standardized outcome indices, while Figure A4 shows the distribution of each outcome component (including number of observations).[17] Appendix Table A3 provides additional results for the treatment effects presented in section 4 using alternative definitions of the experimental outcome components: changing the unit of measurement of time-consistency, using a binary patience measure instead of a continuous one, changing the way we aggregate the decisions in the social preferences experiment, or the sample composition by including children who likely did not fully understand parts of the experiments does not alter our interpretation of results.

[16] z-scores are calculated over all control group observations that enter a regression. Using age-specific z-scores does not change the magnitude of our findings.

[17] Skill components seem to be complements rather than substitutes. For self-control, components are only weakly correlated (Pearson correlation coefficient of -0.037 with p-value < 0.05). For patience, components are significantly correlated (correlation coefficient of 0.353 with p-value < 0.01). Experimentally elicited altruism and survey-assessed prosocial behavior are not significantly correlated.

3.4 Baseline balance and attrition

Baseline imbalance and selective attrition are potential threats to identification in randomized controlled trials. Since we have collected the outcomes of interest (self-control, patience, and prosociality) not only after but also before treatment assignment, we can use these data to provide evidence on successful randomization and the absence of selective attrition.

Baseline balance

Given our randomization procedure, there should be no systematic difference in the pre-treatment outcomes of treatment and control group. As a first balancing test, we regress the pre-treatment outcome measures (assessed in 2018) on the treatment indicator. As expected, Table 4 shows that pre-treatment differences in means and distributions between treatment and control group are small and statistically not significant.

TABLE 4: Balancing results for pre-treatment outcomes

	(1)	(2)	(3)	(4)
	Observations	Treatment/ control group difference	p-value t-test of equal means	p-value K.-Smirnov test
Self-control index, pre-treatment	2,170	−0.034	0.613	0.308
Patience index, pre-treatment	2,162	0.010	0.858	1.000
Prosociality index, pre-treatment	2,186	−0.031	0.697	0.542

NOTES: All variables are standardized such that the mean of the control group in 2018 is 0 and the standard deviation is 1. Point estimates and p-values of t-tests are obtained from regressions of pre-treatment outcomes on the treatment indicator. School-clustered standard errors.

As a second randomization check for covariates, we regress 27 pre-treatment, socio-demographic child and family characteristics on the treatment indicator.[18] Appendix Table A4 gives a detailed account of these variables, their definitions, and the balancing results. Under successful randomization, the actual number of significant differences between treatment and control group (displayed in the second row of Table 5) should be similar to the number of significant differences we expect to find by chance for a given significance level (see the first row of Table 5). This is indeed the case such that, in sum, both checks for baseline balance indicate that our treatment randomization was successful.

Absence of selective attrition

As we interview families twice within 18 months, some families may have opted out of the second interview. With 4.4 percent, the overall rate of attrition is low (4.6 percent in the control and 4.3 percent in the treatment group). Attrition is unlikely to be systematically related to treatment, as the treatment was on school level and there was no visible connection between the treatment in school and the interviews at home. To investigate this further, we regress an attrition indicator (1 if families are interviewed in 2018 but not in 2019, 0 else) on

[18] Unlike the pre-treatment outcomes that were measured in wave 1 in 2018, the covariates were measured in the household survey in wave 2 in 2019 at the beginning of the intervention period, see Figure 1, and are thus available for the second graders as well.

TABLE 5: Overview on covariates' baseline balance

	(1)	(2)	(3)	(4)	(5)
		Significance level			
SET OF VARIABLES	Number of var.	1%	5%	10%	15%
Expected number of significant effects	27	0.27	1.35	2.7	4.05
Actual number of significant effects		0	1	1	4

NOTES: The first row of the table displays the number of significant point estimates we expect by chance when regressing 27 child and family characteristics on the treatment indicator, given the significance level stated in the respective column. The second row gives the number of actually significant point estimates. Appendix Table A4 reports the 27 child and family characteristics, their definition, and the point estimates along with their p-values.

the treatment indicator and children's pre-treatment self-control, patience, and prosociality as well as their interaction. Appendix Table A5 shows the results. All coefficients are economically small, none is significant at the 5 percent level, and only one coefficient, the interaction between pre-treatment self-control and the treatment indicator is significant at the 10 percent level. An F-test for joint significance of treatment, pre-treatment outcomes, and their interactions yields $p = 0.28$. All in all, our results do not suggest that selective attrition is affecting our results.

4 Results

In this section, we first provide causal evidence on the overall treatment effects of the Lions Quest Skills for Growing program on self-control, patience, and prosociality. Section 4.2 then exploits the specific features of our design that enable the identification of sensitive periods in the formation of these skills and presents the corresponding results. Section 4.3 contains robustness checks. This includes a confirmation of our main results using p-values adjusted for multiple hypotheses testing or randomization inference as well as estimations of treatment effects when controlling for pre-treatment outcomes and additional family characteristics.

4.1 Treatment effects on socio-emotional skills

Table 6 displays the treatment effects of the Lions Quest Skills for Growing program on the indices of socio-emotional skills.[19] The indices are standardized, with a control group mean of 0 and a standard deviation of 1. In our preferred specification, we regress the skill indices Y on a treatment indicator (=1 if treated, 0 otherwise) and a full set of strata fixed effects $\boldsymbol{\phi}$:

$$Y = \alpha + \beta \text{treatment} + \boldsymbol{\phi} + \varepsilon \tag{1}$$

Standard errors are clustered at school level. We find that program participation increases children's self-control by 10.7 ($p < 0.01$) and prosociality by 8.8 ($p < 0.05$) percent of a standard deviation. These findings provide evidence in line with hypotheses 1A and 3A. Regarding

[19] Strictly speaking, we estimate intention-to-treat (ITT) effects, as only 98.7 percent children intended to participate in the treatment actually did so (see section 3.2). Given that non-compliance is very limited, we do not qualify estimated parameters as ITT every time, but just talk about treatment effects.

hypothesis 2A on patience, we find a positive, but smaller effect size of 4.4 percent of a standard deviation that is not statistically different from zero. As we show later, however, this aggregate result hides heterogeneity across grades. Also, given that the documentation of the SfG program does not explicitly name patience as a targeted skill, it is not surprising that the treatment effect on patience is smaller.

TABLE 6: Overall treatment effects

	(1)	(2)
DEPENDENT VARIABLE	Number of observations	Treatment effect
Self-control index	3,208	0.107***
		(0.039)
Patience index	3,166	0.044
		(0.042)
Prosociality index	3,219	0.088**
		(0.036)

NOTES: All dependent variables are standardized such that the mean of the control group is 0 and the standard deviation is 1. Regressions include a full set of strata fixed effects. School-clustered standard errors in parentheses. Significance: $^{+}p < 0.15$, $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Putting effect sizes of the significant treatment effects into perspective underlines their economic significance. One way to do so is to look at the control group and compare skill index values of subgroups such as girls and boys, children from households with different income levels, or children in different school grades. For example, regarding self-control, the size of the treatment effect resembles a reduction of the gender gap in self-control (girls in the control group have, on average, 20 percent of a standard deviation higher self-control than control group boys) by half. Moreover, the gap along median income is 14 percent of a standard deviation. So, our treatment effect size is about two-thirds of the gap between children from below-median income households and those above. The treatment-induced increase in prosociality exceeds the average increase in prosociality we observe when students advance one year in age (of about 7 percent of a standard deviation per grade). The treatment effect also accounts for 80 percent of the gender gap in prosociality (11 percent of a standard deviation, again with higher values for control group girls than boys). Please note that these comparisons do not imply that the treatment, for example, closes or reduces the gender gap. For respective statements one would need to look at heterogeneous treatment effects for girls and boys.

For completeness, Table 7 displays the effects of the treatment on each of the six components of the socio-emotional skill indices.[20] It conveys several insights. First, all coefficients are positive, underlining that the treatment tends to uniformly increase the various facets of self-control,

[20]The way we calculate the index measures (as the average of the non-missing components) does not require that the treatment effect on the index equals the unweighted average of the treatment effects on the components. We could achieve this by imputing missing values in the components with the mean value of the components over the treatment and control group before taking the average over the components, as suggested by Kling et al. (2007). Our interpretation of the treatment effects does not change when we apply this strategy. However, we prefer the index aggregation used here, as it does not rely on the additional assumption that missing values in the components are treatment neutral.

patience, and prosociality. Second, the treatment effect on self-control is largely driven by a decrease in children's impulsive behavior. The treatment effect on the reversed Impulsivity Scale amounts to 11 percent of a standard deviation ($p < 0.01$). Given that impulse control and self-discipline are among the SEL-components that are explicitly targeted by the SfG program (see, e.g., the SfG PreK–8 Scope and Sequence Sheet in Figure A1), this confirms our expectations. Third, the increase in prosociality is largely driven by more altruistic behavior as measured in the social preference experiment, with a treatment effect of 9.2 percent of a standard deviation ($p < 0.05$). Although the treatment effect on the prosocial behavior scale is positive, it is neither large nor statistically different from zero. A significant effect on the experimental preference component, but not the survey scale suggests the absence of demand effects[21] since the observed increase in altruistic behavior in the dictator game induces monetary costs to children.

TABLE 7: Treatment effect on socio-emotional skill components

	(1)	(2)
DEPENDENT VARIABLE	Number of observations	Treatment effect
Self-control		
Exp: Time-consistency when patient	3,044	0.054
		(0.043)
Svy: Impulsivity Scale for Children (rev.)	3,077	0.110***
		(0.042)
Patience		
Exp: Patience experiment	3,044	0.016
		(0.044)
Svy: Patience survey item	3,166	0.050
		(0.041)
Prosociality		
Exp: Altruism experiment	3,162	0.092**
		(0.039)
Svy: Prosocial behavior scale	3,109	0.018
		(0.042)

NOTES: All dependent variables are standardized such that the mean of the control group is 0 and the standard deviation is 1. Regressions include a full set of strata fixed effects. School-clustered standard errors in parentheses. Significance: $^{+}p < 0.15$, $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

4.2 Sensitive periods in the formation of socio-emotional skills

We continue by exploiting the specific design elements of our randomized controlled trial that enable the identification of potential sensitive periods in the development of socio-emotional skills—namely that we implemented the same investment (the Lions Quest SfG program) in different school grades (i.e., at different ages), while holding intensity and start and end date of the investment constant. Given these features, comparing the treatment effects across grades informs about possible sensitive periods in the formation of the analyzed skills in the age range we consider (ages 7 to 11). In order to estimate grade-specific treatment effects, we regress the

[21] We take great care to exclude demand effects by design. For example, the treatment takes place in school, the interviews at home and the intervention is not mentioned during the interviews and experiments.

skill indices Y on a full set of grade indicators (1(grade=g)=1 if grade=g with $g \in \{2,3,4,5\}$, 0 otherwise), the grade indicators interacted with the treatment indicator, and a full set of strata fixed effects $\boldsymbol{\phi}$ (the linear treatment indicator is multicollinear):

$$Y = \sum_{g=2}^{5} \Big(\gamma_g 1(\text{grade} = g) \times \text{treatment} + \omega_g 1(\text{grade} = g) \Big) + \boldsymbol{\phi} + \varepsilon \tag{2}$$

Table 8 reports the coefficients of the grade indicators (ω_g) and the grade–treatment interactions (γ_g); Figure 2 illustrates the latter, that is, the grade-specific treatment effects. Grade-specific treatment effects qualify as sensitive periods if they are significant and substantially larger than grade-specific treatment effects in other periods, implying a higher return to investment.

TABLE 8: Sensitive periods as captured by treatment effect heterogeneity along school grades

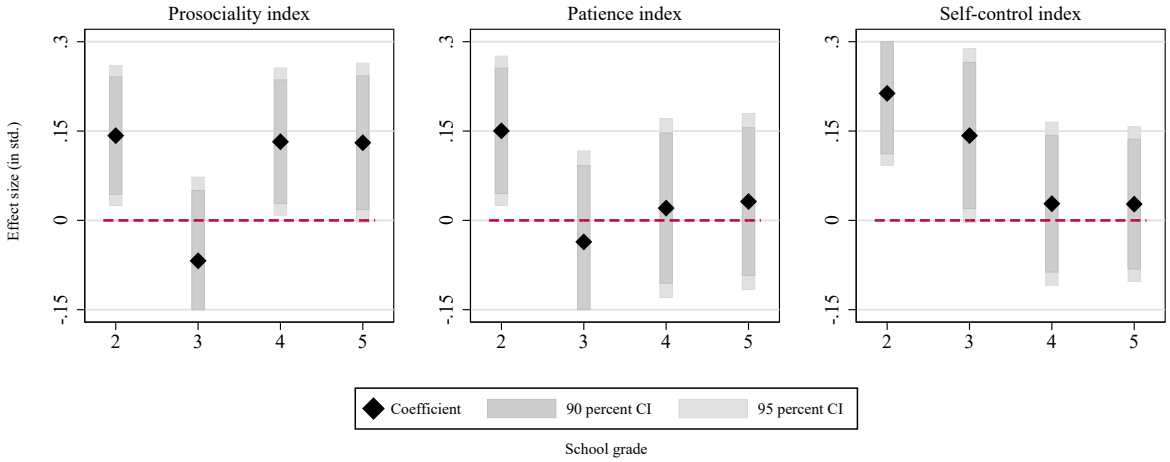
	(1)	(2)	(3)	(4)
	School grade			
DEPENDENT VARIABLE	grade 2	grade 3	grade 4	grade 5
Self-control				
Grade		−0.046	−0.077	0.038
		(0.052)	(0.068)	(0.063)
Treatment × grade	0.213***	0.142*	0.028	0.027
	(0.062)	(0.075)	(0.070)	(0.066)
Patience				
Grade		−0.017	−0.170***	−0.084
		(0.060)	(0.063)	(0.068)
Treatment × grade	0.150**	−0.036	0.021	0.032
	(0.064)	(0.078)	(0.077)	(0.075)
Prosociality				
Grade		0.155**	0.113*	0.262***
		(0.067)	(0.062)	(0.068)
Treatment × grade	0.142**	−0.068	0.132**	0.130*
	(0.060)	(0.072)	(0.063)	(0.068)

NOTES: All dependent variables are standardized such that the mean of the control group is 0 and the standard deviation is 1. Regressions include a full set of strata fixed effects. School-clustered standard errors in parentheses. Significance: $^{+}p < 0.15$, $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Looking at the two dimensions of time preferences, we find that the extent of self-control and patience of the 7- to 11-year-olds in our sample remains rather stable over time. Self-control has been shown to develop at relatively young ages with a first qualitative shift between the ages 3 and 7 (Montroy et al., 2016). Previous evidence from cross-sectional and panel studies on patience is mixed. Most studies document an increase in patience as children grow up for younger children up to age 10 (Bettinger and Slonim, 2007; Angerer et al., 2015; Sutter et al., 2015; Falk et al., 2021), but not for older children (Sutter et al., 2013).

The grade-specific treatment effects on self-control make a strong case for the existence of a sensitive period during early elementary school age. The treatment effects in grades 2 and 3

FIGURE 2: Treatment effects on socio-emotional skills by school grade



NOTES: Figure plots coefficients of interactions displayed in Table 8. Observations: about 880 for grade 2, 780 for grade 3, 760 for grade 4, and 760 for grade 5. For self-control, the grade 2 treatment effect is significantly different (based on t-tests for the equality of coefficients) from treatment effects in grades 4 and 5 ($p \leq 0.05$). For patience, treatment effects in grades 2 and 3 are significantly different ($p = 0.05$). For prosociality, the grade 3 treatment effect is significantly different from the treatment effects in all other grades ($p \leq 0.10$). For all other grade combinations, grade-specific treatment effects are not statistically significantly different.

(21.3 and 14.2 percent of a standard deviation, respectively) are much higher than the treatment effects in grades 4 and 5 (both smaller than 3 percent of a standard deviation and statistically indifferent from zero). Unsurprisingly, given the visual evidence in Figure 2, the point estimates in grade 2 are also statistically different from those in grades 4 and 5 (both $p \leq 0.05$). Moreover, the slight decrease in the treatment effect from grade 2 to grade 3 for self-control suggests that the time period we study (grades 2 to 5) possibly captures the fading out of an even longer sensitive period in the formation of self-control.

For patience, we find a sizeable (15 percent of a standard deviation) treatment effect in grade 2 only ($p < 0.05$). In contrast, treatment effects in grades 3 to 5 are small, ranging from -3.6 to 3.2 percent of a standard deviation, and not significant. The treatment effect in grade 2 is also statistically different from the treatment effect in grade 3 ($p = 0.05$). Together, these findings suggest grade 2 (ages 7 to 8) as a (perhaps fading away) sensitive period in the formation of patience. Our result that sensitive periods for patience are rather in early than later elementary school age has a similar spirit as the cross-sectional evidence that, on average, children's patience increases at younger ages, but no longer once they have turned 10 or older (Sutter et al., 2013).

Similar to the results on self-control, the higher treatment effect in an early period compared to later ones is in line with "the earlier, the better" findings for sensitive periods in the formation of cognitive skills (Shonkoff and Phillips, 2000; J-PAL, 2013; Heckman and Mosso, 2014).

Turning to social preferences, we observe that prosociality increases significantly during elementary school age (ages 7 to 11), as indicated by previous evidence from cross-sectional (see Sutter et al., 2019, for an overview of this literature) and panel data (Kosse et al., 2020). This suggests that prosociality is, in principle, malleable in this age range. Note, however, that cross-sectional patterns in the development of children's economic preferences are not sufficient to identify sensitive periods, i.e., when returns to investments are particularly high. In line

with the observed trajectory during elementary school age, grade-specific treatment effects are relatively large and significant throughout grades 2 to 5, with the exception of grade 3. The treatment effects for grades 2, 4, and 5 are similar in size—ranging between 13 and 14.2 percent of a standard deviation—, while the grade 3 treatment effect is −6.8 percent of a standard deviation and not statistically significant. Overall, these results indicate that grades 2, 4, and 5—and, if the grade 3-specific treatment effect was an outlier, the whole age range between 7 and 11 years—can be considered as a sensitive period for the formation of prosociality.

Overall, our findings provide a first proof of concept that designs such as ours are a valuable tool for identifying sensitive periods. Moreover, they reveal that sensitive periods differ across socio-emotional skills. Earlier investments in self-control and patience, the two dimensions of time preferences, seem to be more effective than the same investments in these skills at later stages (grades 4 and 5). In contrast, prosociality seems to be similarly malleable by the same investment at beginning and end of the age range covered by elementary school. In terms of the SfG program, our results indicate that an implementation in earlier grades is likely to be more effective than in later grades, especially as the higher returns on self-control and patience do not seem to come at the costs of lower prosociality returns.

4.3 Robustness checks

We have run several checks to test the robustness of our findings and confirm our main results. For this, in the following we use p-values that are based on randomization inference or are adjusted for multiple hypothesis testing and we estimate treatment effects when controlling for pre-treatment outcomes and additional family characteristics.

Alternative p-values: Significance levels reported so far are based on conventional critical values of t-tests. Appendix Table A6 reports p-values based on randomization inference as well as from multiple hypothesis testing adjustment.

With randomization inference, the p-values for self-control and prosociality of our overall estimates in Table 6 (referred to as the "pooled specification" in Table A6) are still well below the 5 percent significance level, see column 3 of Table A6. Next, we account for an increasing probability of false positives in the number of tested hypotheses (i.e., outcomes). Column 4 uses the adjustment of critical t-values proposed by Romano and Wolf (2005a,b), column 5 shows the p-values suggested by Westfall and Young (1993), and column 6 reports Anderson (2008)'s q-values, see McKenzie (2020) for a discussion of their different properties. Importantly, p-values for self-control and prosociality do not exceed 0.05 for any method of multiple hypothesis testing adjustment.

Table A6 also gives alternative p-values for our grade-specific treatment effects used to assess sensitive periods. The results in Table 8 in section 4.2 that are repeated in column 1 are based on a more efficient joint estimation of grade-specific treatment effects through the interaction of grade indicators and the treatment indicator. The alternative p-values in column 2 of Table A6 rely on separate estimations (i.e., by restricting the sample to include only second graders, for example). As expected, some of the point estimates are less precise (the p-values for the grade 3 effect on self-control and the grade 4 effect on prosociality become insignificant in the separate

estimations approach). However, this does not alter the overall results and our interpretation of sensitive periods remains unchanged. All grade-specific treatment effects that are significant under separate estimation remain significant under randomization inference and when applying the various forms of multiple hypothesis testing adjustment. All in all, results in Table A6 show that the overall interpretation of our results remains robust for a large number of robustness checks.

Controlling for pre-treatment variables: Appendix Tables A7 and A8 (coefficients from the latter are also plotted in Figure A5 in the appendix) give the overall and grade-specific treatment effects when controlling for (i) pre-treatment outcomes and (ii) pre-treatment outcomes and the socio-demographic family characteristics elicited at the beginning of treatment that are used in the balancing test in Table A4 in the appendix. Neither of these specifications qualitatively changes our treatment effects.

5 Conclusion

This study presents a novel design for identifying sensitive periods in the formation of children's skills and provides a first proof of concept. We propose to identify sensitive periods by investigating heterogeneity in age-(or grade-)specific treatment effects in a randomized controlled trial that assigns a given investment to treatment children of different ages and holds treatment period and intensity constant across all age groups. In such a setup, age-specific treatment effects that are substantially larger than those for the same skill at different ages point towards the existence of a sensitive period. Our design and identification strategy can also be applied in future research aiming at the identification of sensitive periods in skill formation.

The results of this study offer important insights on sensitive periods in the formation of children's time and social preferences. They thereby contribute to filling the knowledge gap regarding the timing of sensitive periods in the formation of children's socio-emotional skills. For self-control and patience, we find that returns to the same investment—participation in the Lions Quest Skills for Growing program—are substantially larger for 7- to 8-year-old than for older children. While the corresponding treatment effects are small and not significant for older children, they range between 15 and 21 percent of a standard deviation for second graders. In contrast, grade-specific treatment effects for prosociality tend to be relatively large and significant throughout elementary school age (13 to 14 percent of a standard deviation), suggesting the whole age range between 7 and 11 years as a sensitive period in the formation of prosociality.

Our findings hold broad significance. First, we contribute to the scarce, causal evidence on drivers of the formation of time preferences and prosociality in childhood (see Alan and Ertac, 2018, and Sorrenti et al., 2020, for time preferences and John and Thomsen, 2015, Alan and Ertac, 2017, Rao, 2019, Cappelen et al., 2020, and Kosse et al., 2020, for prosociality). We complement this literature by adding evidence from Bangladesh, a context that is culturally and economically distinct from those studied in previous work. Our results show that even if returns to investments in children's cognitive skills are low after age 3, returns to investments in socio-emotional skills can still be higher, as proposed by Cunha and Heckman (2007) and Borghans et al. (2008).

Second, our findings imply that the Lions Quest Skills for Growing program provides a valuable input in the process of children's skill formation. This is highly relevant as this program is widely implemented at large scale in dozens of countries worldwide, but lacks rigorous evaluations.

Finally, our findings have important policy implications. Our results on sensitive periods in the formation of time preferences are in line with the "the earlier, the better" hypothesis (see, e.g., Zhou et al., 2021), providing new evidence that earlier investments do indeed have larger returns than later ones in the domain of time preferences. We also document that the timing of sensitive periods differs across socio-emotional skills. During elementary school age, earlier investments seem to be more effective in fostering self-control and patience than the same investments at later stages. In contrast, the malleability of prosociality through the investment under study remains high throughout elementary school age. Our results thus imply that an implementation of the SfG program is likely to be more effective in earlier rather than later grades since the higher returns on patience and self-control do not come at the cost of lower returns on prosociality. As a more general take away, our findings emphasize that the same intervention may be more effective at some ages than others. Identifying sensitive periods is thus a crucial prerequisite for an effective and efficient timing of parental or public investments aimed at fostering children's skills.

References

Alan, S. and S. Ertac (2015). Patience, Self-Control and the Demand for Commitment: Evidence from a Large-Scale Field Experiment. *Journal of Economic Behavior & Organization 115*, 111–122.

Alan, S. and S. Ertac (2017). Belief in Hard Work and Altruism: Evidence from a Randomized Field Experiment. Working Paper.

Alan, S. and S. Ertac (2018). Fostering Patience in the Classroom: Results from Randomized Educational Intervention. *Journal of Political Economy 126*(5), 1865–1911.

Almlund, M., A. L. Duckworth, J. J. Heckman, and T. D. Kautz (2011). Personality Psychology and Economics. In E. A. Hanushek, S. Machin, and L. Woessmann (Eds.), *Handbook of the Economics of Education*, Volume 4, pp. 1–181. Elsevier.

Anderson, M. L. (2008). Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of the American Statistical Association 103*(484), 1481–1495.

Angerer, S., P. Lergetporer, D. Glätzle-Rützler, and M. Sutter (2015). How to Measure Time Preferences in Children: A Comparison of Two Methods. *Journal of the Economic Science Association 1*(2), 158–169.

Bauer, M., J. Chytilová, and J. Morduch (2012). Behavioral Foundations of Microcredit: Experimental and Survey Evidence from Rural India. *American Economic Review 102*(2), 1118–39.

Bauer, M., J. Chytilová, and B. Pertold (2014). Parental Background and Other-Regarding Preferences in Children. *Experimental Economics 17*, 24–46.

Bettinger, E. and R. Slonim (2007). Patience Among Children. *Journal of Public Economics 91*(1-2), 343–363.

Bolton, G. E. and A. Ockenfels (2000). A Theory of Equity, Reciprocity, and Competition. *American Economic Review 90*(1), 166–193.

Borghans, L., A. L. Duckworth, J. J. Heckman, and B. ter Weel (2008). The Economics and Psychology of Personality Traits. *The Journal of Human Resources 43*, 972–1059.

Cappelen, A., J. List, A. Samek, and B. Tungodden (2020). The Effect of Early-Childhood Education on Social Preferences. *Journal of Political Economy 128*(7), 2739–2758.

Chowdhury, S., M. Sutter, and K. F. Zimmermann (2022). Economic Preferences across Generations and Family Clusters: A Large-Scale Experiment in a Developing Country. *Journal of Political Economy*. Forthcoming.

Clarke, D., J. P. Romano, and M. Worlf (2020). The Romano-Wolf Multiple-Hypothesis Correction in Stata. *The Stata Journal 20*(4), 812–843.

Cobb-Clark, D. A., S. C. Dahmann, D. A. Kamhöfer, and H. Schildberg-Hörisch (2019). Self-Control: Determinants, Life Outcomes and Intergenerational Implications.

Cunha, F. and J. J. Heckman (2007). The Technology of Skill Formation. *American Economic Review 97*(2), 31–47.

Cunha, F. and J. J. Heckman (2008). Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation. *Journal of Human Resources 43*, 738–782.

Cunha, F., J. J. Heckman, L. Lochner, and D. V. Masterov (2006). Interpreting the Evidence on Life Cycle Formation. In E. Hanushek and F. Welch (Eds.), *Handbook of the Economics of Education*, Volume 1, pp. 697–812. Elsevier.

Cunha, F., J. J. Heckman, and S. M. Schennach (2010). Estimating the Technology of Cognitive and Noncognitive Skill Formation. *Econometrica 78*(3), 883–931.

DellaVigna, S. and M. D. Paserman (2005). Job Search and Impatience. *Journal of Labor Economics 23*(3), 527–588.

Deming, D. J. (2017). The Growing Importance of Social Skills in the Labor Market. *The Quarterly Journal of Economics 132*(4), 1593–1640.

Durlak, J. A., R. P. Weissberg, A. B. Dymnicki, R. D. Taylor, and K. B. Schellinger (2011). The Impact of Enhancing Students' Social and Emotional Learning: A Meta-Analysis of School-Based Universal Interventions. *Child Development 82*(1), 405–432.

Ericson, K. M. and D. Laibson (2019). Chapter 1—Intertemporal Choice. In B. D. Bernheim, S. DellaVigna, and D. Laibson (Eds.), *Handbook of Behavioral Economics—Foundations and Applications 2*, Volume 2 of *Handbook of Behavioral Economics: Applications and Foundations 1*, pp. 1–67. North-Holland.

Falk, A., A. Becker, T. Dohmen, B. Enke, D. Huffman, and U. Sunde (2018). Global Evidence on Economic Preferences. *The Quarterly Journal of Economics 133*(4), 1645–1692.

Falk, A., F. Kosse, P. Pinger, H. Schildberg-Hörisch, and T. Deckers (2021). Socio-Economic Status and Inequalities in Children's IQ and Economic Preferences. *Journal of Political Economy 129*(9), 2504–2545.

Fehr, E., H. Bernhard, and B. Rockenbach (2008). Egalitarianism in Young Children. *Nature 454*, 1079–1083.

Fehr, E. and K. M. Schmidt (1999). A Theory of Fairness, Competition, and Cooperation. *The Quarterly Journal of Economics 114*(3), 817–868.

Gabard-Durnam, L. and K. A. McLaughlin (2020). Sensitive Periods in Human Development: Charting a Course for the Future. *Current Opinion in Behavioral Sciences 36*, 120–128.

Gol-Guven, M. (2017). The Effectiveness of The Lions Quest Program: Skills for Growing on School Climate, Students' Behaviors, Perceptions of School, and Conflict Resolution Skills. *European Early Childhood Education Research Journal 625*(4), 575–594.

Golsteyn, B. H., H. Grönqvist, and L. Lindahl (2014). Adolescent Time Preferences Predict Lifetime Outcomes. *Economic Journal 124*(580), F739–F761.

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A Research Note. *Journal of Child Psychology and Psychiatry, and Allied Disciplines 38*(5), 581–586.

Goodman, R., T. Ford, H. Simmons, R. Gatward, and H. Meltzer (2000). Using the Strengths and Difficulties Questionnaire (SDQ) to Screen for Child Psychiatric Disorders in a Community Sample. *International Review of Psychiatry (Abingdon, England) 15*, 166–72.

Hartley, C. and W. E. Frankenhuis (2020). Editorial Overview: Sensitive and Critical Periods. *Current Opinion in Behavioral Sciences 36*, iii–v.

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

Heckman, J. J. and S. Mosso (2014). The Economics of Human Development and Social Mobility. *Annual Review of Economics 6*(1), 689–733.

Hertwig, R. and A. Ortmann (2001). Experimental Practices in Economics: A Methodological Challenge for Psychologists? *Behavioral and Brain Sciences 24*(3), 383–403.

Hess, S. (2017). Randomization Inference with Stata: A Guide and Software. *The Stata Journal 17*(3), 630–651.

Hoynes, H., D. W. Schanzenbach, and D. Almond (2016). Long-Run Impacts of Childhood Access to the Safety Net. *American Economic Review 106*(4), 903–934.

J-PAL (2013). J-PAL Youth Initiative Review Paper. Report, Abdul Latif Jameel Poverty Action Lab, Cambridge, MA.

John, K. and S. L. Thomsen (2015). School-Track Environment or Endowment: What Determines Different Other-Regarding Behavior Across Peer Groups? *Games and Economic Behavior 94*, 122–141.

Jones, D., D. Molitor, and J. Reif (2019). What Do Workplace Wellness Programs Do? Evidence from the Illinois Workplace Wellness Study. *The Quarterly Journal of Economics 134*(4), 1747–1791.

Kautz, T., J. J. Heckman, R. Diris, B. ter Weel, and L. Borghans (2014). Fostering and Measuring Skills: Improving Cognitive and Non-Cognitive Skills to Promote Lifetime Success. OECD Education Working Paper No. 110, OECD.

Kidron, Y., M. Garibaldi, E. Anderson, and D. Osher (2015). Lions Quest Skills for Growing: Implementation and Outcome Study in Wood County, West Virginia. Technical report, American Institutes for Research (AIR).

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

Knudsen, E. I. (2004). Sensitive Periods in the Development of Brain and Behavior. *Journal of Cognitive Neuroscience 16*(8), 1412–1425.

Knudsen, E. I., J. J. Heckman, J. L. Cameron, and J. P. Shonkoff (2006). Economic, Neurobiological, and Behavioral Perspectives on Building America's Future Workforce. *Proceedings of the National Academy of Sciences 103*(27), 10155–10162.

Kosse, F., T. Deckers, P. Pinger, H. Schildberg-Hörisch, and A. Falk (2020). The Formation of Prosociality: Causal Evidence on the Role of Social Environment. *Journal of Political Economy 128*(2), 434–467.

Laibson, D. (1997). Golden Eggs and Hyperbolic Discounting. *Quarterly Journal of Economics 112*(2), 443–478.

LCIF (2016). Lions Quest Universal Program Guide. Guide, Lions Clubs International Foundation, Oak Brook, IL.

Lundborg, P., D.-O. Rooth, and J. Alex-Petersen (2021). Long-Term Effects of Childhood Nutrition: Evidence from a School Lunch Reform. *The Review of Economic Studies*.

Maalouf, W., M. Stojanovic, M. Kiefer, G. Campello, H. Heikkila, and Z. El-Khatib (2019). Lions Quest Skills for Adolescence Program as a School Intervention to Prevent Substance Use—A Pilot Study Across Three South East European Countries. *Prevention Science 20*(4), 555–565.

Matischek-Jauk, M., G. Krammer, and H. Reicher (2018). The Life-Skills Program Lions Quest in Austrian Schools: Implementation and Outcomes. *Health Promotion International 33*, 1022–1032.

McKenzie, D. (2020). If It Needs a Power Calculation, Does It Matter for Poverty Reduction? *World Development 127*, 104815.

Moffitt, T. E., L. Arseneault, D. Belsky, N. Dickson, R. J. Hancox, H. Harrington, R. Houts, R. Poulton, B. W. Roberts, S. Ross, M. R. Sears, W. M. Thomson, and A. Caspi (2011). A Gradient of Childhood Self-Control Predicts Health, Wealth, and Public Safety. *Proceedings of the National Academy of Sciences 108*(7), 2693–2698.

Montroy, J. J., R. P. Bowles, L. E. Skibbe, M. M. McClelland, and F. J. Morrison (2016). The Development of Self-Regulation Across Early Childhood. *Developmental Psychology 52*(11), 1744–1762.

O'Donoghue, T. and M. Rabin (1999). Doing It Now or Later. *American Economic Review 89*(1), 103–124.

Ostrom, E., T. Dietz, N. Dolsak, P. C. Stern, S. Stonich, and E. U. Weber (2002). *The Drama of the Commons*. Washington, DC: National Academy of Sciences.

Rao, G. (2019). Familiarity Does Not Breed Contempt: Generosity, Discrimination, and Diversity in Delhi Schools. *American Economic Review 109*(3), 774–809.

Rodríguez-Planas, N. (2012). Mentoring, Educational Services, and Incentives to Learn: What Do We Know About Them? *Evaluation and Program Planning 35*(4), 481–490.

Romano, J. P. and M. Wolf (2005a). Exact and Approximate Stepdown Methods for Multiple Hypothesis Testing. *Journal of the American Statistical Association 100*(469), 94–108.

Romano, J. P. and M. Wolf (2005b). Stepwise Multiple Testing as Formalized Data Snooping. *Econometrica 73*(4), 1237–1282.

Shonkoff, J. P. and D. A. Phillips (2000). *From Neurons to Neighborhoods: The Science of Early Child Development*. Washington, DC: National Academy Press.

Sorrenti, G., U. Zölitz, D. Ribeaud, and M. Eisner (2020). The Causal Impact of Socio-Emotional Skills Training on Educational Success. IZA DP No. 13087, IZA Institute of Labor Economics.

Sutter, M., M. G. Kocher, D. Glätzle-Rützler, and S. T. Trautmann (2013). Impatience and Uncertainty: Experimental Decisions Predict Adolescents' Field Behavior. *American Economic Review 103*(1), 510–531.

Sutter, M., L. Yilmaz, and M. Oberauer (2015). Delay of Gratification and the Role of Defaults—An Experiment with Kindergarten Children. *Economics Letters 137*, 21–24.

Sutter, M., C. Zoller, and D. Glätzle-Rützler (2019). Economic Behavior of Children and Adolescents—A First Survey of Experimental Economics Results. *European Economic Review 111*, 98–121.

Tangney, J. P., R. F. Baumeister, and A. L. Boone (2004). High Self-Control Predicts Good Adjustment, Less Pathology, Better Grades, and Interpersonal Success. *Journal of Personality 72*(2), 271–324.

Taylor, R. D., E. Oberle, J. A. Durlak, and R. P. Weissberg (2017). Promoting Positive Youth Development Through School-Based Social and Emotional Learning Interventions: A Meta-Analysis of Follow-Up Effects. *Child Development 88*(4), 1156–1171.

The World Bank (2018). Program Appraisal Document on a Proposed Credit for the Quality Learning for All Program for Fourth Primary Education Development Porgram . Report No. 122705-BD, The World Bank.

The World Bank (2021). World Development Indicators Database: GDP per Capita, PPP (current international $)—Bangladesh (last accessed August 13, 2021). Online Database, The World Bank.

Tsukayama, E., A. Duckworth, and B. Kim (2013). Domain-Specific Impulsivity in School-Age Children. *Developmental Science 16*, 879–893.

Westfall, P. H. and S. S. Young (1993). *Resampling-Based Multiple Testing: Examples and Methods for p-value Adjustment*, Volume 279. John Wiley & Sons.

Zeanah, C. H., M. R. Gunnar, R. B. McCall, J. M. Kreppner, and N. A. Fox (2011). VI. Sensitive Periods. *Monographs of the Society for Research in Child Development 76*(4), 147–162.

Zhou, J., A. Baulos, J. J. Heckman, and B. Liu (2021). The Economics of Investing in Early Childhood: Importance of Understanding the Science of Scaling. In J. A. List, D. Suskind, and L. H. Supplee (Eds.), *The Scale-Up Effect in Early Childhood and Public Policy: Why Interventions Lose Impact at Scale and What We Can Do About It*. New York: Routledge, Taylor & Francis.

Appendix

Figure A1: Lions Quest Skills for Growing PreK–8 Scope and Sequence Sheet

Lions Quest
PreK-8
SCOPE AND SEQUENCE

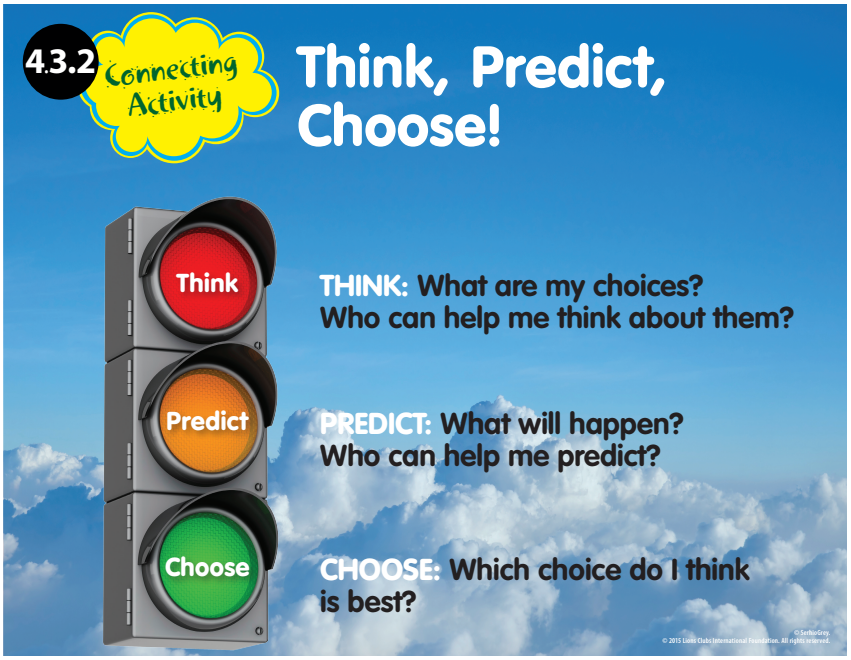
	TOPIC 1	TOPIC 2	TOPIC 3	TOPIC 4	TOPIC 5	TOPIC 6	TOPIC 7	TOPIC 8
UNIT 1: A POSITIVE LEARNING COMMUNITY	**Making Introductions** SEL Component: Self-awareness Skill: Accurate self-perception, self-confidence, clarifying your values	**Establishing Classroom Agreements** SEL Component: Self-management Skill: Impulse control	**Building Relationships and Community** SEL Component: Relationship skills Skills: Communication, social engagement, building relationships, working cooperatively					
UNIT 2: PERSONAL DEVELOPMENT	**Clarifying Your Values** SEL Component: Self-awareness Skills: Accurate self-perception, recognizing strengths	**Assessing Strengths and Growth Opportunities** SEL Component: Self-awareness Skill: Accurate self-perception	**Building Self-Confidence and Self-Respect** SEL Component: Self-awareness Skill: Self-confidence	**Motivating Yourself** SEL Component: Self-management Skill: Self-motivation	**Setting Positive Goals** SEL Component: Self-management Skill: Goal setting	**Labeling Your Emotions** SEL Component: Self-awareness Skill: Resolving conflicts	**Managing Stress and Strong Emotions** SEL Component: Self-management Skills: Stress management, impulse control, self-discipline	**Recognizing the Thoughts, Emotions, & Action Connection** SEL Component: Self-management Skills: Impulse control, self-discipline
UNIT 3: SOCIAL DEVELOPMENT	**Listening** SEL Component: Relationship skills Skill: Communication	**Respecting Others** SEL Component: Social awareness Skills: Empathy, respect for others, perspective-taking, appreciating diversity	**Communicating with "What, Why, and How" messages** SEL Component: Social awareness, relationship skills Skills: Empathy, seeking help	**Working Together** SEL Component: Relationship skills Skill: Working cooperatively	**Building Healthy Relationships** SEL Component: Relationship Skills Skill: Social engagement	**Handling Conflict in Relationships** SEL Component: Relationship skills Skill: Resolving conflicts	**Dealing with Bullying Behavior** SEL Component: Relationship skills Skills: Communication, resolving conflicts, seeking help	**Dealing with Bullying Behavior** SEL Component: Relationship skills Skills: Communication, resolving conflicts, seeking help
UNIT 4: HEALTH AND PREVENTION	**Choosing Healthy Living** SEL Component: Responsible decision making Skills: Ethical responsibility, problem identification, situation analysis	**PreK – 2: Staying Away from Poison Substances** **3 – 8: Staying Away from Alcohol** SEL Component: Responsible decision making Skills: Problem identification, situation analysis, problem solving	**Growing in Responsibility** SEL Component: Responsible decision making Skill: Ethical responsibility	**PreK – 2: Being Careful Around Medicines** **3 – 8: Staying Away from Tobacco** SEL Component: Responsible decision making Skill: Problem identification, situation analysis, problem solving	**Making Good Decisions - Part 1** SEL Component: Responsible decision making Skill: Problem solving	**PreK-2: Making Good Decisions – Part 2** **3 – 8: Staying Away from Other Drugs** SEL Component: Responsible decision making Skill: Problem identification, situation analysis, problem solving	**Standing Up to Social Pressure** SEL Component: Responsible decision making Skill: Problem identification, situation analysis, problem solving	**6 – 8 only: Reinforcing and Modeling a Healthy, Drug-Free Lifestyle** SEL Component: Responsible decision making Skill: Problem identification, situation analysis, problem solving
UNIT 5: LEADERSHIP AND SERVICE	**Serving Your School and Community** SEL Component: Relationship skills, responsible decision making Skills: Helping/seeking help, ethical responsibility	**Assessing Classroom Assets and Interests for Service-Learning** SEL Component: Relationship skills, responsible decision making skills, relationship skills Skills: Communication skills, working together, problem solving,	**Identifying Classroom, School, and Community Issues and Needs** SEL Component: Relationship skills, responsible decision making Skills: Communication, helping/seeking help, problem identification, situation analysis, problem solving	**Deciding Together on a Service-Learning Project** SEL Component: Relationship skills, responsible decision making Skills: Communication, helping/seeking help, problem identification, situation analysis, problem solving	**Planning a Service-Learning Project to Meet School or Community Needs** SEL Component: Relationship skills, responsible decision making Skills: Communication, helping/seeking help, problem identification, situation analysis, problem solving	**Implementing the Service-Learning Project** SEL Component: Relationship skills Skills: Communication, social engagement, building relationships, working cooperatively, resolving conflicts, helping and seeking help	**Reflecting on and Demonstrating the Service-Learning Project** SEL Component: Responsible decision making Skill: Reflection	**Demonstrating Service** SEL Component: Relationship skills, responsible decision making Skills: Social engagement, reflection, evaluation
UNIT 6: REFLECTION AND CLOSURE	**Reflecting on Learning, Experience, and Goals** SEL Component: Responsible decision making Skills: Reflection, evaluation	**Celebrating Class Successes and Acknowledging Contributions** SEL Component: Responsible decision making Skills: Reflection, evaluation						

6 7

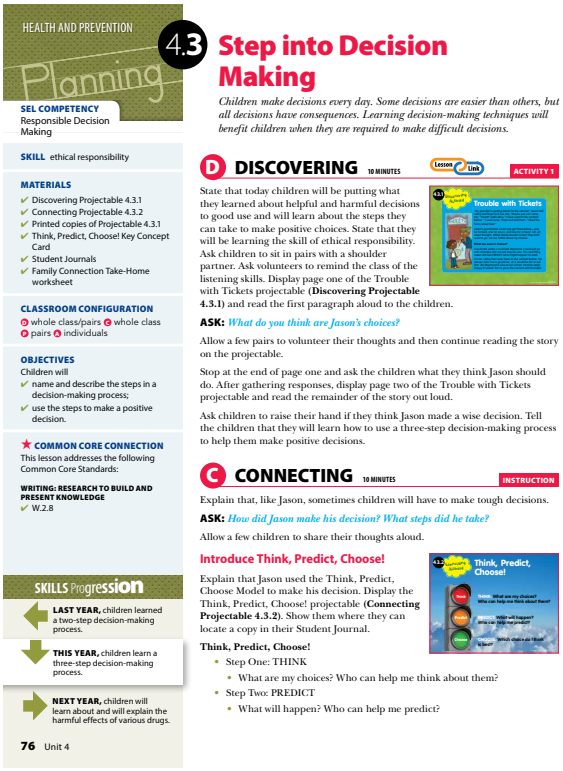
Notes: Lions Quest Skills for Growing resources.

FIGURE A2: Example of learning materials

(A) Instruction materials



(B) Teachers' resource guide


HEALTH AND PREVENTION

4.3 Planning

Step into Decision Making

Children make decisions every day. Some decisions are easier than others, but all decisions have consequences. Learning decision-making techniques will benefit children when they are required to make difficult decisions.

SEL COMPETENCY
Responsible Decision Making

SKILL ethical responsibility

MATERIALS
- ✔ Discovering Projectable 4.3.1
- ✔ Connecting Projectable 4.3.2
- ✔ Printed copies of Projectable 4.3.1
- ✔ Think, Predict, Choose! Key Concept Card
- ✔ Student Journals
- ✔ Family Connection Take-Home worksheet

CLASSROOM CONFIGURATION
whole class/pairs whole class
pairs individuals

OBJECTIVES
Children will
- ✔ name and describe the steps in a decision-making process;
- ✔ use the steps to make a positive decision.

★ COMMON CORE CONNECTION
This lesson addresses the following Common Core Standards:

WRITING: RESEARCH TO BUILD AND PRESENT KNOWLEDGE
- ✔ W.2.8

SKILLS Progression

LAST YEAR, children learned a two-step decision-making process.

THIS YEAR, children learn a three-step decision-making process.

NEXT YEAR, children will learn about and will explain the harmful effects of various drugs.

76 Unit 4

D DISCOVERING 10 MINUTES — Lesson Link — ACTIVITY 1

State that today children will be putting what they learned about helpful and harmful decisions to good use and will learn about the steps they can take to make positive choices. State that they will be learning the skill of ethical responsibility. Ask children to sit in pairs with a shoulder partner. Ask volunteers to remind the class of the listening skills. Display page one of the Trouble with Tickets projectable **(Discovering Projectable 4.3.1)** and read the first paragraph aloud to the children.

ASK: *What do you think are Jason's choices?*

Allow a few pairs to volunteer their thoughts and then continue reading the story on the projectable.

Stop at the end of page one and ask the children what they think Jason should do. After gathering responses, display page two of the Trouble with Tickets projectable and read the remainder of the story out loud.

Ask children to raise their hand if they think Jason made a wise decision. Tell the children that they will learn how to use a three-step decision-making process to help them make positive decisions.

C CONNECTING 10 MINUTES — INSTRUCTION

Explain that, like Jason, sometimes children will have to make tough decisions.

ASK: *How did Jason make his decision? What steps did he take?*

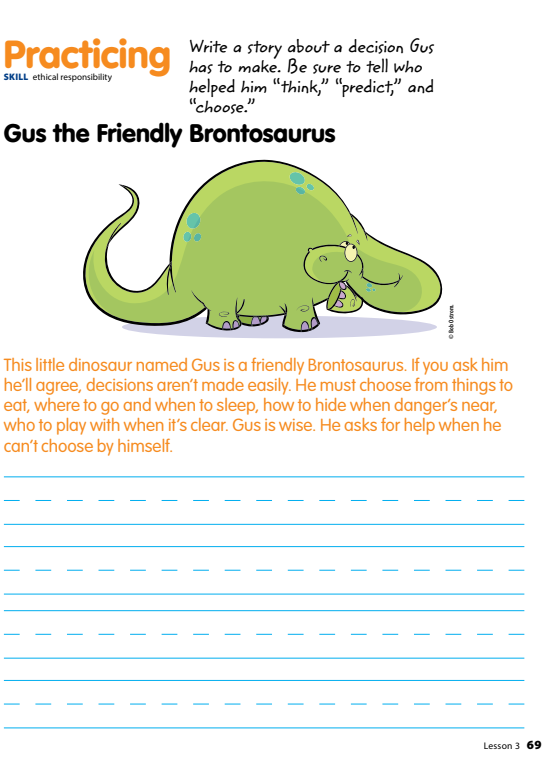
Allow a few children to share their thoughts aloud.

Introduce Think, Predict, Choose!

Explain that Jason used the Think, Predict, Choose Model to make his decision. Display the Think, Predict, Choose! projectable **(Connecting Projectable 4.3.2)**. Show them where they can locate a copy in their Student Journal.

Think, Predict, Choose!
- Step One: THINK
  - What are my choices? Who can help me think about them?
- Step Two: PREDICT
  - What will happen? Who can help me predict?

(C) Student journal


Practicing
SKILL ethical responsibility

Write a story about a decision Gus has to make. Be sure to tell who helped him "think," "predict," and "choose."

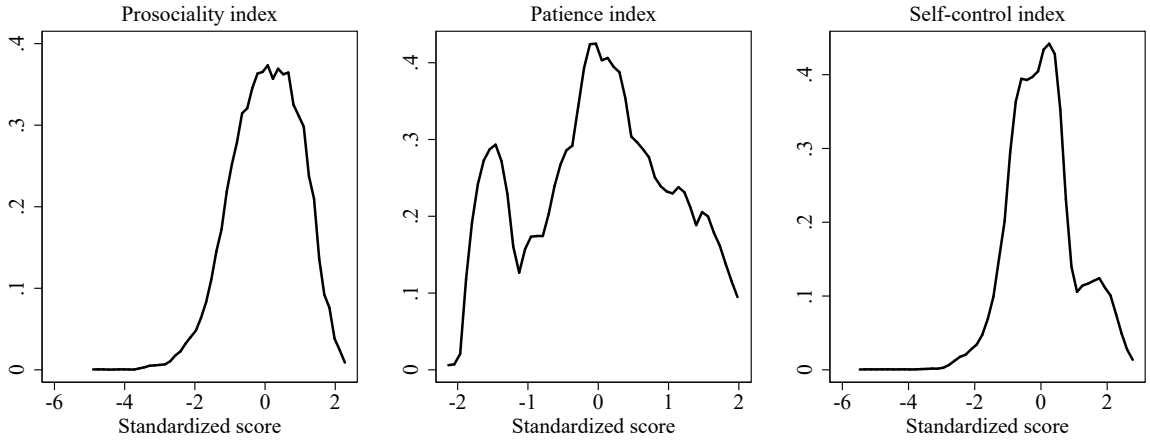
Gus the Friendly Brontosaurus

This little dinosaur named Gus is a friendly Brontosaurus. If you ask him he'll agree, decisions aren't made easily. He must choose from things to eat, where to go and when to sleep, how to hide when danger's near, who to play with when it's clear. Gus is wise. He asks for help when he can't choose by himself.

Lesson 3 69

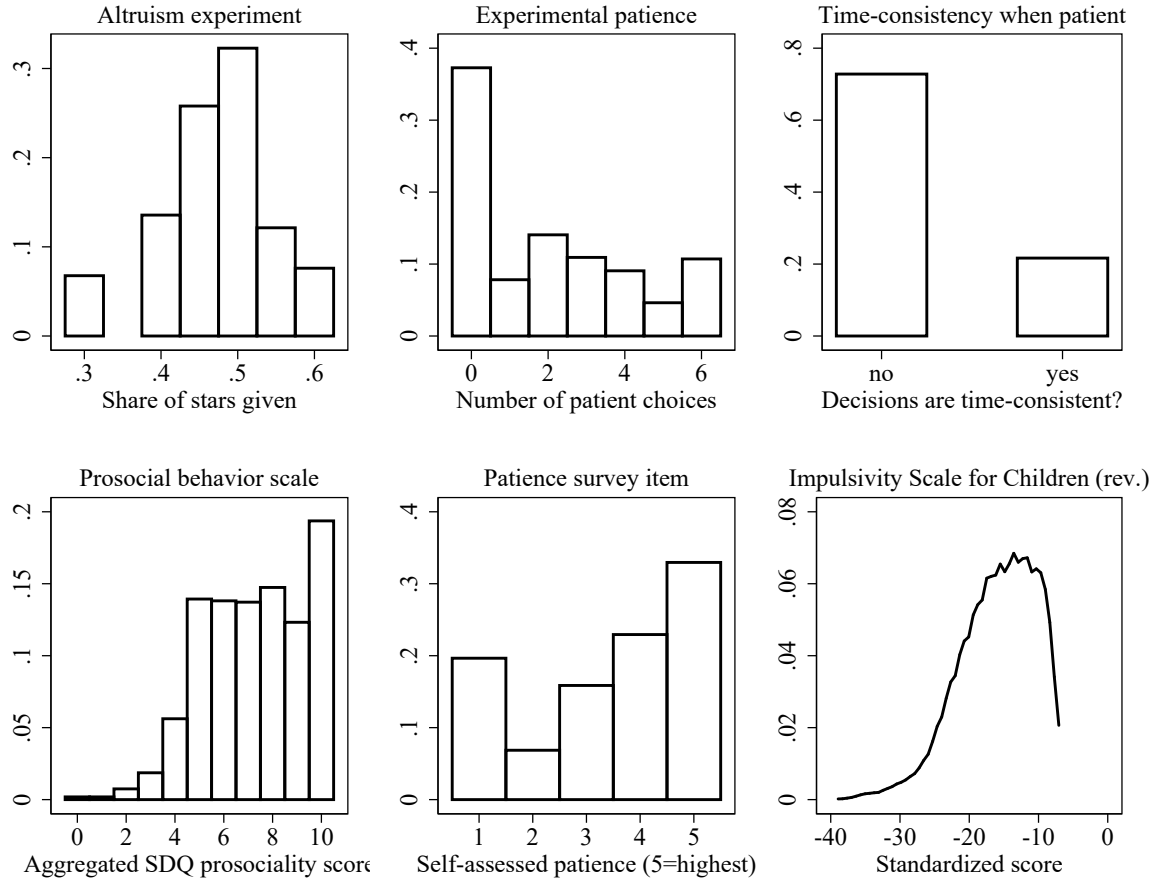
NOTES: Lions Quest Skills for Growing resources, grade 2, unit 4. English translation.

FIGURE A3: Distribution of outcome measures



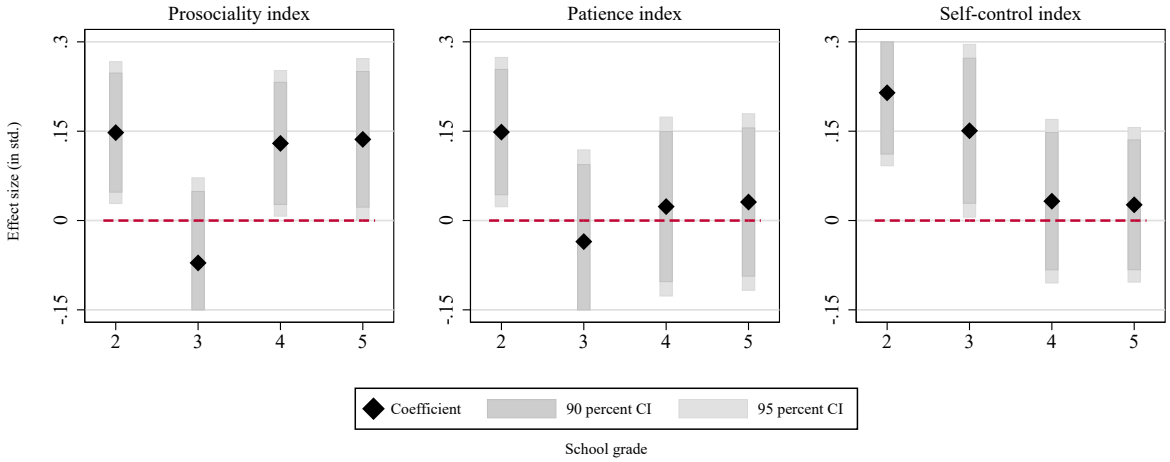
NOTES: Observations for self-control index: 3,215; patience index: 3,166; prosociality index: 3,219. All measures are standardized to have control group mean 0 and standard deviation 1. Exact aggregation procedures are described in section 3.3. Higher values indicate higher degrees of self-control, patience, or prosociality, respectively.

FIGURE A4: Distribution of outcome components



NOTES: Observations for experimental time-consistency: 3,160; ISC: 3,077; experimental patience: 3,160; patience survey item: 3,166; altruism experiment: 3,162; prosocial behavior scale: 3,109.

FIGURE A5: Treatment effects on socio-emotional skills by school grade: controlling for pre-treatment outcomes



NOTES: Figure plots coefficients displayed in column 1 of Table A8. Observations: about 880 for grade 2, 780 for grade 3, 760 for grade 4, and 760 for grade 5.

TABLE A1: Items of the Impulsivity Scale for Children

	ITEM
	Mothers assess how often the following behaviors occur on the scale 1 "almost never", 2 "about once a month", 3 "about 2 to 3 times per month", 4 "about once per week", and 5 "at least once per day":
1	My child interrupts other people.
2	My child says something rude.
3	My child loses temper.
4	My child talks back when upset.
5	My child forgets something needed for school.
6	My child cannot find something because of mess.
7	My child does not remember what someone said to do.
8	My child's mind wanders.

NOTES: Impulsivity Scale for Children (ISC) taken from Tsukayama et al. (2013).

TABLE A2: Items of the Strengths and Difficulties Questionnaire prosociality scale

	ITEM
	Mothers assess the following statements on the scale 1 "not true", 2 "somewhat true", and 3 "certainly true":
1	My child is considerate of other people's feelings.
2	My child shares readily with other children (treats, toys, pencils, etc.).
3	My child is helpful if someone is hurt, upset or feeling ill.
4	My child is kind to younger children.
5	My child often volunteers to help others (parents, teachers, children).

NOTES: Strengths and Difficulties Questionnaire (SDQ) taken from Goodman (1997). The SDQ covers 25 items in total. The displayed items are items 1, 4, 9, 17, and 20.

TABLE A3: Using alternative definitions of experimental outcome measures for treatment effect estimations

	(1)	(2)
	Treatment effect on the	
DEPENDENT VARIABLE	experimental measure	corresponding index
Panel A: Time-consistency and self-control		
Time-consistency (std., main specification)	0.054	0.107***
	(0.043)	(0.039)
Time-consistency, binary	0.022	
	(0.018)	
Time-consistency, not conditional on patience (std.)	0.024	0.087**
	(0.030)	(0.036)
Time-consistency, full sample (std.)	0.032	0.099**
	(0.042)	(0.039)
Panel B: Patience		
Experimental patience (std., main specification)	0.016	0.044
	(0.044)	(0.042)
Indicator for at least 1 patience decision (std.)	−0.002	0.035
	(0.042)	(0.041)
Experimental patience, full sample (std.)	0.011	0.035
	(0.044)	(0.041)
Panel C: Altruism measures and prosociality		
Altruism experiment (std., main specification)	0.092**	0.088**
	(0.039)	(0.036)
Altruism experiment: average share (std.)	0.076*	0.075**
	(0.040)	(0.037)
Altruism experiment: number of altruistic choices (std.)	0.082**	0.080**
	(0.038)	(0.036)

NOTES: Column 1 gives the treatment effects on alternative definitions of the experimental outcome components. Column 2 reports the treatment effects on the aggregated socio-emotional skill indices when using the experimental components as stated on the left of the table. The first row of each panel displays our preferred specification also presented in Tables 6 and 7. **Panel A.** Second row: binary time-consistency indicator that conditions on some degree of patience; third row: same as preferred specification but not conditioning on a minimum of patience; fourth row: same as preferred specification but using the full sample of children, including those who fail to answer the control question of the time preferences experiment correctly and those who prefer the later outcome over the sooner one when the interest is low, but not when the interest is high (we exclude these children in the preferred specification as this likely reflects that they have misunderstood the experiment). **Panel B.** Second row: binary indicator that is 1 if the child makes any patient choice (and 0 otherwise); third row: number of patient decisions (as in preferred specification), but using the full sample of children, as in fourth row of Panel A. **Panel C.** Second row: avg. of the share of stars given to the other child over the four games; third row: counting the number of games with altruistic decisions. If measures are standardized (indicated in table), then such that the mean of the control group is 0 and the standard deviation is 1. Regressions include a full set of strata fixed effects. School-clustered standard errors in parentheses. Significance: $^{+}p < 0.15$, $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

TABLE A4: Balancing results for covariates (measured in 2019 at the beginning of the treatment period)

VARIABLE	Definition	Uncond. mean	Treatment difference
C: female	=1 if C is female, 0 else	0.514	−0.012
C: age	C's age in years	9.430	−0.006
Number of siblings	Number of siblings (whether they live in the HH or not)	2.342	0.059
C: days of absence	C's days of absence from school due to ill health in the last 12 months	7.534	0.204
C: grade 2	=1 if C is in grade 2 during intervention, 0 else	0.276	0.003
C: grade 3	=1 if C is in grade 3 during intervention, 0 else	0.245	-0.018^{+}
C: grade 4	=1 if C is in grade 4 during intervention, 0 else	0.237	0.007
C: grade 5	=1 if C is in grade 5 during intervention, 0 else	0.242	0.008
Dwelling: brick wall	=1 if the family's dwelling has a brick wall, 0 else (usually tin wall)	0.144	0.047^{**}
Dwelling: brick floor	=1 if the family's dwelling has a brick floor, 0 else (usually mud floor)	0.143	0.022
Dwelling: area	Area family's dwelling (unit: decimal/dismil; 1 dismil ≈ 40 square meters)	9.726	−0.164
Dwelling: electricity	=1 if family's dwelling is connected to the national power grid, 0 else	0.901	−0.030
M: age	M's age in years	34.969	0.029
M: Islam	=1 if M's religion is Islam, 0 else (usually Hindu)	0.801	0.084^{+}
M: literate	=1 if M can read and write, 0 else	0.718	0.037
M: chronic illness	M's days of absence from work due to ill health in the last 12 months	10.110	0.343
F: age	F's age in years	42.591	0.146
F: Islam	=1 if F's religion is Islam, 0 else (usually Hindu)	0.789	0.092^{+}
F: literate	=1 if F can read and write, 0 else	0.577	−0.005
F: days of absence	F's days of absence from work due to ill health in the last 12 months	13.948	−0.043
F: no interview	=1 if F did not answer questionnaire, 0 else	0.243	0.011
HH income	Total HH income in 10,000 Taka (10,000 Taka ≈ US-$118), including wages, salaries, in-kind benefits, net value of agricultural products (can be negative), and cash transfers	22.976	−4.730
M: working	=1 if M is working (usually looking after live stocks or poultry, but no formal employment), 0 else	0.876	0.008
F: agricultural work	=1 if F is working in an agricultural occupation, 0 else	0.424	0.028
F: high-status job	=1 if F is a wholesale trader, labor contractor, (service-sector) employee, doctor, advocate, tutor, Imam, or receives rent; 0 else	0.089	−0.004
F: not working	=1 if F is not working (unemployed, disabled, retired), 0 else	0.007	−0.002
F: other occupation	=1 if F's occupation indicators above are all 0, 0 else	0.481	−0.023

NOTES: C=child, M=mother, F=father, HH=household. The second to the right column 3 gives the unconditional mean. The rightmost column states the treatment effect when the variable on the left is regressed on a treatment indicator. School-clustered standard errors. Significance level of the treatment effect: $^{+}p < 0.15$, $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

TABLE A5: Sample attrition

	Dependent variable:
	Attrition indicator
Panel A: Attrition w.r.t. treatment	
Treatment indicator	−0.003
	(0.008)
Constant	0.046***
	(0.006)
Panel B: Attrition w.r.t. treatment and outcomes	
Treatment indicator	−0.001
	(0.009)
Self-control index, pre-treatment	0.008
	(0.006)
– interaction with treatment	−0.015*
	(0.009)
Patience index, pre-treatment	0.002
	(0.008)
– interaction with treatment	−0.011
	(0.010)
Prosociality index, pre-treatment	0.004
	(0.007)
– interaction with treatment	0.005
	(0.009)
Constant	0.042***
	(0.007)

NOTES: Point estimates from regressions of an attrition indicator (1 if families are interviewed in 2018 but not in 2019, 0 else) on the treatment indicator (**Panel A**) and additionally on children's pre-treatment self-control, patience, and prosociality, as well as their interaction (**Panel B**). Standard errors in parentheses, clustered at school level for Panel B specification. Significance: $^{+}p < 0.15$, $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Number of observations for attrition w.r.t. treatment only: 2,716; for attrition w.r.t. treatment and outcomes: 2,665. p-value of an F-test of joint significance of treatment, pre-treatment outcomes, and interactions: 0.28.

TABLE A6: Treatment effects p-values

	(1)	(2)	(3)	(4)	(5)	(6)
	CONVENTIONAL p-VALUE		RANDOMIZATION INFERENCE	MULTIPLE HYPOTHESIS TESTING ADJUSTED		
	joint estimation	separate estimation		Romano–Wolf	Westfall–Young	Anderson's q-value
Self-control						
Pooled	0.006		0.014	0.036	0.001	0.019
Grade 2	0.001	0.000	0.003	0.004	0.000	0.001
Grade 3	0.058	0.178	0.197	0.445	0.199	0.414
Grade 4	0.689	0.891	0.881	0.978	0.969	1.000
Grade 5	0.680	0.613	0.641	0.706	0.445	0.692
Patience						
Pooled	0.295		0.358	0.306	0.129	0.110
Grade 2	0.020	0.016	0.040	0.030	0.001	0.011
Grade 3	0.644	0.653	0.690	0.649	0.509	0.414
Grade 4	0.788	0.857	0.856	0.978	0.969	1.000
Grade 5	0.676	0.471	0.540	0.706	0.445	0.692
Prosociality						
Pooled	0.015		0.036	0.046	0.004	0.019
Grade 2	0.019	0.007	0.012	0.030	0.001	0.008
Grade 3	0.345	0.195	0.275	0.445	0.199	0.414
Grade 4	0.038	0.209	0.261	0.480	0.420	1.000
Grade 5	0.058	0.017	0.039	0.074	0.002	0.054

NOTES: **The first two columns** give the p-values of the treatment effects (on the outcome stated in the panel) using the conventional critical values of the t-test. In each panel, the first row gives the p-value that corresponds to the overall results in Table 6. The other four rows ("Grade 2" to "Grade 5") state the p-values for the estimation of sensitive periods, i.e., the grade-specific effects in Figure 2. For each grade-specific effect we report two p-values: The first p-value (in column 1) results from a joint estimation, that is, the treatment indicator is interacted with grade indicators (cf. Table 8 or Figure 2). Alternatively, the grade-specific effects can be obtained through separate estimations (i.e., by restricting the sample to include only second graders, for example). The alternative p-values discussed in the following (columns 3 to 6) use the separate estimations approach. **The third column** states the p-values based on randomization inference, when we randomly assign the treatment indicator to schools within the strata 1,000 times. Randomization inference rests on the Stata `ritest` ado-file introduced by Hess (2017). **Columns 4 and 5** report the p-values when adjusting the critical t-value of the treatment indicator for multiple hypothesis testing using the techniques suggested by Romano and Wolf (2005a,b) and Westfall and Young (1993), respectively. The number of hypotheses is three, cf. section 2.3. Romano–Wolf p-values are calculated using the `rwolf` ado-file provided by Clarke et al. (2020). Westfall–Young p-values are obtained using the `wyoung` ado-file by Jones et al. (2019). In both cases bootstrapping was repeated 10,000 times as suggested by Westfall and Young (1993), see McKenzie (2020). **Column 6** gives the q-value suggested by Anderson (2008) using the accompanying syntax. All errors are our responsibility. Romano–Wolf and Westfall–Young adjustments both account for the probability of making any type-I error. This family-wise error rate (FWER) allows for a correlation of the p-values between the tested outcomes. As we do not necessarily assume that the treatment effects correlate across outcomes, we also calculate the false discovery rate (FDR) q-values. The FDR gives the expected proportion of false rejections (type-I errors) based on the number of hypotheses and their conventional p-values.

TABLE A7: Treatment effects on socio-emotional skills when controlling for pre-treatment outcomes and family characteristics

	(1)	(2)	(3)	(4)
	Controlling for pre-treatment outcomes		...additionally controlling for pre-treatment characteristics	
DEPENDENT VARIABLE	Treatment effect	Pre-treatment outcome	Treatment effect	Pre-treatment outcome
Self-control index	0.109^{***}	0.085^{***}	0.103^{***}	0.078^{***}
	(0.039)	(0.024)	(0.038)	(0.025)
Patience index	0.044	-0.014	0.045	-0.011
	(0.042)	(0.019)	(0.042)	(0.019)
Prosociality index	0.090^{**}	0.047^{***}	0.092^{**}	0.038^{**}
	(0.036)	(0.018)	(0.036)	(0.018)

NOTES: All dependent variables are standardized such that the mean of the control group is 0 and the standard deviation is 1. Regressions include a full set of strata fixed effects. School-clustered standard errors in parentheses. Significance: $^{+}p < 0.15$, $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Observations: 3,208 for self-control; 3,166 for patience; and 3,219 for prosociality. Missing values in pre-treatment outcomes are imputed with the mean and we control for the imputation.

TABLE A8: Grade-specific treatment effects on socio-emotional skills when controlling for pre-treatment outcomes and family characteristics

	(1)	(2)	(3)	(4)
	Controlling for pre-treatment outcomes		...additionally controlling for pre-treatment characteristics	
DEPENDENT VARIABLE	Treatment effect	Pre-treatment outcome	Treatment effect	Pre-treatment outcome
Self-control				
Grade 2	0.214^{***}	0.096^{*}	0.197^{***}	0.085^{+}
	(0.062)	(0.057)	(0.062)	(0.057)
Grade 3	0.151^{**}	0.068^{+}	0.151^{**}	0.053
	(0.074)	(0.041)	(0.072)	(0.042)
Grade 4	0.031	0.103^{**}	0.034	0.097^{**}
	(0.070)	(0.047)	(0.070)	(0.048)
Grade 5	0.027	0.083^{**}	0.017	0.084^{***}
	(0.066)	(0.032)	(0.065)	(0.032)
Patience				
Grade 2	0.150^{**}	0.072	0.147^{**}	0.070
	(0.064)	(0.058)	(0.065)	(0.058)
Grade 3	−0.035	-0.047^{+}	−0.031	-0.047^{+}
	(0.078)	(0.031)	(0.078)	(0.031)
Grade 4	0.021	−0.011	0.028	−0.011
	(0.077)	(0.035)	(0.076)	(0.035)
Grade 5	0.030	−0.005	0.022	−0.003
	(0.076)	(0.037)	(0.075)	(0.036)
Prosociality				
Grade 2	0.141^{**}	0.079^{+}	0.147^{**}	0.079^{+}
	(0.060)	(0.053)	(0.060)	(0.053)
Grade 3	−0.067	0.004	−0.066	0.000
	(0.073)	(0.032)	(0.074)	(0.033)
Grade 4	0.128^{**}	0.058^{**}	0.128^{**}	0.057^{**}
	(0.063)	(0.028)	(0.063)	(0.029)
Grade 5	0.135^{**}	0.041	0.146^{**}	0.034
	(0.069)	(0.033)	(0.068)	(0.033)

NOTES: All dependent variables are standardized such that the mean of the control group is 0 and the standard deviation is 1. Regressions include a full set of strata fixed effects. School-clustered standard errors in parentheses. Significance: $^{+}p < 0.15$, $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Observations: about 880 for grade 2, 780 for grade 3, 760 for grade 4, and 760 for grade 5. Missing values in pre-treatment outcomes are imputed with the mean and we control for the imputation.