MAX PLANCK INSTITUTE
FOR RESEARCH ON
COLLECTIVE GOODS

Laura Breitkopf
Shyamal Chowdhury
Daniel A. Kamhöfer
Hannah Schildberg-Hörisch
Matthias Sutter

Discussion Paper
2025/8

THE RIGHT TIMING MATTERS: SENSITIVE PERIODS IN THE FORMATION OF SOCIO-EMOTIONAL SKILLS

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

Laura Breitkopf[a], Shyamal Chowdhury[b,e], Daniel A. Kamhöfer[c,e,f], Hannah Schildberg-Hörisch[d,a,e,f], and Matthias Sutter[a,f,g]

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

[b] *Australian National University*

[c] *University of Kaiserslautern-Landau*

[d] *Heinrich Heine University Düsseldorf*

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

[f] *CESifo, Munich*

[g] *University of Cologne and University of Innsbruck*

June 18, 2025

Abstract

Identifying sensitive periods in which the returns to investments into skills are especially high is challenging, but crucial for an effective and efficient timing of parental or public investments aimed at fostering children's skills. We can detect sensitive periods with a novel design by implementing the same investment in different school grades and examining grade-specific treatment effects. Based on a randomized controlled trial with more than 3,200 Bangladeshi children in grades 2 to 5, we find sensitive periods in the formation of self-control and patience in grade 2 (age 7–8), while prosociality remains similarly malleable throughout grades 2 to 5 (age 7–11).

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, J13

---

*Corresponding author: Hannah Schildberg-Hörisch, schildberg-hoerisch@dice.hhu.de. Financial support from the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) through grant no. SCHI 1377/1 and SCHI 1377/2 and 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 Alexander Bertermann, Enzo Brox, Sarah Cattan, Sebastian Findeisen, Miriam Gensowski, Henning Hermes, Tim Kaiser, Katja Kaufmann, Fabian Kosse, Petter Lundborg, Gerhard Riener, Guido Schwerdt, Anya Samek, Astrid Marie Jorde Sandsør, Joachim Winter, Maximiliaan Thijssen, Ludger Wössmann, Jonathan Zinman, and many seminar and conference participants. 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 elementary schools, all Lions Quest trainers from India, and translators of the LQ materials 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 Lions Quest Skills for Growing 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 conditions at conception, parental characteristics, and parental and 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—specific developmental stages in which investments are particularly effective. While previous work provides evidence on the timing of sensitive periods in physical and neurological development and cognitive skills (e.g., Rutter, 1998; O'Connor et al., 2000; Knudsen, 2004; Van IJzendoorn et al., 2007; Cunha and Heckman, 2008; Barham et al., 2013; van den Berg et al., 2014; Attanasio et al., 2020), empirical evidence on sensitive periods in the formation of socio-emotional skills and in particular economic preferences is lacking. This paper aims at addressing this 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, identifying sensitive periods requires comparing returns to the *same investment into skills for children in different developmental stages.* As an important prerequisite for clean inference, the investment needs to be implemented during the same time period for all children to ensure that potential period effects do not confound treatment effects specific to certain developmental stages. Moreover, both the investment intensity and length must be held constant across different developmental stages.

In this paper, we propose and implement a novel design to empirically study sensitive periods that can be applied more broadly in future work. To enable causal inference, we set up a randomized controlled trial (RCT) in which we assign a given investment to children in different school grades (as a proxy for different stages of childhood), but not to their control group counterparts. Treatment period, length, and intensity are held constant across grades. We then measure heterogeneity in the treatment effect along grades. Following Cunha et al. (2006), we interpret grade-specific treatment effects that are substantially larger than those for the same skill in other grades as indicative of a sensitive period in the formation of this skill.

The investment we implement is a well-established social and emotional learning (SEL) program, the Lions Quest (LQ) Skills for Growing program, which is designed to provide the same investment to children in different school grades. The program aims at supporting young people confronted with the challenges of growing up. Children learn how to manage their emotions, achieve their goals, care about and build healthy relationships with others, and act responsibly. The wide-spread curriculum has a longstanding history and was designed by experts on

[1] Skills produced at one stage of childhood do not only persist but may also augment 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 are predicted to beget skills (Cunha and Heckman, 2007, 2008; Cunha et al., 2010).

child development on behalf of the Lions Clubs International Foundation. It comprises lessons on personal development, social development, as well as responsible decision-making. Since the program is implemented by the children's school teachers in the classroom environment, it relies on existing education infrastructure and does not require budget-intensive investments. Drawing on the program's detailed documentation, such as the "Universal Program Guide" (LCIF, 2016), we hypothesize that program participation increases three important socio-emotional skills: children's self-control, patience, and prosociality.

Self-control and patience are both integral to people's time preferences and intertemporal decision-making. Often modeled as present-focus, self-control problems influence 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., 2022). Patience reflects the long-run discount factor in intertemporal utility and has been shown to predict education, labor market and health outcomes, as well as savings (DellaVigna and Paserman, 2005; Sutter et al., 2013; Golsteyn et al., 2014; Alan and Ertac, 2015). Prosociality, which captures non-selfish behaviors, has been linked to both individual-level outcomes such as labor market success (Deming, 2017; Kosse and Tincani, 2020) and societal outcomes such as the provision of public goods and management of commons (Ostrom et al., 2002). To measure these skills, we combine children's revealed preferences elicited in incentivized experiments and validated survey scales answered by children or their mothers. This synthesis of lab-in-the-field and survey assessments of skills reflects the multi-dimensional nature of these underlying skills and provides a more comprehensive characterization of individuals (Falk et al., 2018; Kosse et al., 2020). Moreover, our approach reduces measurement error and potential demand effects (Hertwig and Ortmann, 2001).

In total, about 10,000 children in 69 treatment schools in Bangladesh participated in the LQ Skills for Growing program for 28 weeks with one 30-minute lesson each week. Among all second to fifth graders in the 69 treatment and 66 control schools, we randomly chose five children and their families from each class to be part of our sample. We thus collected data on self-control, patience, and prosociality of about 3,200 children in grades 2 to 5 (typically aged 7 to 11). By comparing grade-specific treatment effects of the same investment, we can learn about sensitive periods in the formation of self-control, patience, and prosociality between the ages of 7 to 11.

Our main findings can be summarized as follows. Overall, participation in the LQ Skills for Growing program significantly enhances self-control and prosociality in elementary school children. Averaging treatment effects across grades yields increases of 10.1 percent of a standard deviation for self-control and 8.3 percent for prosociality. While positive as well, the overall treatment effect on patience is smaller (3.3 percent of a standard deviation) and not significantly different from zero. Comparing treatment effects across grades reveals substantial heterogeneity. For self-control, treatment effects are substantially larger for children in grades 2 and 3 (19.5 and 11.9 percent of a standard deviation) than in grades 4 and 5 (below 5 percent), suggesting a sensitive period in the formation of self-control around age 7 to 9. Similarly, a significantly

larger treatment effect on patience in grade 2 compared to grades 3 to 5 suggests that ages 7 to 8 are a sensitive period in the formation of time preferences more generally. Grade-specific treatment effects for prosociality amount to 11 to 12 percent of a standard deviation throughout grades 2 to 5, with the exception of grade 3 (the non-significant treatment effect here is likely driven by an initial imbalance that occurred by chance). Thus, prosociality seems to be equally malleable between 7 and 11 years of age. However, malleability alone is not a sufficient condition for the existence of sensitive periods. This leads us to conclude that, while malleability is high, there is no evidence for sensitive periods in the formation of prosociality within the age range we consider.

The contribution of this study is threefold. First, as a conceptual contribution, we propose a new design for empirically assessing sensitive periods in the formation of children's skills and provide first evidence on sensitive periods in the development of children's self-control, patience, and prosociality. 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 and economic preferences, our research design can also be applied in future work. We consider our approach as complementary to the work of Cunha and Heckman (2008) and Attanasio et al. (2020). They study the impact of more abstract, simulated investments in structural estimation frameworks to investigate, among other things, sensitive periods in cognitive skills and maternally assessed behavioral problems (as a proxy for socio-emotional skills).[2] In contrast, our study relies on a tailor-made combination of experimental and survey measures of important, specific socio-emotional skills capturing time and social preferences (self-control, patience, and prosociality)[3] and focuses on an actual, easily scalable investment.

Second, our findings have important policy implications. 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). We extend the evidence that earlier investments often have larger returns than later ones to the domain of time preferences. Our results show that even if returns to investments in the cognitive skills of disadvantaged children beyond age 3 are low (Cunha et al., 2006), returns to investments in socio-emotional skills can still be higher, as hypothesized by Cunha and Heckman (2007) and Borghans et al. (2008).[4] More generally, our findings underline that the same investment may be more effective at some developmental stages than others. Therefore, knowledge about sensitive periods is crucial for an effective and efficient timing of

[2]Del Bono et al. (2020) show that parental assessments of children's socio-emotional skills are directly affected by the skills of the parents which may affect the estimates of skill production functions.

[3]See Deming (2022) for a criticism that studies on socio-emotional skills typically use whatever measures are at hand rather than relating them conceptually to particular skills as we do.

[4]Our conclusion relies on returns to the *same* intervention delivered to children of different ages. Kaiser et al. (2023) meta-analyze 11 interventions on financial decision-making capacities of children and adults and reach a similar conclusion across studies that use different interventions with different age groups: earlier interventions increase respondents' patience, while interventions with older respondents typically have null effects. By contrast, Hendren and Sprung-Keyser (2020) challenge the "the earlier, the better" paradigm. In a comparative welfare analysis of 133 historical policy changes in the United States, covering policies in social insurance, education and job training, taxes and cash transfers, and in-kind transfers, they calculate each policy's Marginal Value of Public Funds (MVPF). They conclude that MVPFs for education and health policies are large for children of all ages rather than observing diminishing marginal returns throughout childhood.

parental or public investments, including interventions that aim at enhancing children's socio-emotional skills.

Finally, we evaluate the LQ Skills for Growing program with respect to its impact on self-control, patience, and prosociality. Although the effect sizes we document are slightly smaller than those for intensive, proof-of-concept programs, such as the Perry Preschool program (Heckman et al., 2010a,b; Conti et al., 2016), that target strongly disadvantaged children only, our results indicate that available, large-scale programs can be 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 are rare (Rodríguez-Planas, 2012 and Kautz et al., 2014). For studies in the school context focusing on self-control, grit, time-consistency, or patience, see Alan and Ertac (2018), Lührmann et al. (2018), Santos et al. (2022), Schunk et al. (2022), Sutter et al. (2023), and Sorrenti et al. (2025); for results on prosociality, see John and Thomsen (2015), Alan and Ertac (2017), Rao (2019), Cappelen et al. (2020), and Kosse et al. (2020). The school-based intervention evaluated in Algan et al. (2022) documents positive effects on children's self-control and social skills, combined with comprehensive follow-up data until adulthood. Castillo et al. (2024) present results of an intervention targeting cognitive skills and executive functioning, but also measure time, risk, and social preferences. None of the studies focuses on sensitive periods, however. In light of the frequent implementation of the LQ programs in over 100 countries on the globe, rigorous evaluations, especially of the LQ Skills for Growing program targeting elementary school children, are surprisingly scarce and suffer from methodological limitations and small sample sizes. Only two studies (Kidron et al., 2015; Gol-Guven, 2017) have evaluated the LQ Skills for Growing program using a sound treatment–control group research design. Both find positive effects on school environment.[5] However, no study evaluates the LQ Skills for Growing program's impact on socio-emotional skills such as time and social preferences or in terms of sensitive periods in the formation of skills.

The remainder of this paper is organized as follows. Section 2 introduces the conceptual framework underlying sensitive periods and summarizes previous evidence. In section 3, we discuss the design of our study, the LQ Skills for Growing program, its implementation, and our hypotheses. Section 4 provides details on sampling, data collection, randomization and describes our experimental and survey measures of self-control, patience, and prosociality. Section 5 presents and discusses our results and robustness checks, before we conclude in section 6.

[5]Kidron et al. (2015) evaluate a two-year implementation in grades 3 to 5 in eight elementary schools in one county in West Virginia compared to eight control schools. They find positive effects on students' perception of their school environment as safe and supportive, self-reported interpersonal skills, and less disruptive behavior at school. However, implementation quality was so low that in the end the program had to be 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, data were only elicited for subsamples of classrooms and students), documenting positive effects on school climate, student behaviors, and conflict resolution skills. Several studies deal with the LQ 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).

2 Sensitive periods: conceptual framework and empirical implementation

In neuroscience, periods are referred to as sensitive "whenever the effects of experience on the brain are unusually strong during a limited period in development" (Knudsen, 2004, p. 1412). Cunha et al. (2006) and Cunha and Heckman (2007) integrate the notion of sensitive periods into the economic concept of skill production functions. Starting point is the general technology of skill formation (using the notation of Cunha et al., 2006): $S_t = f_t(S_{t-1}, I_t)$, where S_t is a vector of the stock of cognitive and socio-emotional skills at stage t of childhood. This stock is a concave and differentiable function $f_t(\cdot)$ that increases in contemporaneous investments, vector I_t. Vector S_{t-1} denotes the past stock of all skills which is a sufficient statistic for all past investments. In this framework, Cunha et al. (2006, p. 702) describe sensitive periods as "stages [of childhood] that are more effective in producing certain skills" than other stages.[6] Formally, according to Cunha et al. (2006, p. 803), stage t of childhood is a sensitive period for skill j when

$$\left.\frac{\partial S_{t+k,j}}{\partial I_{t+k}}\right|_{S_{t+k-1}=s, I_{t+k}=i} < \left.\frac{\partial S_{t,j}}{\partial I_t}\right|_{S_{t-1}=s, I_t=i}, \qquad \forall k \neq 0. \tag{1}$$

That is, the marginal effect of an investment in period t on skill j evaluated in period t exceeds the marginal effect of the *same investment* in period $t+k$ on skill j evaluated in period $t+k$ for all $k \neq 0$.[7]

An important feature of sensitive periods is that the investment under study is held constant but administered at different stages of childhood. In the next section, we discuss that the content of the LQ Skills for Growing program is carefully designed to deliver the same investment in different grades of elementary school and that all features of its implementation are held constant across grades. Given this setup, we will rely on heterogeneity in treatment effects along school grades to learn about sensitive periods. 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.

The definition of sensitive periods implicitly holds (i) the stock of skills in the production function and (ii) all other investments constant. We address (i) by using an RCT design which ensures that the stock of skills is comparable across treatment and control group. Moreover, patterns in self-control and patience of control group children are rather flat and not significant as children advance in grade; only for prosociality, we observe an increasing trend in grade (see Appendix Figure A1). In section 5.4, we add covariates including pre-treatment skills that proxy S_{t-1} to our parsimonious main specification and demonstrate that results remain robust. Regarding (ii), no other changes in formal investments were implemented during the intervention

[6] Critical periods are a special case of sensitive periods. If an individual does not receive a stimulus during a critical period, it may be impossible to develop a specific skill later in life. Despite this distinction, some studies use the terms sensitive period and critical period interchangeably.

[7] In addition to defining sensitive periods on the same-period outcome, Cunha et al. (2006, p. 804) provide an alternative definition that solves out the stock of skill $S_{t,j}$ as a function of the initial conditions and all lagged investments as we discuss in Appendix A.1.

period. Interpreting the return to investment I in Eq. (1) in the narrowest possible way, it holds all other investments constant. This includes possible adjustments of investments that may—perhaps even typically—result from the narrowly defined treatment (the content of the LQ Skills for Growing program) such as compensating or reinforcing behavior of parents, teachers, and children. In section 5.3, we use comprehensive measurements of observable investments of parents, teachers, and children to provide evidence that they merely change in reaction to treatment and thus cannot provide an alternative explanation for the sensitive-periods pattern that we document for time preferences.

3 Intervention, hypotheses, and implementation

Our study design builds on the LQ Skills for Growing program for two main reasons: First, the program provides well-established means to target the same skills with the same investments at different stages of childhood. It thus enables a design that allows investigating the existence and timing of sensitive periods in the formation of skills. Second, 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 (e.g., Almlund et al., 2011; Moffitt et al., 2011; Falk et al., 2018; Cobb-Clark et al., 2022).

3.1 The Lions Quest Skills for Growing program

Developed by the Lions Clubs International Foundation, a global non-profit organization, LQ programs have a longstanding history. Together with its sister programs, LQ Skills for Adolescence for middle schoolers and LQ Skills for Action for high schoolers, the LQ Skills for Growing program for elementary school children has been implemented in schools in more than 100 countries worldwide (Maalouf et al., 2019). LQ programs are classroom-based social and emotional learning (SEL)[8] programs that aim at helping young people to find their way by learning how to manage their emotions, achieve their goals, have supportive relationships with others, and act in a responsible and caring manner. For web content on the LQ Skills for Growing program, see `https://lions-quest.org` (last accessed on June 8, 2025).

According to the program's "Scope and Sequence Sheet" in Figure 1, the LQ Skills for Growing program comprises six units that each consist of several lessons (labeled as "topics"): 1) building a positive learning community, 2) personal development, 3) social development, 4) health and prevention, 5) leadership and service (optional and not implemented in the 28-week schedule that we employ), and 6) a reflection section on what has been learned. The program aims at promoting the five "SEL competencies" self-awareness, self-management, social awareness, relationship skills, and responsible decision-making. The documentation of the program translates these SEL competencies, in turn, into skills like self-discipline, impulse control, goal-setting, working cooperatively, empathy, and self-confidence, which are targeted through the various program units.

[8]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/fundamentals-of-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 June 8, 2025).

The LQ program is 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 in all grades.[9] The program's official Scope and Sequence Sheet in Figure 1 underlines that program lessons in each unit have the same goals and contents and target the same skills from pre-kindergarten up to grade 8. This feature allows us to introduce the same investment in different grades and hence to investigate possible sensitive periods in the formation of children's targeted skills.

Each LQ lesson lasts for about 30 minutes and is divided into four parts. First, the teacher uses LQ instruction materials to present an everyday situation, like a short story or pictures of someone getting bullied, and identifies together with the students why this situation is problematic ("discovery phase"). The instruction materials (e.g., example stories used) to illustrate a given issue are partly adjusted to better reflect students' cognitive development and everyday environment in the respective school grades. 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 a decision, they are trained to follow a "traffic light approach": to step back, calm down (red light), reflect on their options and the consequences (yellow light), and make a deliberate decision and implement it (green light). They also discuss how to keep up their motivation for tedious tasks by not following immediate impulses but reminding themselves of why a goal is important to them or where they have been successful in the past. Third, students reenact the presented situations in role plays or solve tasks in pairs or small groups, employing the solutions and strategies they have talked about ("practicing phase"). Finally, teachers assign 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 of the instruction materials, teachers' resource guides, and student journals (in English).

3.2 Hypotheses

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

HYPOTHESIS 1A. Participation in LQ Skills for Growing increases self-control.

HYPOTHESIS 2A. Participation in LQ Skills for Growing 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-consistency. Quasi-hyperbolic discounting (Laibson, 1997; O'Donoghue and Rabin, 1999) is the most commonly used model of intertemporal choice in behavioral economics (Ericson and Laibson, 2019), using 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 $0 < \delta \leq 1$. Parameter δ represents long-run discounting ("patience"). The parameter β indicates how much an individual favors the current period over later periods ("present focus"). If $\beta < 1$, preferences

[9] Defining inputs that deliver the same investment at different developmental stages is not a trivial task. Money, for example, can be quantified objectively, but the same amount may not constitute the same investment for children at different developmental stages. For example, 50 US-$ for school lunch per month may not be considered the same investment for a 7- and a 12-year-old child that have different recommended caloric intakes.

are time-inconsistent because individuals place more relative weight on the current period once it has arrived than in any previous period. They are thus more likely to give in to temptations and impulses in the here and now. Individuals with $\beta < 1$ have self-control problems, time-consistent individuals with $\beta = 1$ do not.

We expect the LQ program to foster impulse control and making responsible, forward-looking decisions through the lessons on personal development taught in unit 2 as well as lessons on how to best restrain oneself 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 oneself, and build healthy habits. This is likely to be reflected in a higher level of self-control. Although not directly mentioned in the LQ's Scope and Sequence Sheet, the lessons that relate to intertemporal decision-making and in particular to setting longer-term goals may also affect patience.

HYPOTHESIS 3A. Participation in LQ Skills for Growing 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 LQ unit on building a positive learning environment and particularly unit 3 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 of treated children.[10]

We are not aware of empirical evidence on sensitive periods in the formation of self-control, patience, and prosociality that could provide guidance for deriving hypotheses on the timing of sensitive periods. Yet, following the "the earlier, the better"-findings on cognitive skills, we hypothesize:

HYPOTHESES 1B TO 3B. Participation in LQ Skills for Growing has a larger impact in earlier than later grades for the formation of

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

3.3 Implementation of the Lions Quest Skills for Growing program

Despite their widespread implementation on a global scale, LQ programs had not been realized in Bangladesh before the start of our intervention. In cooperation with the Lions Clubs International Foundation, learning materials got translated and adapted to the local context (for instance, pictures of children in the materials were changed to depict Bangladeshi rather than US American or Indian children). Throughout the program's implementation, we collaborated with the Lions Clubs International Foundation to stick to standard procedures. The

[10] We also expect the LQ Skills for Growing program to increase self-esteem and accuracy of self-perception, but lacked suitable measurement tools for them. This prevents us from considering them as further outcome variables. For example, self-esteem is usually assessed by children themselves (as opposed to their parents or teachers) through survey scales. In pre-tests using the common Rosenberg scale that is generally applied from age 10 onwards, most children in our sample were indeed too young to answer its items meaningfully.

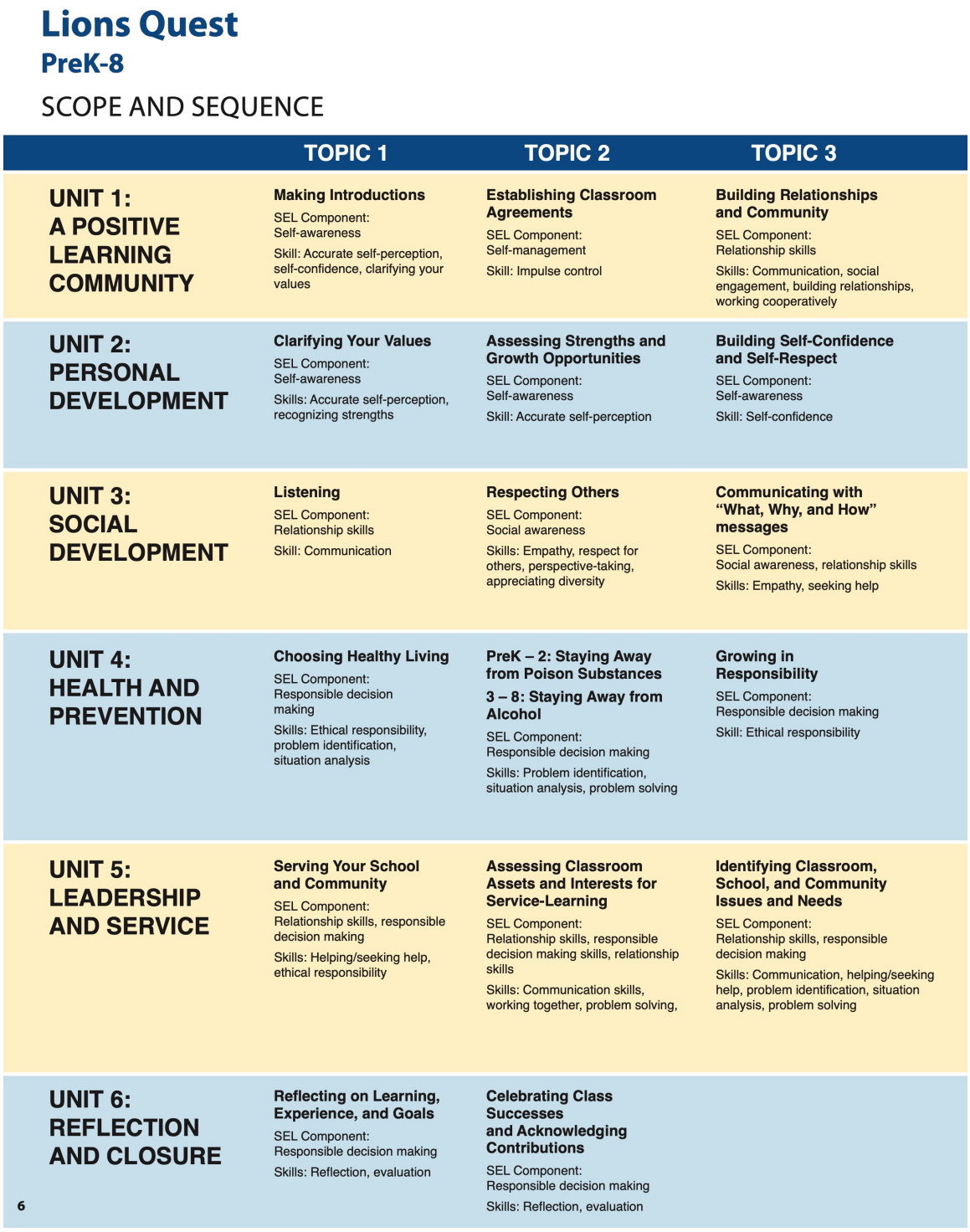
Ministry of Primary and Mass Education of Bangladesh supported the program's implementation, directing treatment schools to teach the LQ Skills for Growing program. The program was implemented by the children's elementary school teachers following the pre-existing curriculum. In the course of program adoption, school teachers got trained as LQ teachers in intensive three-day workshops by qualified LQ trainers from India and received the instruction materials and the teachers' resource guide, a 126-pages textbook with detailed instructions. Children got student journals to summarize the topics and do homework. Parents were invited to a single mid-program meeting with the LQ teachers and local education authorities in their children's schools to receive basic information about the program. We hired four research assistants who served as contact persons for the schools, monitored the program implementation, and visited treatment and control schools to conduct teacher interviews.

In order to allow for clean inference, our implementation has two further important features. First, the LQ program was implemented during the same time period for children in different grades, ruling out possibly different period effects. Second, treatment intensity was the same for all children in the treatment group, i.e., all treated children were equally long exposed to the intervention which consisted of one LQ lesson per week over the course of one school year (except for school holidays and exam periods). In order to fit the LQ program into one school year, the optional leadership and service unit was not implemented, as the Lions Club recommends for a shorter 28-week schedule.[11] The program ran from the end of January to the end of October 2019. LQ lessons were held during classes that teachers could use flexibly. Control group teachers filled these lessons in whatever way they wanted as there was no specific curriculum. Often, they were used for art classes, physical education, or children simply played in the school yard. Thus, our treatment effect has to be interpreted vis-à-vis this counterfactual. Importantly, total instruction time was constant across treatment and control schools.

Elementary school in Bangladesh is compulsory and covers grades 1 to 5, starting at age 6. Public elementary education is free of charge. Our implementation of the LQ program focused on public elementary schools' grades 2 to 5 to target young children, but, at the same time, give children in grade 1 time to accustom themselves to the new school environment before being exposed to the intervention. In general, children in grades 2 to 5 are between 7 and 11 years old. Some children are slightly older if they had to repeat classes. Following the standard implementation of the program, teachers in the treatment schools were not randomly assigned to become LQ teachers but the decision of who taught the program was left with the schools' head teachers. Given this decentralized assignment, it is not surprising that some LQ teachers' characteristics differ from those of average teachers in the control schools. LQ teachers are significantly more likely to be male, older, and more experienced, see Appendix Table A1. However, teachers' predetermined characteristics are barely correlated with students' skill outcomes (Table A2), suggesting that the decentralized teacher assignment did not influence the quality of program implementation.

[11] Given the unit's content, we do not expect that including it would do much in terms of fostering self-control, patience, and prosociality. In case it does, the estimated effects of the 28-week schedule likely represent lower bounds of the effects of the extended program.

FIGURE 1: Lions Quest Skills for Growing PreK–8 Scope and Sequence Sheet



Lions Quest
PreK-8
SCOPE AND SEQUENCE

	TOPIC 1	TOPIC 2	TOPIC 3
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
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
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
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
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

Figure continues on next page.

TOPIC 4	TOPIC 5	TOPIC 6	TOPIC 7	TOPIC 8
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
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
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
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

7

Notes: LQ Skills Universal Program Guide, pages 6–7, provided by the Lions Clubs International Foundation. In order to fit the LQ program into one school year with one lesson per week, the optional unit 5 (leadership and service) was not implemented, as the Lions Club recommends for a shorter 28-week schedule.

4 Data

This section first provides details on sampling, data collection, and the randomization procedure. We continue by describing our experimental and survey measures of self-control, patience, and prosociality, and how we construct the corresponding outcome indices. Finally, we document the balancedness of our sample and the absence of selective attrition.

4.1 Sampling, randomization, and data collection

We drew a new sample of children and their families from 135 elementary schools in four districts of Bangladesh, Netrokona, Sunamganj, Chandpur, and Gopalganj. These districts represent four of the eight administrative divisions of the country. In the course of a previous survey in 2014 and 2016 (see Chowdhury et al., 2022), 11 subdistricts were chosen based on the availability of NGOs willing to collaborate and 150 villages were randomly drawn from the 11 subdistricts.[12] In 2018, the 150 villages were visited again and a public elementary school suitable for sampling school children was chosen. Most villages have only one elementary 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 some schools serve multiple villages, the process resulted in a selection of 135 elementary schools. Based on power calculations, we decided to randomly sample five students in each of the grades 2 to 5 in each of the 135 schools using 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. Interviews with the sampled student, both parents (if available), and one randomly selected sibling (if available) were conducted by a specialized, local survey firm (ECONS Evaluation & Consulting Services Limited) at the families' homes.

We used a stratified randomization procedure to divide the 135 schools into treatment schools that implemented the LQ Skills for Growing program in grades 2 to 5 and control schools that did not. Strata are based on three criteria: the 11 subdistricts as well as median splits along the villages' (i) literacy rates (measured in a preceding village survey in 2016) and (ii) distance to the subdistrict capitals. The village literacy rate proxies a village population's educational level. In some districts, all villages were either above or below the median literacy rate. Therefore, we have 35 instead of 44 ($11\times2\times2$) strata. Larger distance to subdistrict capital may reflect lower school quality, as more rural schools are less attractive for teachers and less effectively supervised by education officials. Each school belongs to one stratum and, within each stratum, each school was randomly assigned to the treatment or control group using a random number generated in Stata. As a result, 69 schools form the treatment group, the remaining 66 schools

[12] These 150 villages were also part of a randomized study in the context of arsenic water contamination, which, however, did not directly affect our newly sampled households. After arsenic was discovered in underground drinking water, Bangladesh's government, supported by UNICEF, the World Bank, and similar organizations, tested tubewells nationwide for arsenic between 1999 and 2002, labeling each well either as green or red (indicating safe or unsafe levels of arsenic in the drinking water). Between 2014 and 2016, all tubewells in the study villages were tested and labeled again (which, according to regulations, should happen regularly anyhow but is not always enforced) as part of an RCT. This RCT was conducted in our sample villages to assess the effectiveness of a public information campaign about possible arsenic contamination of drinking water, similar to an earlier government campaign, as well as the promotion of arsenic filters to encourage households to switch to arsenic-free drinking water. Since we drew a new sample for this study, the households in our sample were not exposed to the repeated information campaign or the option to purchase a filter.

serve as control group.[13] As the treatment was part of the curriculum in grades 2 to 5 of the treatment schools, children could not drop out of the treatment single-handedly. With 98.7 percent the compliance rate is very high.[14]

Due to unanticipated delays, program implementation started only in 2019 and the originally sampled students were one grade higher than initially expected. To maintain the intended sample composition, the post-treatment data collection comprised a randomly selected refreshment sample of new second graders for whom we do not have pre-treatment, baseline information. Our final sample consists of 3,263 children from 2,842 families, see Table 1. The randomly sampled siblings have the same treatment status as the initially sampled students if they attend any of grade 2 to 5 in the same elementary school in 2019.

Table 1: Sample overview

	(1) All	(2) Treatment group	(3) Control group
Number of schools	135	69	66
Number of students	3,263	1,647	1,616
– sampled students	2,842	1,439	1,403
– siblings	421	208	213
By grade			
– grade 2	896	455	441
– grade 3	803	391	412
– grade 4	773	393	380
– grade 5	791	408	383

As a result of this recruitment process, we have a large sample of families in which we comprehensively measured both children's and parents' skills. Our analysis focuses on two waves of data on children's and parents' skills, see Figure 2: a baseline wave before the treatment in 2018 and a short-term, immediate post-treatment skill assessment (11/2019–2/2020).[15] All skill assessments elicited economic preferences (time, risk, and social preferences), personality traits, and cognitive skills via paper-and-pencil interviewing for the sampled children, up to

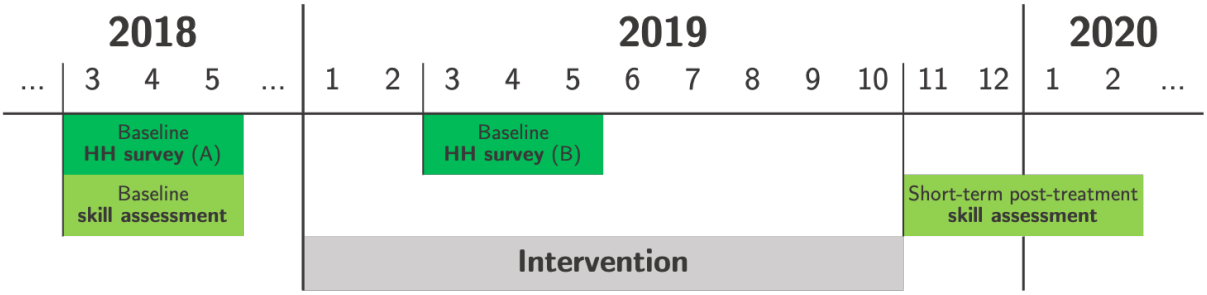
[13]Since not all strata contained an even number of schools, randomization within strata led to an unequal number of treatment and control group schools.

[14]Teachers of one treatment school with 22 out of 1,645 sampled students did not receive the LQ teacher training due to miscommunication and the school did not implement the treatment. We present conservative intention-to-treat estimates throughout the paper, i.e., students in this school still belong to the treatment group.

[15]Note that the post-treatment assessment ended before the outbreak of Covid-19. We have also collected two further waves of panel data, about 15 and 32 months after the treatment, respectively. However, for two reasons, we do not analyze these data here. First, our focus is investigating sensitive periods according to the definition of Cunha et al. (2006, p. 803) in Eq. (1) that refers to *immediate* skill returns of the same investment at different stages of childhood—from that perspective, the additional data are not helpful. Second and most importantly, both further waves of data were elicited during the Covid-19 pandemic that constitutes a huge, negative shock on children's skill formation. In Bangladesh, schools were closed for a total of 20 months and many children, in particular boys, were forced to leave school permanently to start working in order to contribute to the decreasing family income (TIME magazine, 2022). As a consequence, the additional data cannot inform us about the longer-run treatment effects under normal conditions.

one sibling, and their parents. Mothers also answered a questionnaire about their children, assessing, among other things, their children's strengths and difficulties (including prosociality), their self-control, and their own parenting style.[16] Moreover, we conducted several household surveys (see Figure 2 for their timing) to collect survey information on socio-demographics, income, expenditures, employment, land ownership, credits and savings, and assets. Household surveys were 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. Importantly, there was no visible connection between the treatment at school and the interviews at home. Neither the skill assessments nor the household surveys included any reference to the LQ Skills for Growing program and the interviewers were not aware of it to avoid social desirability bias or interviewer demand effects.

FIGURE 2: Timeline of data collection and intervention



NOTES: Own representation.

Appendix Table A6 provides a detailed description of our sample based on the 2019 household survey. Children's mean age is 9.5 years and 51.4 percent are girls. On average, yearly household income is around 229,000 Taka (approximately 2,700 US-$ in 2019). An electricity connection is available in 90.0 percent of households. Fathers' mean age is 42.5, mothers' mean age is 35.1. 59.4 percent of fathers and 72.0 percent of mothers can read and write. Almost all fathers and 80.4 percent of mothers are working. The latter are usually looking after the family's live stocks or poultry instead of being formally employed.

4.2 Outcome variables

We measure children's self-control, patience, and prosociality in a particularly comprehensive manner, using both children's revealed preferences in incentivized experiments and well-established survey scales. This combination of lab-in-the-field and survey assessment results in measures that (i) reflect the multi-dimensional nature of the underlying skills, (ii) combine the advantages of incentive-compatible experiments and validated psychological survey questionnaires, and (iii) use information from multiple sources—children and their mothers (Falk et al., 2018; Kosse et al., 2020).

[16] 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 LQ program. Following the pre-registration, we do not analyze treatment effects on these skills.

Experiments: Time and social preferences

Although we relied on well-established measurement tools to elicit time and social preferences, we carefully pre-tested all items in our context. We used standardized control questions to verify that participating children understood the instructions.[17] The order of the experiments was randomly determined by rolling a die. Children earned stars which were transformed into money after the experiments using age-specific exchange rates proportional to pocket money (depending on children's age, one star's value ranged between 10 and 30 Taka, 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 others' 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 made 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 very soon (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) after the interview, 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 our experimental measure of self-control, we use children's *time-consistency* in choice sets 2 and 3 which have the same three-months time delay but different early payments dates, keeping payment amounts fixed. 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. To disentangle time-consistency from extreme impatience, we only classify children as time-consistent if they exhibit some degree of patience by choosing the larger, but later reward at least once in choice sets 2 and 3. Our findings do not hinge on this restriction as we show in Appendix Table A3.

[17] Interviewers asked children to repeat their explanations (four times during the time preferences game and once in the social preferences game). Each time, the interviewer noted down whether the child understood the game after the first, second, or third explanation. In our main specification, we keep children who answered all control questions correctly after at most three explanations and drop those who did not understand a game even after three repeated explanations. Including children who failed the control questions does not alter our findings, see Appendix Table A3.

Our experimental measure of *patience* is a simple count of the number of patient choices in all six decisions, i.e., the number of larger, but later reward choices. This measure ranges from 0 to 6.

Social preferences: altruism. We followed an experimental protocol by Fehr et al. (2008) which got extended by Bauer et al. (2014) to measure 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 (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 calculate the share of stars a child gave to the other child across all four games relative to the overall number of stars a child allocated to herself and to the other child. This share varies between 0.29 and 0.58.

Survey measures

Self-control: Impulsivity Scale for Children. We use the mother-rated, reversely coded Impulsivity Scale for Children (ISC) (Tsukayama et al., 2013) that captures children's impulsive behavior in social contexts and with respect to schoolwork. Appendix Table B1 lists its eight items that are rated on a five-point Likert scale. For example, mothers stated how often their child loses temper. All items receive equal weights when they are combined 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., a larger reward)" applies to them on a five-point Likert scale from 1 ("not at all right") to 5 ("absolutely right"). This survey item is a child-adjusted version of the one used in the Global Preference Survey by Falk et al. (2023).

Prosocial behavior. We make use of the prosociality scale of the well-established 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 their children's prosocial behavior such as "Considerate of other people's feelings" or "Shares readily with other children (treats, toys, pencils, etc.)" on a three-point scale, see Appendix Table B2 for all five items. The equally-weighted answers are combined into one scale.

Aggregation into indices

To assess children's self-control, patience, and prosociality comprehensively, we combine information from experiments and surveys into three aggregate indices that reflect the multidimensional nature of each underlying skill. The indices 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 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 two components is missing for a child, we use the remaining component only. This results in an increase from 3,075 to 3,258 observations for the self-control index, from 3,207 to 3,213 observations for the patience index, and from 3,109 to 3,262 observations for the prosociality index.

Self-control. The self-control index combines experimentally elicited time-consistency with the reversed ISC.

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

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

Appendix Figure A3 displays the post-treatment distribution of the standardized outcome indices, while Figure A4 shows the distribution of each outcome component.[18] Appendix Table A5 demonstrates that children's socio-emotional skills are positively correlated with parents' socio-emotional skills, as one would expect (see Appendix B.4 for how parents' socio-emotional skills are measured).

4.3 School and teacher questionnaires

In addition to surveying children and their families, we collected data on the 135 schools in our sample and, in general, two teachers per school (277 teachers in total). In control schools, we randomly selected two teachers, whereas we interviewed the LQ teachers in treatment schools.[19] The school questionnaire includes information on the school's size and facilities (see Appendix

[18]The self-control components as well as the prosociality components seem to be complements rather than substitutes. The Pearson correlation coefficient is -0.004 ($p = 0.862$) for the self-control components and 0.042 ($p = 0.101$) for the prosociality components. The correlation between the patience components is positive, medium in size, and significantly different from zero (0.344, $p < 0.01$).

[19]As we interviewed children at home, ensuring that there is a barrier between the interview and the intervention, we cannot assign students directly to their teachers. Instead, we assign students to teachers based on the school and grade. Out of the 540 school–grade combinations (135 schools, 4 grades per school), 112 are taught by only one of the interviewed teachers; for 415 combinations, we take the average characteristics of the teachers in this school and grade; and 13 school–grade combinations (encompassing 78 students) are not taught by any of the interviewed teachers, see Appendix Figure A5.

Table B3 for all variables and their definitions) and was answered by the head teachers. The teacher questionnaire includes questions about teachers' socio-demographics (see Table B4) as well as their teaching style (Table B5). Moreover, LQ teachers answered an additional questionnaire on the experiences they made with the LQ program (Table B6).

We use the school information to assess the balancedness of treatment and control schools' characteristics. Teacher information allows us to gauge how schools implemented the program and whether the program changed teaching styles.

4.4 Baseline balance and attrition

Baseline imbalance and selective attrition are potential threats to identification in randomized controlled trials. Since we have collected information on the self-control, patience, and prosociality indices 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.

As a first balancing test, we regress the pre-treatment indices (assessed in 2018, half a year before the intervention started) 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 Kolmogorov-Smirnov test
Self-control index, pre-treatment	2,504	−0.052	0.305	0.207
Patience index, pre-treatment	2,483	0.031	0.510	0.867
Prosociality index, pre-treatment	2,513	−0.051	0.308	0.488

Notes: Own calculations based on the estimation sample. 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 and strata fixed effects. School-clustered standard errors.

As a second randomization check, we regress 36 pre-treatment, socio-demographic child and family characteristics on the treatment indicator. Panel A of Table 5 summarizes the results. Appendix Table B9 provides definitions for all 36 variables, Table A6 states their means, the treatment coefficients, and corresponding p-values. Under successful randomization, the actual number of significant differences between treatment and control group (displayed in the second row of columns (2) to (5) of Table 5) should be similar to the number of significant differences we expect to observe by chance for a given significance level (see first row of Panel A). Indeed, the actual number of significant treatment coefficients is in line with the expected one. Column (6) approaches the multiple-testing problem when comparing "treatment effects" across a large number of variables from another angle and shows the p-value of a 'stacked F-test' that assesses the joint significance of differences in the 36 variables between treatment and control group (Lee and Lemieux, 2010). The coefficients of the "treatment effects" are not jointly significantly different from zero ($p = 0.61$). Panel B of Table 5 repeats this procedure for the grade-specific

"treatment effects" on the pre-treatment outcome indices. One out of twelve pre-treatment outcomes differs significantly between treatment and control group: treatment children in grade 3 are, on average, significantly less prosocial than their counterparts in control schools. Although this difference has to be kept in mind when analyzing sensitive periods, it is not entirely surprising given that we test a total of twelve hypotheses. The corresponding stacked F-test yields $p = 0.06$. Drawing on interviews with the head teachers in Panel C, we are also able to consider the balancedness of treatment and control schools' characteristics (see Appendix Table A7 for detailed results). In line with expectations for a successful randomization, only one of seven elicited characteristics differs at the 10 percent level. In treatment schools, students in different grades tend to be less likely to be taught in the same classroom. The corresponding stacked F-test yields $p = 0.86$. In sum, using rich information on pre-treatment outcome indices, child, family, and school characteristics, treatment and control group appear to be balanced.

TABLE 5: Balancing overview for family and child characteristics, pre-treatment outcomes on grade level, and school characteristics

	(1)	(2)	(3)	(4)	(5)	(6)
		Significance level				F-test
Set of variables	Number of variables	1%	5%	10%	15%	p-value
Panel A: Family and child characteristics						
Expected number of significant effects	36	0.36	1.8	3.6	5.4	
Actual number of significant effects		0	1	2	4	0.610
Panel B: Pre-treatment outcomes on grade level						
Expected number of significant effects	12	0.12	0.6	1.2	1.8	
Actual number of significant effects		0	1	1	1	0.061
Panel C: School characteristics						
Expected number of significant effects	7	0.08	0.35	0.7	1.05	
Actual number of significant effects		0	0	1	1	0.860

NOTES: Own calculations based on the estimation sample and the school data. This tables summarizes the results of the balancing checks. In **Panel A**, we regress a total of 36 child and family characteristics on the treatment indicator and strata fixed effects, using school-clustered standard errors. The first row gives the number of variables (out of the 36 child and family characteristics) for which we expect the treatment indicator to be significant at the respective significance level in the column header. The second row gives the actual number of variables for which the treatment indicator is significant. The number of variables accumulates from the left to the right, i.e., an indicator that is significant at the 1 percent level is also counted as significant at the other three levels. Table B9 provides definitions of all 36 variables, while Table A6 states their means, treatment coefficients, and corresponding p-values. Column (6) displays the p-value of a 'stacked F-test' as suggested by Lee and Lemieux (2010). It tests for joint significance of differences in the variables between treatment and control group by testing for joint significance of interaction terms of the variables and the treatment indicator in a stacked regression. The characteristics in Panel A are taken from the baseline household survey (B) in 2019, at the beginning of the treatment period, see Figure 2. We use the household survey (B) because it contains more observations, but most characteristics do not change over time. **Panel B** summarizes the balancing checks for the pre-treatment outcome indices by grade (comparable to the analysis in Table 4). We regress the pre-treatment self-control index, patience index, and prosociality index on the treatment indicator interacted with grade indicators and strata fixed effects, using school-clustered standard errors. This results in a total of 12 coefficients. The interpretation of the columns is the same as in Panel A. **Panel C** summarizes the balancing results at the school level, using the seven variables elicited in interviews with the head teachers. The specification is the same as in Panel A. Table B3 defines the seven school variables and Table A7 displays their means, treatment coefficients, and corresponding p-values.

Attrition is 14.5 percent between the pre- and post-treatment skill assessments that were about 20 months apart (see Figure 2). 80.0 percent of overall attrition is due to children changing

or leaving school or repeating grade 1 such that they were not part of our post-treatment intervention sample in 2019 (i.e., not attending grades 2 to 5 in one of the 135 sampled schools). Children's non-participation in the second skill assessment although they were still part of our intervention sample is causing the remaining 20.0 percent of overall attrition. Attrition is thus 2.9 percent when focusing on non-participation only. Appendix Table A8 demonstrates that our results are not driven by selective attrition. It shows the results of regressing the overall attrition indicator on the treatment indicator in column (1), the treatment indicator and strata fixed effects in column (2), and additionally children's pre-treatment self-control, patience, and prosociality as well as their interaction in column (3). All coefficients are economically small and none is significant.

5 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 5.2 then exploits the specific features of our design that enable estimating sensitive periods in the formation of these skills. Section 5.3 discusses our findings, section 5.4 contains numerous robustness checks.

5.1 Overall treatment effects on socio-emotional skills

Table 6 displays the treatment effects of the LQ Skills for Growing program on children's self-control, patience, and prosociality, measured between one and four months after the end of the intervention. The corresponding skill 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 $\boldsymbol{\phi}$:

$$Y = \alpha + \beta \text{treatment} + \boldsymbol{\phi} + \varepsilon. \quad (2)$$

Coefficient β reflects the overall treatment effect. Standard errors are clustered at the school level.

In line with hypotheses 1A and 3A, we find that program participation increases children's self-control by 10.1 ($p < 0.01$) and prosociality by 8.3 ($p < 0.05$) percent of a standard deviation. Regarding hypothesis 2A on patience, we find a positive, but smaller effect size of 3.3 percent of a standard deviation that is not statistically different from zero. As we show later, these aggregate results hide substantial heterogeneity across grades.

Putting the effect sizes into perspective underlines their economic significance. For example, the size of the treatment effect on self-control is equivalent to about half of the gender gap in self-control in the control group (on average, girls have 21.1 percent of a standard deviation higher self-control than boys). Moreover, the treatment effect corresponds to about two-thirds of the self-control gap between control group children from below- and above-median income households (15.1 percent of a standard deviation). The treatment-induced increase in prosociality exceeds the average increase in prosociality that we observe when students advance one grade (about 7.5 percent of a standard deviation per grade, see the panel on prosociality in

TABLE 6: Overall treatment effects

Dependent variable	(1) Number of observations	(2) Treatment effect
Self-control index	3,258	0.101***
		(0.038)
Patience index	3,213	0.033
		(0.042)
Prosociality index	3,262	0.083**
		(0.036)

NOTES: Own calculations based on the estimation sample. Outcome variables are defined as described in the text based on an experimental and a survey measure. The number of observations is given in column (1); fewer observations compared to Table 1 are due to missing information in the outcome variables (if both components are missing). The estimation results stated in column (2) follow model (2), that is, we regress the outcome variable on the treatment indicator and a full set of strata fixed effects. All dependent variables are standardized such that the mean of the control group is 0 and the standard deviation is 1. School-clustered standard errors in parentheses. Significance at $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Appendix Figure A1). The size of the treatment effect also corresponds to about three-quarters of the gender gap in prosociality (11.3 percent of a standard deviation, with higher values for control group girls than boys).[20] Appendix Table A3 uses alternative definitions of the experimental outcome components to provide robustness checks for the overall treatment effects and demonstrates that they do not alter the interpretation of our results.

For completeness, Appendix Table A4 displays the treatment effects on each of the six components of the socio-emotional skill indices.[21] It conveys several insights. First, all coefficients are positive, underlining that the treatment tends to uniformly increase the various facets of self-control, patience, and prosociality. Second, the increase in self-control is largely driven by a decrease in children's impulsive behavior. The treatment effect on the reversed Impulsivity Scale amounts to 10.9 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 LQ program (see Figure 1), this confirms expectations. Third, the increase in prosociality mainly originates from more altruistic behavior as measured in the dictator game experiments, with a treatment effect of 9.4 percent of a standard deviation ($p < 0.05$). Although the treatment

[20] These comparisons do not imply that the treatment closes or reduces the gender gap. Appendix Table A9 documents that treatment effects on self-control, patience, and prosociality do not differ statistically significantly for boys and girls, by parents' literacy, or by a median split in family income. In Table A10, we re-run the treatment effect estimations for children with below- and above-median pre-treatment socio-emotional skill and IQ levels as well as their parents' socio-emotional skills. Treatment effects on self-control and patience are significantly larger (at the 10 percent and 5 percent level, respectively), if the child's father has an above-median level of this skill; and the one on prosociality is significantly larger (at the 10 percent level), if the mother's prosociality is above the median. This could indicate that later investments (the LQ program) are more effective, when building on higher earlier investments (parents' inputs as role models)—as hypothesized by Cunha et al. (2006).

[21] Since we calculate the indices as the average of the non-missing components, the treatment effect on the index does not equal the unweighted average of the treatment effects on the components.

effect on the prosocial behavior scale is positive, it is small and not statistically different from zero. A significant effect on the experimental preference component, but not the survey scale suggests the absence of experimenter demand effects since the observed increase in altruistic behavior in the dictator game incurs monetary costs on children.

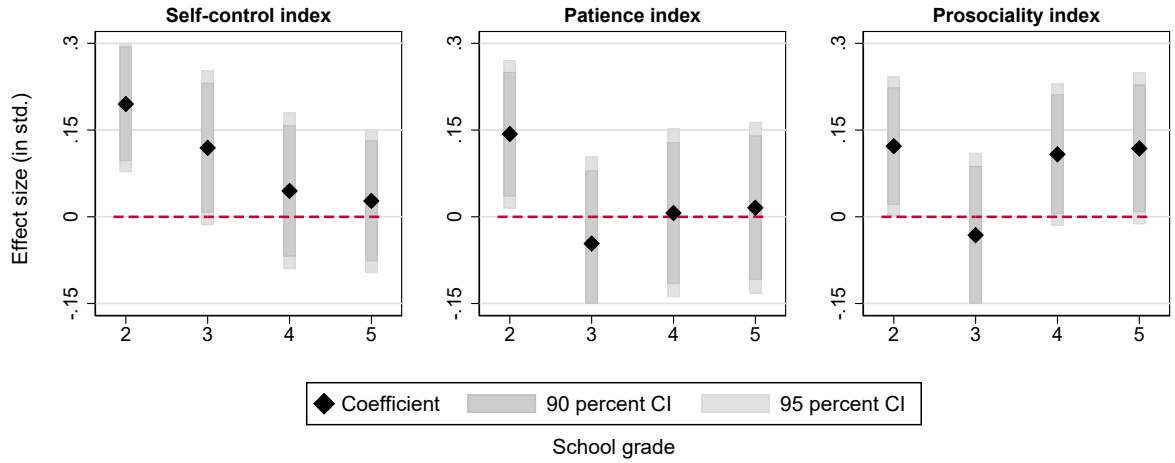
5.2 Sensitive periods in the formation of socio-emotional skills

We continue by exploiting the specific design elements of our randomized controlled trial that enable us to investigate sensitive periods in the development of socio-emotional skills—namely that we implemented the same investment (the LQ Skills for Growing program) in different school grades (i.e., at different stages of childhood), while holding intensity as well as start and end date of the investment constant. Given these features, comparing treatment effects across grades informs us about possible sensitive periods in the formation of self-control, patience, and prosociality in the age range we consider (ages 7 to 11). In order to estimate grade-specific treatment effects, we regress the skill indices Y on grade indicators (with grade 2 as omitted category), all four grade indicators interacted with the treatment indicator, and strata fixed effects ϕ:

$$Y = \sum_{g=3}^{5} \Big(\omega_g 1(\text{grade} = g)\Big) + \sum_{g=2}^{5} \Big(\gamma_g 1(\text{grade} = g) \times \text{treatment}\Big) + \phi + \varepsilon, \tag{3}$$

where 1(grade=g)=1 if grade=g, 0 otherwise. Table 7 reports the coefficients of the grade indicators (ω_g) and the grade–treatment interactions (γ_g); Figure 3 illustrates the grade-specific treatment effects. Grade-specific treatment effects qualify as sensitive periods if they are substantially larger than treatment effects in other grades, implying a higher immediate return to investment.

FIGURE 3: Graphical representation of sensitive periods



NOTES: Own calculations based on the estimation sample. This figure plots the coefficients of the interaction terms reported in Table 7. Vertical bars indicate the 90 percent (darker gray) and the 95 percent significance level (lighter gray). Appendix Figure A6 repeats the results, additionally reporting p-values for differences between the grade-specific treatment effects.

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

TABLE 7: 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 index				
Grade		−0.057	−0.099	0.014
		(0.050)	(0.068)	(0.063)
Treatment × grade	0.195***	0.119*	0.045	0.027
	(0.060)	(0.068)	(0.069)	(0.064)
N	896	803	771	788
Patience index				
Grade		−0.049	−0.183***	−0.094
		(0.060)	(0.064)	(0.070)
Treatment × grade	0.143**	−0.046	0.007	0.016
	(0.065)	(0.077)	(0.074)	(0.076)
N	891	793	757	772
Prosociality index				
Grade		0.113	0.122*	0.255***
		(0.070)	(0.062)	(0.068)
Treatment × grade	0.122**	−0.032	0.108*	0.118*
	(0.062)	(0.073)	(0.063)	(0.067)
N	896	803	773	790

NOTES: Own calculations based on the estimation sample. The estimation follows model (3), that is, we regress the outcome variable on grade indicators (grade 2 as omitted baseline), indicators for the treatment–grade interaction, and strata fixed effects. All dependent variables are standardized such that the mean of the control group is 0 and the standard deviation is 1. *N* refers to the number of students observed in each grade, not the number of observations in the regression. School-clustered standard errors in parentheses. Significance at $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Figure 3 illustrates the coefficients of the interaction terms graphically. Appendix Figure A6 additionally reports the *p*-values for the differences of the grade-specific treatment effects.

(19.5 and 11.9 percent of a standard deviation, respectively) are much larger than the treatment effects in grades 4 and 5 (both smaller than 5 percent of a standard deviation and statistically indistinguishable from zero). The point estimate in grade 2 is also statistically different from the one in grade 5 ($p < 0.05$), see Appendix Figure A6. The slight decrease in the treatment effect from grade 2 to 3 suggests that the stages in childhood we study (grades 2 to 5) possibly capture the fading out of a longer sensitive period in the formation of self-control.

For patience, we find a sizeable (14.3 percent of a standard deviation) treatment effect in grade 2 only ($p < 0.05$). Treatment effects in grades 3 to 5 are small, ranging from −4.6 to 1.6 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. Similar to the results on self-control, the larger treatment effect in an early stage 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; Heckman and Mosso, 2014).

Looking at cross-sectional patterns in the two dimensions of time preferences (see Appendix Figure A1), we find that the extent of self-control and patience of the 7- to 11-year-olds in the control group remains rather stable as children grow. 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 stems from WEIRD (Western, Educated, Industrialized, Rich, Democratic) countries and is mixed. Most studies document an increase in patience as children become older up to age 10 (Bettinger and Slonim, 2007; Angerer et al., 2015; Sutter et al., 2015; Falk et al., 2021), but not beyond age 10 (Sutter et al., 2013). However, cross-sectional patterns in the development of children's skills are not informative about sensitive periods, i.e., when returns to investments are particularly high.

Turning to prosociality, we observe a significant increase from age 7 to 11, in line with previous evidence from cross-sectional and panel data (Sutter et al., 2019; Kosse et al., 2020; List et al., 2023). This suggests that prosociality is, in principle, malleable in this age range. Moreover, 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 10.8 and 12.2 percent of a standard deviation—, while the grade 3 treatment effect is -3.2 percent of a standard deviation and not statistically significant. Likely, the latter is driven by an imbalance we observe at baseline: children in grade 3 of treatment schools have, on average, a significantly lower, pre-treatment prosociality than their counterparts in control schools. This initial difference amounts to 17.0 percent of a standard deviation. Not observing a significant difference between treatment and control group in grade 3 after the treatment implies that treated children caught up with an effect size similar to the ones we see in the three other grades. In that sense, grade-specific treatment effects are comparable in size during the whole age range between 7 and 11 years. We therefore refrain from interpreting our results as unequivocal evidence in favor of a sensitive period in the formation of prosociality during elementary school age, but summarize instead that prosociality seems to be equally malleable in the age range we consider.

In sum, our findings provide a first proof of concept that designs such as ours provide a valuable tool for learning about sensitive periods. Moreover, they suggest that sensitive periods differ across socio-emotional skills. Earlier investments in self-control and patience seem to be more effective than the same investments in these skills at later stages. In contrast, prosociality seems to be similarly malleable throughout elementary school age. For the LQ program, our results indicate that an implementation in earlier grades is more effective than in later grades, as the higher returns on self-control and patience do not come at the cost of lower prosociality returns.

5.3 Testing alternative explanations for treatment effect heterogeneity across grades

In the following, we discuss to which extent the grade-specific treatment effects on skills are driven by the differential impact of the same investment (the content of the LQ Skills for Growing program) at different stages of childhood—as opposed to other possible differences between treatment and control group children across grades. While the randomization ensures initial, pre-treatment balancedness, we cannot a priori exclude that children in different grades

have more or less time or motivation to engage with the program content, or that parents' or teachers' responses to the LQ Skill for Growing program differ across grades. We collected comprehensive, additional data on children, parents, and teachers to be able to address these possible alternative interpretations of our findings on sensitive periods.

Child involvement in the LQ Skills for Growing program. Children's grade-specific treatment effects on skills might be affected by their extent of involvement with the program. We consider three domains of involvement. First, treatment effects might be driven by how much children like the program. Younger children may be more receptive to learn through short stories and role plays than older children. We therefore asked LQ teachers to assess how much their students like the LQ lessons (see the top row of Appendix Table B6).[22] Column (1) of Table A11 shows that the average popularity of the program across all grades is rather high: 5.4 on a seven-point scale (where 7 indicates the highest popularity). Columns (2) to (5) demonstrate that there is no indication that students' differential enthusiasm about the program contributes to the grade-specific treatment effects. Second, children in higher grades may experience more time constraints and have less time to engage with the content of the LQ program. Our data allow exploring several dimensions of children's time use: how much time they spend studying for school at home, how much they help their parents at home, whether they work and how many days at school they miss due to work (see Table B10 for the full list of variables and their definitions). Table A12 demonstrates that, as one would expect, the treatment merely affects these variables. Most importantly, grade-specific heterogeneity in treatment effects on these variables is small and not significant, and thus cannot explain grade-specific treatment effects on socio-emotional skills. Finally, apart from missing days of school for work, older children may be more resilient and less prone to missing school due to illness. However, Table A12 shows that there are no grade-specific treatment effects on absence from school for health reasons.

Parental responses. The LQ intervention could, in principle, change parents' behavior towards their children because they respond to their children's changed behavior (for example, parents of less impulsive children may need to be less strict). To test this, we use a battery of six parenting style dimensions (Thönnissen et al., 2019), each comprising three survey items (see Appendix Table B11): emotional warmth, inconsistent parenting, monitoring, negative communication, psychological control, and strict control. There are no significant differences in four out of the six parenting styles between treatment and control group children, see column (2) of Table A13. Mothers in the treatment group show slightly less emotional warmth and strict control than control group mothers ($p < 0.10$). Importantly, style differences do not vary much by grade, as shown in columns (3) to (6). Only for strict control, the p-value of an F-test for equality of the coefficients in column (7) indicates that the grade-specific treatment effects are different from each other at the 10 percent level. However, the pattern of the grade-specific treatment effects on strict control does not resemble the ones of self-control and patience in section 5.2 (for treated children, strict control is significantly lower in both grade 2 and 5).

[22] We did not ask children themselves to avoid experimenter demand effects and to keep child interviews including the skills assessment completely detached from the intervention.

Teacher responses. As the LQ program is a classroom-based intervention, teachers' implementation of the program and their attitudes towards it are essential. To investigate whether differences in implementation across grades may contribute to the grade-specific heterogeneity in skill development, we asked the LQ teachers about the average duration of LQ lessons, problems with the program implementation, and how much they liked the LQ program, see Appendix Table B6. Table A11 compares average answers across grades. By regressing teachers' answers to these variables on grade indicators, we assess whether teachers' grade-specific reactions to the program exhibit a similar pattern as the grade-specific treatment effects on children's socio-emotional skills. Both the duration of the lessons and the share of teachers who report that they encountered problems when implementing the program do not differ significantly across grades. Teachers' attitudes differ across grades, in particular if we compare grade 5 to grade 2. However, we do not observe patterns which could account for stronger treatment effects in earlier grades: teachers' stronger identification with the LQ Skills for Growing program in grade 5 would rather suggest that treatment effects on socio-emotional skills in grade 5 instead of grade 2 should be more pronounced.[23]

Controlling for child involvement, parent and teacher responses. As a final check whether grade-specific responses may drive the patterns in sensitive periods, we re-estimate Table 7 including all variables that measure child involvement, parent and teacher responses that we assessed for the treatment and control group (i.e., no LQ-specific factors) as control variables. Appendix Table A15 shows that including them does not alter our main conclusions. The grade-specific treatment effects remain similar, indicating that grade-specific differences in child involvement, parent and teacher responses do not "explain away" the sensitive periods pattern we document for the formation of time preferences.

Overall, the comprehensive investigation of the role of child involvement as well as parent and teacher responses does not offer an alternative explanation for the larger treatment effects on self-control and patience in earlier grades. This reinforces the interpretation of the grade-specific heterogeneity in treatment effects as evidence of sensitive periods.

5.4 Robustness checks

We run several additional checks to test—and confirm—the robustness of our findings.

Alternative p-values. Significance levels reported so far are based on conventional critical values of t-tests. Appendix Table A16 reports p-values that are based on randomization inference or adjusted for multiple hypothesis testing.

With randomization inference, the p-values for self-control and prosociality of our overall estimates in Table 6 are still well below 0.05, see column (3) of Table A16. Next, we account for an increasing probability of false positives in the number of tested hypotheses (i.e., outcomes).

[23] We also collected information on treatment and control group teachers' days of absence, teaching style, and teacher attitudes, allowing us to gauge whether there are treatment effects on teachers. Appendix Table A14 shows the results. The treatment does not seem to operate through increased teacher supervision as the treatment effect on teacher absence from school is positive (on average, treatment group teachers miss 3.4 more days than control group teachers) and does not differ significantly across grades. While treatment group teachers do not differ in their use of visual teaching aids, they are somewhat more likely to combine textbook and real-word examples. Since the LQ program is implemented this way, this finding is reassuring.

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, with the exception of the Romano–Wolf p-value for the treatment effect on prosociality ($p = 0.058$).

Table A16 also displays alternative p-values for the grade-specific treatment effects used to assess sensitive periods. The results in Table 7 in section 5.2 are based on a joint estimation of grade-specific treatment effects through the interaction of grade indicators and the treatment indicator. Column (2) of Table A16 repeats this analysis but restricts the sample to treatment and control group children in the same grade, separately for each grade. 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 are no longer significant at conventional levels 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, with the exception of the grade 5 effect for prosociality using the Roman–Wolf adjustment ($p = 0.102$). All in all, results in Table A16 show that the overall interpretation of our results remains unchanged when using several alternative ways of inference.

Controlling for pre-treatment outcome variables, child and family characteristics. Appendix Table A17 gives the overall and grade-specific treatment effects when sequentially adding (i) pre-treatment outcomes and (ii) child and family characteristics (see Table B9) as additional control variables. Neither of these specifications qualitatively changes the estimates of our treatment effects.

Potential ceiling effects. There is no evidence for ceiling effects in higher grades that could explain the results on sensitive periods. Appendix Table A18 shows that the fraction of children who score the highest possible value on each measurement is low and, importantly, does not differ much between grades. It is thus not the case that older children are reaching the top values of our measurements and hence have no more room for improvement through the treatment.

Alternative sample restrictions. Appendix Table A19 shows the estimated treatment effects when we restrict the sample (i) to non-missing observations in both components of a given outcome index (instead of using only one component to calculate the outcome index when the other component is missing) and (ii) to observations with non-missings in all three outcome indices. Qualitatively, the results remain unchanged.

Sensitive periods along age. In our preferred specification, we estimate grade-specific treatment effects to learn about sensitive periods. Using the school grade matches the way the LQ intervention is implemented. Moreover, school grades are likely to reflect the stock of children's skills and are measured consistently across various sources in our data. While we also

elicit children's age, age is less important as a concept in Bangladesh than it is in other cultures. Therefore, the age information in our data is prone to measurement error. Nevertheless, Appendix Figure A7 shows age-specific treatment effects. They confirm the general pattern that treatment effects on time preferences are larger for younger children, albeit with decreased precision.

6 Conclusion

This study proposes a novel approach to learn about sensitive periods in the formation of children's skills: setting up a randomized controlled trial that assigns the same investment to children at different developmental stages of childhood, while holding treatment period and intensity constant. With such a design, a treatment effect that is substantially larger than those for the same skill at different developmental stages points towards the existence of a sensitive period. We believe that our approach will prove useful in future research that aims at broadening our knowledge base about sensitive periods in skill formation.

Our results offer important first insights into sensitive periods in the formation of time and social preferences, starting to fill the knowledge gap on the timing of sensitive periods in the development of children's economic preferences and socio-emotional skills more broadly. For self-control and patience, we document a sensitive period for 7- to 8-year-olds rather than older children. Our results on prosociality document its high malleability throughout elementary school age. But since malleability alone is not a sufficient condition for the existence of a sensitive period, we conclude that there is no evidence for sensitive periods in the formation of prosociality when children are between 7 and 11 years old.

Future research should bolster and extend our findings. For example, future work could build on our approach to investigate sensitive periods in economic preferences beyond the elementary school years. We cover a substantial age range that is a plausible candidate for sensitive periods as previous evidence from cross-sectional (Sutter et al., 2019; List et al., 2023, for an overview of this literature) and panel data (Kosse et al., 2020; Falk et al., 2021) documents that time and social preferences change as children grow older in the age range we consider. Still, studying children below age 7 and above age 11 could provide valuable information into possible, further sensitive periods in the formation of children's skills. Moreover, while we have no specific reason to believe that the sensitive periods we document are investment-specific, future research could aim at expanding our results to other investments or programs targeting self-control, patience, and 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. While most previous randomized controlled trials on the formation of socio-economic skills focus on early childhood (see the overviews by Kautz et al., 2014, and Durlak et al., 2011), we provide evidence from the elementary school years. 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. Furthermore, we complement the literature on childhood interventions with evidence from rural Bangladesh, a context that is culturally and economically distinct from the Western

ones studied in previous work.[24] Demonstrating the effectiveness of interventions that target children's economic preferences beyond Western, high income countries is particularly relevant as economic preferences seem to be similarly important for individual life outcomes and country-level outcomes such as economic development, entrepreneurship, charitable activities, or violent conflict around the world (Falk et al., 2018).

Second, our results imply that the LQ Skills for Growing program provides a valuable input in the process of children's skill formation. This is highly relevant as the program is already widely implemented at a large scale in dozens of countries worldwide, but, surprisingly, lacked rigorous evaluations. Since the LQ Skills for Growing program is well-established and supported by the Lions Clubs International Foundation, scaling it up even further should be relatively easy compared to setting up newly designed and often more expensive, proof-of-concept programs.

Finally, our findings on sensitive periods have important policy implications. Our results on 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 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. In contrast, the malleability of prosociality remains high throughout elementary school age. Jointly, these results suggest that an implementation of the LQ program is 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 takeaway, our findings emphasize that the same intervention may be more effective at some developmental stages of childhood 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.

[24] Only Rao (2019) exploits a natural experiment in schools in Delhi to demonstrate that being integrated with poorer students increases the prosociality of richer students.

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.

Algan, Y., E. Beasley, S. Cote, J. Park, R. E. Tremblay, and F. Vitaro (2022). The Impact of Childhood Social Skills and Self-Control Training on Economic and Non-Economic Outcomes: Evidence from a Randomized Experiment using Administrative Data. *American Economic Review 112*(8), 2553–2579.

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.

Attanasio, O., R. Bernal, M. Giannola, and M. Nores (2020). Child Development in the Early Years: Parental Investment and the Changing Dynamics of Different Domains. CSEF Working Paper 626, Centre for Studies in Economics and Finance (CSEF).

Barham, T., K. Macours, and J. A. Maluccio (2013). Boys' Cognitive Skill Formation and Physical Growth: Long-Term Experimental Evidence on Critical Ages for Early Childhood Interventions. *American Economic Review: Papers & Proceedings 103*(3), 467–471.

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.

Castillo, M., J. A. List, R. Petrie, and A. Samek (2024). Detecting Drivers of Behavior at an Early Age: Evidence from a Longitudinal Field Experiment. *Journal of Political Economy 132*(12), 3942–3977.

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 130*(9), 2361–2410.

Cobb-Clark, D. A., S. C. Dahmann, D. A. Kamhöfer, and H. Schildberg-Hörisch (2022). The Predictive Power of Self-Control for Life Outcomes. *Journal of Behavior & Organization 197*, 725–744.

Conti, G., J. J. Heckman, and R. Pinto (2016). The Effects of Two Influential Early Childhood Interventions on Health and Healthy Behaviour. *The Economic Journal 126*(596), F28–F65.

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.

Del Bono, E., J. Kinsler, and R. Pavan (2020). Skill Formation and the Trouble with Child Non-Cognitive Skill Measures. IZA Discussion Paper 13713, Institute of Labor Economics (IZA). Revision requested at Quantitative Economics.

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.

Deming, D. J. (2022). Four facts about human capital. *Journal of Economic Perspectives 36*(3), 75–102.

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., A. Becker, T. Dohmen, D. Huffman, and U. Sunde (2023). The Preference Survey Module: A Validated Instrument for Measuring Risk, Time, and Social Preferences. *Management Science 69*(4), 1935–1950.

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.

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.

Heckman, J. J., S. H. Moon, R. Pinto, P. A. Savelyev, and A. Yavitz (2010a). Analyzing Social Experiments as Implemented: A Reexamination of the Evidence from the HighScope Perry Preschool Program. *Quantitative Economics 1*(1), 1–46.

Heckman, J. J., S. H. Moon, R. Pinto, P. A. Savelyev, and A. Yavitz (2010b). 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.

Hendren, N. and B. Sprung-Keyser (2020). A unified welfare analysis of government policies. *The Quarterly Journal of Economics 135*(3), 1209–1318.

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

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.

Kaiser, T., L. Menkhoff, and L. Oberrauch (2023). Is Patience Malleable via Educational Intervention? Evidence on the Role of Age in Field Experiments. IZA Discussion Paper 16399, Institute of Labor Economics (IZA).

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, 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. Report, American Institutes for Research (AIR).

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 (PNAS) 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.

Kosse, F. and M. M. Tincani (2020). Prosociality Predicts Labor Market Success Around the World. *Nature Communications 11*(1), 1–6.

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

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

Lee, D. S. and T. Lemieux (2010). Regression Discontinuity Designs in Economics. *Journal of Economic Literature 48*(2), 281–355.

List, J. A., R. Petrie, and A. Samek (2023). How Experiments with Children Inform Economics. *Journal of Economic Literature 61*(2), 504–564.

Lührmann, M., M. Serra-Garcia, and J. Winter (2018). The Impact of Financial Education on Adolescents' Intertemporal Choices. *American Economic Journal: Economic Policy 10*(3), 309–332.

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 (PNAS) 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'Connor, T. G., M. Rutter, C. Beckett, L. Keaveney, J. M. Kreppner, English, R. A. S. Team, et al. (2000). The Effects of Global Severe Privation on Cognitive Competence: Extension and Longitudinal Follow-Up. *Child Development 71*(2), 376–390.

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.

Rutter, M. (1998). Developmental Catch-Up, and Deficit, Following Adoption After Severe Global Early Privation. *The Journal of Child Psychology and Psychiatry and Allied Disciplines 39*(4), 465–476.

Santos, I., V. Petroska-Beska, P. Carneiro, L. Eskreis-Winkler, A. Munoz Boudet, I. Berniell, C. Krekel, O. Arias, and A. Duckworth (2022). Can Grit Be Taught? Lessons from a Nationwide Field Experiment with Middle-School Students. IZA Discussion Paper 15588, Institute of Labor Economics (IZA).

Schunk, D., E. M. Berger, H. Hermes, K. Winkel, and E. Fehr (2022). Teaching Self-Regulation. *Nature Human Behaviour 6*(12), 1680–1690.

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 (2025). The Causal Impact of Socio-Emotional Skills Training on Educational Success. *The Review of Economic Studies 92*(1), 506–552.

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., M. Weyland, A. Untertrifaller, M. Froitzheim, and S. O. Schneider (2023). Financial Literacy, Experimental Preference Measures and Field Behavior – A Randomized Educational Intervention. IZA Discussion Paper 16102, Institute of Labor Economics (IZA).

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.

Thönnissen, C., B. Wilhelm, P. Alt, H. Greischel, and S. Walper (2019). Scales and Instruments Manual, Waves 1 to 10; Release 10.0. Technical Report, Panel Analysis of Intimate Relationships and Family Dynamics.

TIME magazine (2022). Tens of Thousands of Boys in Bangladesh Were Forced into Work During the Pandemic. Now School Is Resuming Without Them. `https://time.com/6170432/bangladesh-child-labor-pandemic/`. TIME article by Corinne Redfern and Ali Ashan from April 26, 2022; last accessed June 8, 2025.

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

van den Berg, G. J., P. Lundborg, P. Nystedt, and D.-O. Rooth (2014). Critical Periods During Childhood and Adolescence. *Journal of the European Economic Association 12*(6), 1521–1557.

Van IJzendoorn, M. H., M. J. Bakermans-Kranenburg, and F. Juffer (2007). Plasticity of Growth in Height, Weight, and Head Circumference: Meta-Analytic Evidence of Massive Catch-Up After International Adoption. *Journal of Developmental & Behavioral Pediatrics 28*(4), 334–343.

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.

A Additional information

A.1 Alternative definition: ceteris paribus versus composite sensitive periods

In addition to defining sensitive periods on the same-period outcome as in Eq. (1), Cunha et al. (2006, p. 804) provide the following alternative definition of sensitive periods. Formally, t^* is a sensitive period for $S_{t,j}$ if for $k \neq t^*$ and $k, t^* \leq t$,

$$\left.\frac{\partial S_{t,j}}{\partial I_k}\right|_{S_0=s_0, I_k=i_k, k=1,\ldots,t, k\neq t^*} \leq \left.\frac{\partial S_{t,j}}{\partial I_{t^*}}\right|_{S_0=s_0, I_k=i_k, k=1,\ldots,t}, \tag{4}$$

where $S_{t,j}$ fixes the period in which skill j is measured. The inequality in Eq. (4) is strict for at least one period.

Eq. (1) and (4) differ in the point in time when skill j is assessed. Eq. (1) compares the effect of the same investment in two periods, say, the effect of a given investment in grade 2 on students' skills at the end of grade 2 with the effect of the same investment in grade 3 on students' skills at the end of grade 3 (when using grades to proxy stages of childhood). Eq. (4) compares the effect of an investment in grade 2 with the effect of the same investment in grade 3 on students' skills at same later stage, say, the end of grade 6.

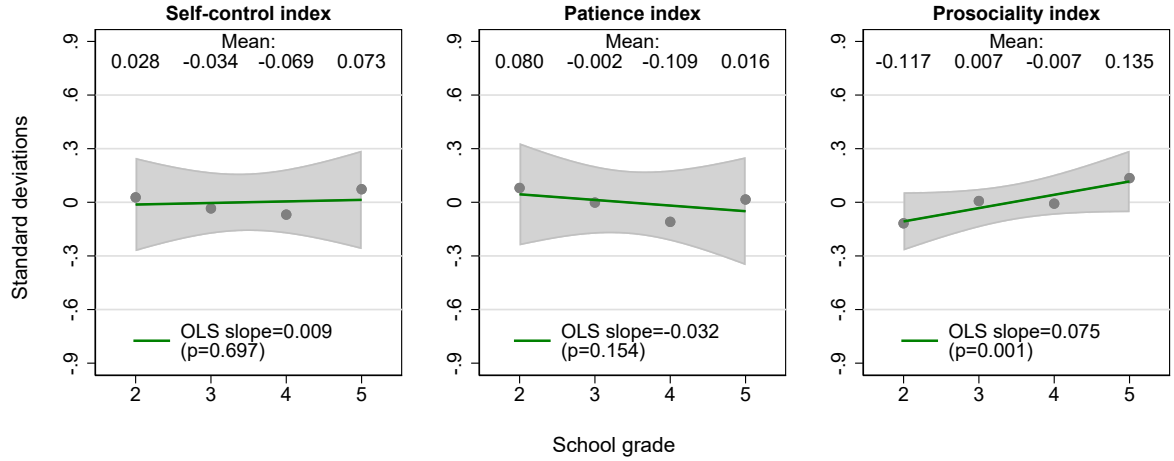
While this may seem like a technicality at first sight, the interpretation of sensitive periods derived from estimates of Eq. (1) and (4) differs. An important feature of the technology of skill formation is to take multiplier effects into account, i.e., mechanisms through which skills beget skills. Since skill formation is subject to self-productivity and complementarity of skills over time, sensitive periods in Eq. (4) reflect this process on top of the immediate return to investment. They compound the initial investment effect with possible multiplier effects that evolved over time. Therefore, we refer to sensitive periods as defined by Eq. (4) as *composite sensitive periods*. In contrast, Eq. (1) defines *ceteris paribus sensitive periods* that reflect the immediate effect of the investment. Our terminology is inspired by Todd and Wolpin (2003) who distinguish ceteris paribus treatment effects that abstract from compensating or reinforcing reactions from policy treatment effects that do not. Previous studies estimated exclusively composite sensitive periods (most likely due to data availability), whereas our research design allows for the estimation of ceteris paribus sensitive periods. We therefore make this distinction that was introduced by Cunha et al. (2006) and Cunha and Heckman (2007) explicit.

Both ceteris paribus and composite sensitive periods are of obvious policy relevance and we consider them complementary. Since we rely on an RCT design with an exogenously assigned investment, causal inference on investment returns is straightforward. By assessing skills immediately after the investment, we directly estimate the marginal effects in Eq. (1). By contrast, estimating composite sensitive periods requires imposing assumptions on the functional form of the process of skill formation, including its dynamics over time that depend on initial skill levels, possible responses in other investments to the investment under study, and potential further shocks after the initial investment that may affect different subsamples differently. Our approach of inferring ceteris paribus sensitive periods circumvents this challenge by evaluating

skills shortly after the investment. Moreover, previous studies have estimated composite sensitive periods based on investments that are not restricted to one period but sustained until measuring outcomes at a fixed age in the future. In such an environment, exploiting differences in initial age at exposure compounds age at exposure with length of exposure. Disentangling length of exposure (the per-year effect in non-sensitive "normal" periods) and age at exposure (a higher per-year effect in sensitive periods) requires assuming a functional form for the per-year effect and its dynamics over time and interpreting deviations from this functional form as evidence for sensitive periods.

A.2 Additional figures and tables

FIGURE A1: Trends in means of outcomes indices across school grades

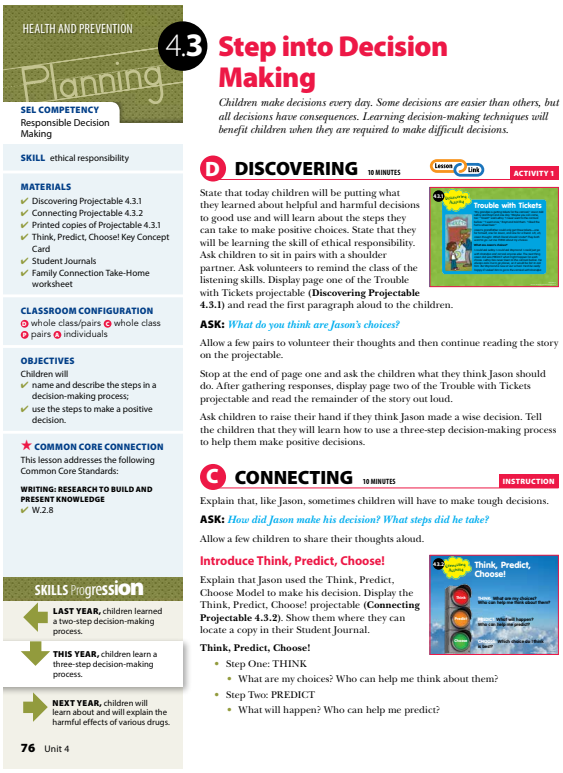


NOTES: Own illustration based on the control group of the estimation sample. The markers represent the mean values of the outcome variables across school grades. The numerical values are reported at the top of each panel. The gray area indicates the 95 percent confidence interval, centered around the mean. We standardized the data across all grades (mean 0, standard deviation 1). The green line gives the slope of a linear regression of the outcome variable on a scalar for the school grade. The coefficient is reported at the bottom of the plot, with its p-value in parentheses.

FIGURE A2: Example of Lions Quest learning materials

(A) Instruction materials

(B) Teachers' resource guide



HEALTH AND PREVENTION

Planning

4.3 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?*

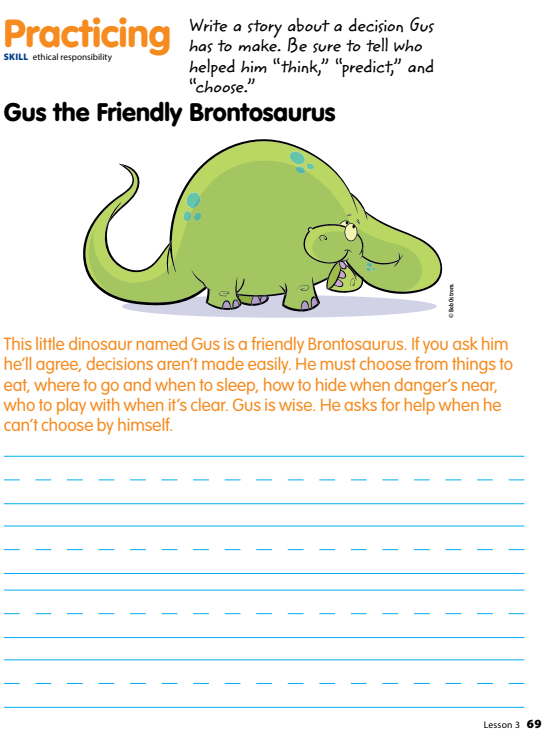
Allow a few children to share their thoughts aloud.

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."

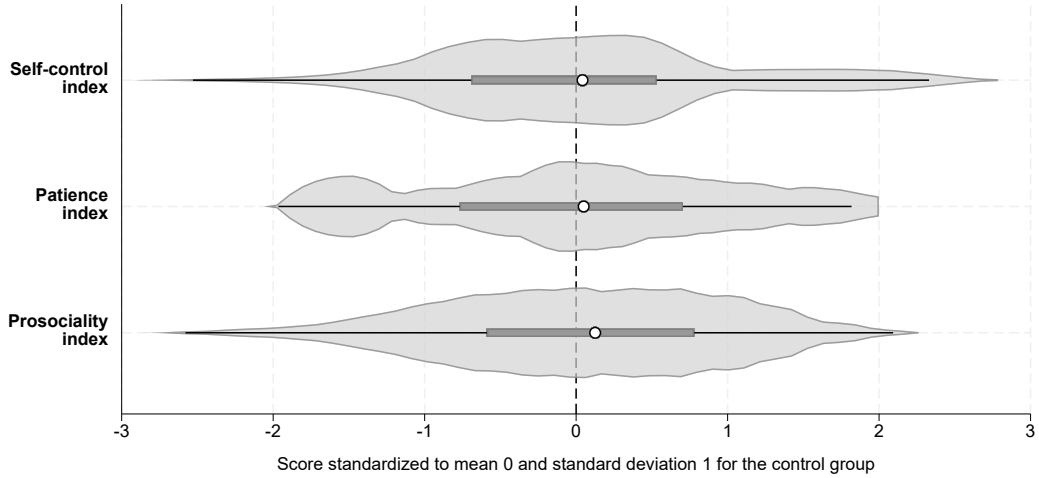
Gus the Friendly Brontosaurus

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**

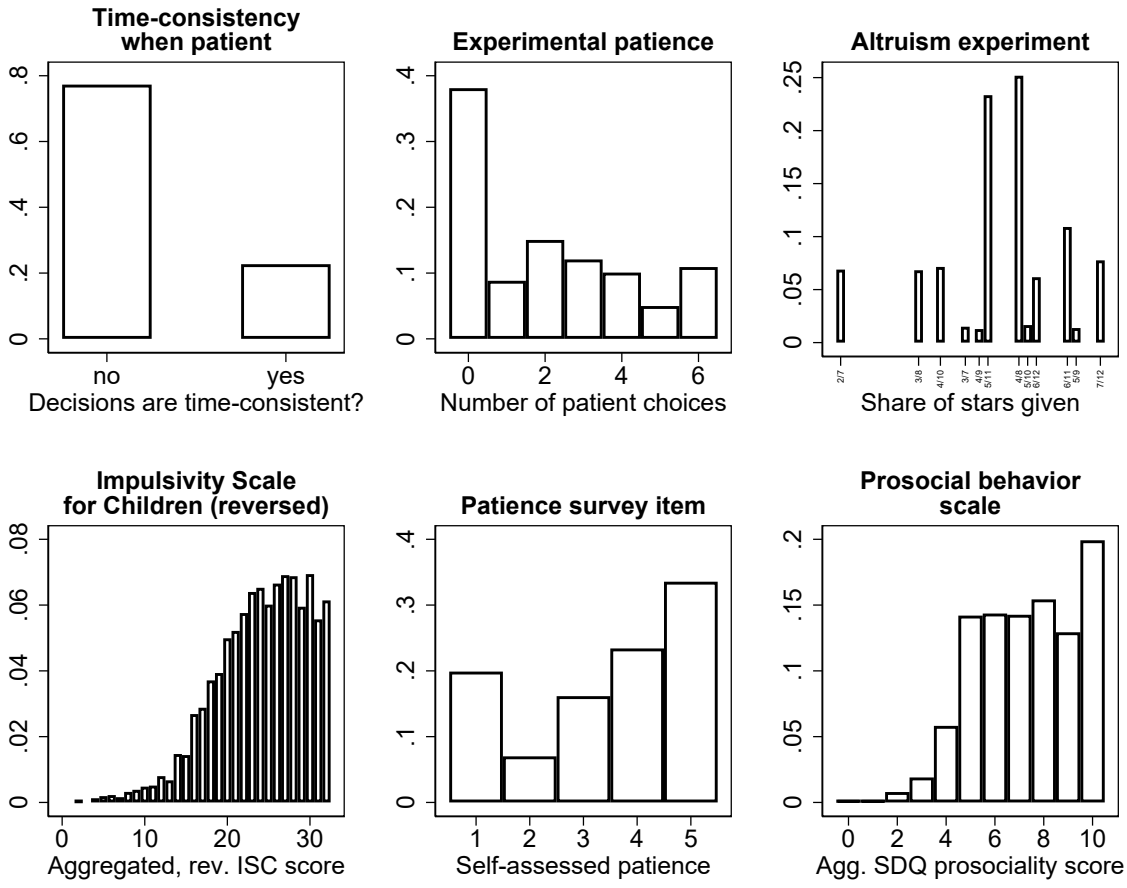
NOTES: LQ Skills for Growing resources, grade 2, unit 4. English translation. Provided by the Lions Clubs International Foundation.

FIGURE A3: Distribution of outcome measures



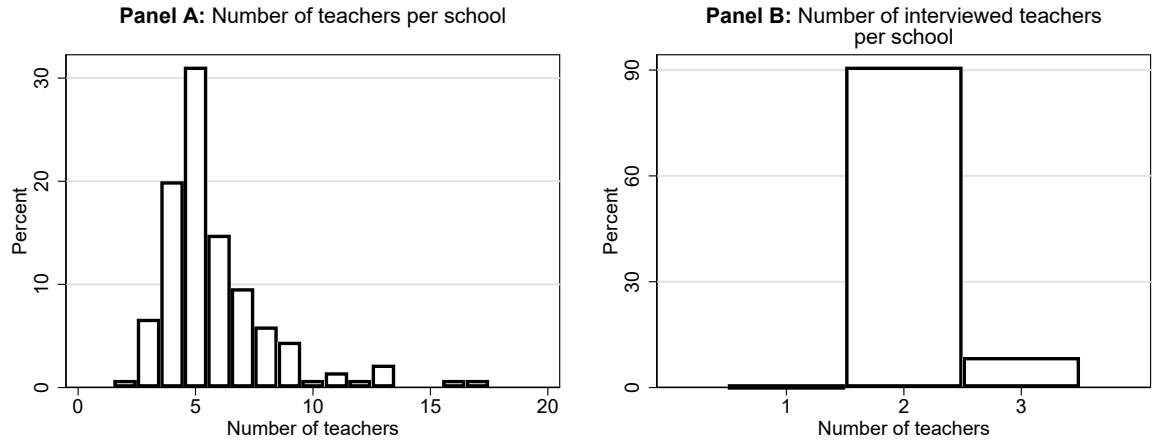
NOTES: Own calculations based on the estimation sample. All measures are standardized to have control group mean 0 and standard deviation 1. Exact aggregation procedures are described in section 4.2. Higher values indicate higher degrees of self-control, patience, or prosociality, respectively. The violin plot displays the distribution and density of the data, with the width of each violin representing the kernel density estimate, while the central bar marks the interquartile range, and the white dot indicates the median.

FIGURE A4: Distribution of outcome components



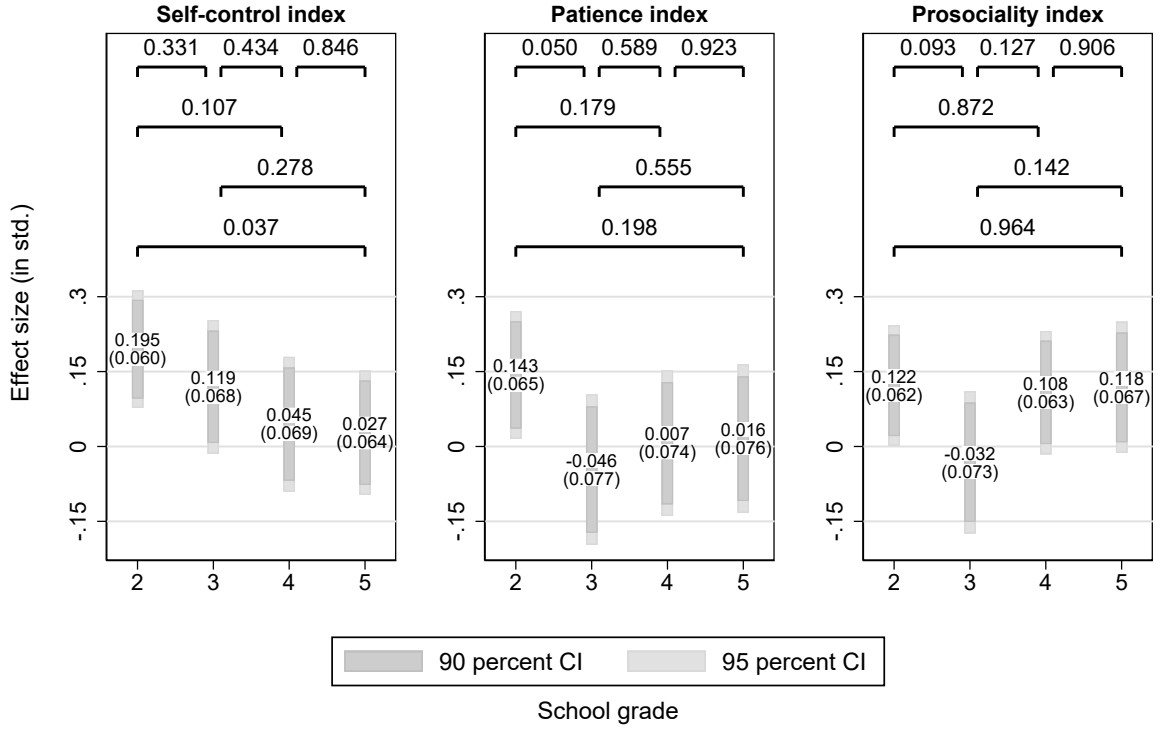
NOTES: Own calculations based on the estimation sample. Higher values indicate higher levels of the plotted variable.

FIGURE A5: Descriptive statistics on teachers



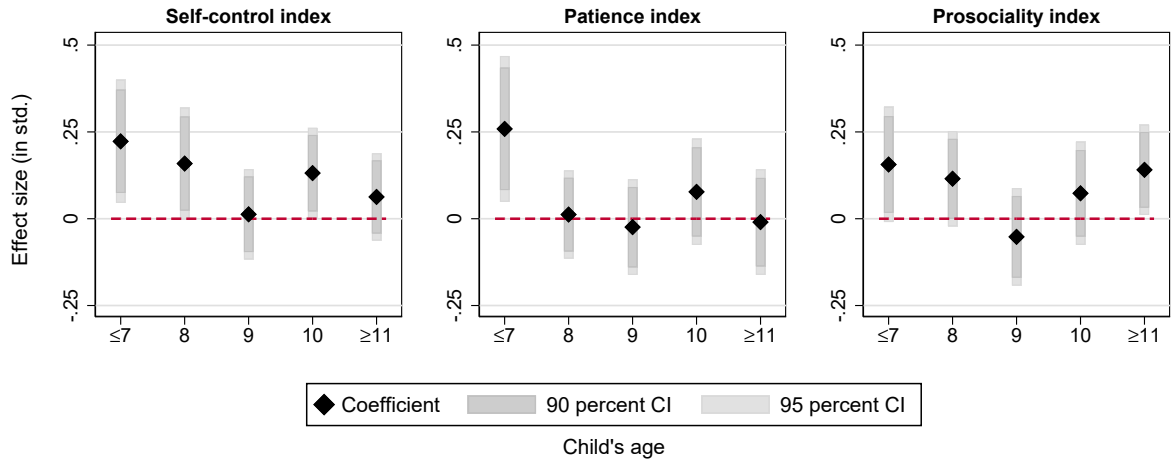
NOTES: Own calculations. **Panel A** plots the number of teachers per school as reported in the school questionnaire, based on all 135 schools. **Panel B** plots the number of interviewed teachers per school, 277 teachers in total.

FIGURE A6: Sensitive periods, stating coefficient sizes and p-values for the difference between treatment–grade interaction terms



NOTES: Own calculations based on the estimation sample. This figure repeats Figure 3 (that displays the results reported in Table 7) and additionally states the p-values for the difference between the treatment–grade interaction terms calculated by a Wald test.

FIGURE A7: Sensitive periods by age (in years)



NOTES: Own calculations based on the estimation sample. The figure plots age-specific treatment effects, similar to the grade-specific treatment effects plotted in Figure 3. Age-specific treatment effects are estimated as in Equation (3), but use children's age instead of grade interactions with the treatment indicator.

TABLE A1: Teacher characteristics by treatment status

Teacher characteristic	(1) Treatment group mean	(2) Control group mean	(3) p-value treatment effect
Female	0.563	0.704	0.002
Age	40.823	38.378	0.002
Working experience	15.071	13.311	0.045
Completed teacher training	0.894	0.867	0.175
Reason: love of job	0.810	0.696	0.188
Lives in village of school	0.289	0.304	0.646
Observations	142	135	

NOTES: Own calculations based on teacher data, 277 observations, 276 observations due to missing information for age and working experience. Columns (1) and (2) report the mean of teacher characteristics for the treatment and the control group, respectively. Column (3) reports the p-value of the treatment effect when the teacher characteristic on the left is regressed on the treatment indicator and strata fixed effects, with standard errors clustered at the school level. Teacher characteristics are defined in Table B4.

TABLE A2: Relationship between teacher characteristics and student outcomes

	(1)	(2)	(3)
	Dependent variable		
Teacher characteristic	Self-control index	Patience index	Prosociality index
Female	−0.010	0.111*	−0.004
	(0.056)	(0.061)	(0.058)
Age	0.007	0.014**	−0.004
	(0.007)	(0.006)	(0.007)
Working experience	−0.002	−0.009	0.001
	(0.006)	(0.006)	(0.006)
Completed teacher training	−0.038	0.061	−0.013
	(0.088)	(0.097)	(0.094)
Reason: love of job	0.125	0.005	0.094
	(0.079)	(0.082)	(0.080)
Lives in village of school	−0.086*	−0.077	−0.036
	(0.052)	(0.063)	(0.057)
Observations	3,184	3,135	3,184

NOTES: Own calculations based on the estimation sample augmented with the teacher data. We have slightly fewer observations compared to Table 6 due to school–grade combinations that were not taught by an interviewed teacher. Each column reports the output of one regression, where the variable in the header (self-control, patience, and prosociality, respectively) is regressed on teacher characteristics and strata fixed effects. Teacher characteristics are defined in Table B4. Standard errors are clustered at school–grade level. Significance at $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

TABLE A3: Using alternative experimental outcome measures for estimating overall treatment effects

	(1)	(2)
	Treatment effect on the	
Dependent variable	experimental measure	corresponding index
Panel A: time-consistency		
Time-consistency (std., main specification)	0.032	0.101***
	(0.042)	(0.038)
Time-consistency, not std., binary	0.013	
	(0.017)	
Time-consistency, not conditional on patience (std.)	0.000	0.079**
	(0.031)	(0.035)
Time-consistency, sample incl. failed control questions (std.)	0.030	0.099***
	(0.041)	(0.038)
Panel B: patience		
Experimental patience (std., main specification)	0.010	0.033
	(0.043)	(0.042)
Exp. patience, indicator for at least 1 patient decision (std.)	0.006	0.031
	(0.041)	(0.041)
Experimental patience, sample incl. failed control questions (std.)	0.007	0.032
	(0.043)	(0.042)
Panel C: altruism		
Altruism experiment (std., main specification)	0.094**	0.083**
	(0.038)	(0.036)
Altruism experiment, avg. share (std.)	0.070*	0.065*
	(0.038)	(0.037)
Altruism experiment, number of altruistic choices (std.)	0.085**	0.077**
	(0.037)	(0.036)
Altruism experiment, sample incl. failed control questions (std.)	0.094**	0.082**
	(0.038)	(0.036)

NOTES: Own calculations based on the estimation sample. Column (1) displays the treatment effects for alternative definitions of the experimental outcome components. Column (2) reports treatment effects on the indices when using the experimental components as stated on the left of the table. The first row of each panel displays our preferred specification that is also presented in Tables 6 and A4. **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 some degree of patience: Fourth row: same as preferred specification but using the full sample of children (3,213 observations in column (1), 3,258 observations in column (2)), including those who fail to answer the control questions of the time preferences experiment correctly. **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, including those who fail to answer the control questions of the time preferences experiment correctly (3,213 observations in the columns (1) and (2)). **Panel C.** Second row: share of stars given to the other child averaged over the four games. Third row: counting the number of games with altruistic decisions. Fourth row: same as preferred specification but using the full sample of children, including those who fail to answer the control questions of the social preferences experiment correctly (3,213 observations in the column (1), 3,262 observations in column (2)). If measures are standardized (indicated in table), 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 at $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

TABLE A4: Treatment effects on socio-emotional skill components

Dependent variable	(1) Number of observations	(2) Treatment effect
Self-control		
Experiment: Time-consistency	3,207	0.032
		(0.042)
Survey scale: Impulsivity Scale for Children (rev.)	3,126	0.109***
		(0.042)
Patience		
Experiment: Patience	3,207	0.010
		(0.043)
Survey scale: Patience survey item	3,213	0.046
		(0.040)
Prosociality		
Experiment: Altruism in dictator games	3,209	0.094**
		(0.038)
Survey scale: Prosocial behavior scale	3,162	0.018
		(0.042)

NOTES: Own calculations based on the estimation sample. Outcome variables are defined as described in the text. The number of observations for each outcome is given in column (1); fewer observations compared to Table 1 are due to missing information. The estimation result stated in column (2) follows model (2), that is, we regress the outcome variable on the treatment indicator and a full set of strata fixed effects. All dependent variables are standardized such that the mean of the control group is 0 and the standard deviation is 1. School-clustered standard errors in parentheses. Significance at $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

TABLE A5: Correlation of children's socio-emotional skills with parents' skills

	(1)	(2)	(3)
	Child's socio-emotional skills (in std.)		
Family characteristic ↓ (in std.)	Self-control index	Patience index	Prosociality index
Mother's self-control index	0.101***		
	(0.025)		
N	1,579		
Mother's patience index		0.276***	
		(0.024)	
N		1,553	
Mother's prosociality index			0.070***
			(0.025)
N			1,582
Father's self-control index	0.018		
	(0.028)		
N	1,247		
Father's patience index		0.214***	
		(0.028)	
N		1,231	
Father's prosociality index			0.158***
			(0.028)
N			1,250

NOTES: Own calculations based on the control group observations of the estimation sample. The table shows the Pearson correlation coefficients between children's socio-emotional skills in the columns and parental skills on the left of the table. We obtain the correlation coefficients by regressing children's skills on the single regressor on the left without any further control variables and without an intercept. We standardize all variables to mean 0 and standard deviation 1 in each sample (the number of observations is given by N), such that the regression coefficients equal the Pearson correlation coefficients. Standard errors from the regression in parentheses. Significance at $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

TABLE A6: Balancing results for family and child covariates (measured in 2019 at the beginning of the treatment period)

VARIABLE	(1) Number of observations	(2) Uncond. mean	(3) Treatment difference	(4) p-value of difference
Female	3,263	0.514	−0.016	0.260
Age	3,263	9.479	0.063	0.265
Number of siblings	3,213	2.345	0.035	0.561
Muslim	3,152	0.799	0.057	0.177
Grade 2	3,263	0.275	0.005	0.663
Grade 3	3,263	0.246	-0.021^{*}	0.071
Grade 4	3,263	0.237	0.001	0.893
Grade 5	3,263	0.242	0.015^{+}	0.142
HH income (in 1,000 Taka)	3,263	229.446	−38.760	0.186
Dwelling: tin wall	3,263	0.790	−0.013	0.493
Dwelling: tin roof	3,263	0.955	0.002	0.801
Dwelling: mud floor	3,263	0.849	−0.023	0.269
Dwelling: area	3,250	9.734	−0.173	0.718
Dwelling: electricity	3,263	0.900	−0.013	0.590
M: age	3,237	35.073	−0.028	0.890
M: literate	3,235	0.720	0.032^{+}	0.108
M: primary educ. or less	2,331	0.233	0.025	0.156
M: secondary educ.	2,331	0.731	−0.021	0.228
M: post-sec. educ.	2,331	0.036	−0.005	0.538
M: high-status job	3,243	0.009	−0.004	0.159
M: routine job	3,243	0.007	−0.002	0.411
M: agricultural work	3,243	0.789	0.023	0.300
M: not working	3,243	0.196	−0.017	0.425
M: missed work due to illness	3,235	0.323	0.027	0.175
M: no interview	3,263	0.006	0.002	0.538
F: age	3,102	42.549	0.144	0.596
F: literate	3,089	0.594	−0.007	0.755
F: primary educ. or less	1,836	0.279	0.013	0.559
F: secondary educ.	1,836	0.645	−0.007	0.754
F: post-sec. educ.	1,836	0.075	−0.005	0.710
F: high-status job	3,152	0.131	0.013	0.361
F: routine job	3,152	0.457	−0.028	0.294
F: agricultural work	3,152	0.380	0.019	0.460
F: not working	3,152	0.032	−0.003	0.623
F: missed work due to illness	3,089	0.381	0.052^{**}	0.014
F: no interview	3,263	0.034	−0.001	0.922

NOTES: Own calculations based on the estimation sample. This table shows the balancing results for family and child covariates, summarized in Table 5. See Table B9 for variable definitions. M = mother, F = father, HH = household. Column (1) gives the number of non-missing observations. Column (2) reports the unconditional mean. Column (3) states the treatment effect when the variable on the left is regressed on a treatment indicator and strata fixed effects, and column (4) gives the corresponding p-value of the coefficient. School-clustered standard errors. Significance level of the treatment effects: $^{+}p < 0.15$, $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

TABLE A7: Balancing results for school-level covariates

VARIABLE	(1) Number of observations	(2) Uncond. mean	(3) Treatment difference	(4) p-value of difference
Number teachers	135	5.867	0.064	0.877
Female teachers	135	3.911	−0.173	0.644
Number class rooms	135	4.393	0.383	0.238
Separate rooms	135	0.607	0.127^{*}	0.068
Sanitary	135	0.422	0.030	0.738
Furniture	135	0.852	0.070	0.217
Overhead projector	135	0.511	−0.006	0.946

NOTES: Own calculations based on the school data. This table shows the balancing results for school-level covariates, summarized in Table 5. See Table B3 for variable definitions. Column (1) gives the number of non-missing observations. Column (2) reports the unconditional mean. Column (3) states the treatment effect when the variable on the left is regressed on a treatment indicator and strata fixed effects, and column (4) gives the corresponding p-value of the coefficient. Significance level of the treatment effects: $^{+}p < 0.15$, $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

TABLE A8: Sample attrition

	(1)	(2)	(3)
	Dependent variable: attrition indicator		
	Attrition w.r.t. treatment	Attrition w.r.t. treatment and strata fixed effects	Attrition w.r.t. treatment and strata fixed effects and pre-treatment outcomes
Treatment indicator	0.000	0.006	0.005
	(0.016)	(0.014)	(0.014)
Self-control index, pre-treatment			0.001
			(0.009)
– interaction with treatment			−0.007
			(0.014)
Patience index, pre-treatment			−0.008
			(0.010)
– interaction with treatment			0.001
			(0.014)
Prosociality index, pre-treatment			−0.004
			(0.010)
– interaction with treatment			0.012
			(0.014)
Constant	0.145***	0.251***	0.240***
	(0.012)	(0.067)	(0.063)
N	2,789	2,789	2,737

NOTES: Own calculations based on all children sampled in 2018. The outcome variable in all three specifications is attrition. The attrition indicator equals 1 if a child attending a sample elementary school in grades 1 to 4 and being interviewed in 2018 does not attend a sample elementary school in grades 2 to 5 and/or is not interviewed in 2019, 0 else. In column (1), the attrition indicator is regressed on the treatment indicator. In column (2), we additionally include strata fixed effects. The specification in column (3) regresses the attrition indicator on the treatment indicator, pre-treatment socio-emotional skill outcomes (assessed in 2018), treatment–skill interaction terms, as well as strata fixed effects. Pre-treatment skill measures are standardized such that the mean of the control group is 0 and the standard deviation is 1. The number of observations (N) is smaller in column (3) due to item non-response in the 2018 skill measures. Standard errors in parentheses, clustered at school level. Significance at $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. For column (3), the p-value of an F-test of joint significance of the treatment indicator, pre-treatment outcomes, and interactions is 0.94. For F-tests of joint significance of skills and their respective interactions with the treatment, p-values are 0.88 for self-control, 0.59 for patience, and 0.67 for prosociality.

TABLE A9: Treatment effect heterogeneity by socio-demographic characteristics

	(1)	(2)	(3)	(4)
		Treatment effect on		
	Mean	Self-control index	Patience index	Prosociality index
Panel A: treatment effect by child's gender				
Male	0.514	0.136***	0.044	0.096*
		(0.052)	(0.050)	(0.049)
N		1,583	1,557	1,586
Female		0.072	0.025	0.070*
		(0.050)	(0.053)	(0.039)
N		1,675	1,656	1,676
Equality (p-value)		0.266	0.481	0.968
Panel B: treatment effect by mother's literacy				
Illiterate	0.720	0.150**	0.099	0.056
		(0.068)	(0.062)	(0.067)
N		903	894	905
Literate		0.076*	−0.014	0.092**
		(0.040)	(0.044)	(0.040)
N		2,327	2,291	2,329
Equality (p-value)		0.401	0.114	0.673
Panel C: treatment effect by father's literacy				
Illiterate	0.594	0.149***	0.077	0.110**
		(0.053)	(0.062)	(0.051)
N		1,251	1,237	1,253
Literate		0.090*	0.014	0.087**
		(0.051)	(0.048)	(0.044)
N		1,833	1,810	1,835
Equality (p-value)		0.229	0.258	0.236
Panel D: treatment effect by family income				
Below median	0.500	0.124**	0.060	0.072
		(0.051)	(0.053)	(0.047)
N		1,629	1,606	1,631
Above median		0.076	0.012	0.089*
		(0.047)	(0.051)	(0.049)
N		1,629	1,607	1,631
Equality (p-value)		0.331	0.449	0.677

NOTES: Own calculations based on the estimation sample. The table shows the treatment effect estimated separately for binary socio-demographic characteristics stated in the panel headings. Column (1) states the share of individuals in the estimation sample for whom the characteristic is 1. Specifications in columns (2) to (4) are similar to Table 6, but estimations were run separately for each value of the indicator (e.g., girls and boys). N indicates the number of observations that enter the regressions. The bottom row of each panel reports the p-value of a Wald test for the equality of the coefficients. Standard errors are clustered at school level. Significance at $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

TABLE A10: Treatment effect heterogeneity by children's pre-treatment socio-emotional skills and parents' socio-emotional skills

	(1)	(2)	(3)
	Treatment effect on		
	Self-control index	Patience index	Prosociality index
Panel A: treatment effect by children's pre-treatment skills			
Below median	0.125**	0.016	0.071
	(0.060)	(0.064)	(0.056)
N	1,176	1,140	1,197
Above median	0.055	0.029	0.028
	(0.051)	(0.063)	(0.046)
N	1,325	1,304	1,316
Equality (*p*-value)	0.427	0.582	0.470
Panel B: treatment effect by children's pre-treatment IQ			
Below median	0.046	0.084	0.055
	(0.066)	(0.067)	(0.051)
N	1,094	1,103	1,145
Above median	0.051	−0.030	0.039
	(0.054)	(0.054)	(0.056)
N	1,386	1,341	1,338
Equality (*p*-value)	0.845	0.183	0.605
Panel C: treatment effect by mother's skills			
Below median	0.095*	0.032	0.039
	(0.053)	(0.062)	(0.044)
N	1,586	1,469	1,760
Above median	0.081	0.024	0.105*
	(0.051)	(0.045)	(0.054)
N	1,600	1,672	1,430
Equality (*p*-value)	0.790	0.951	0.080
Panel D: treatment effect by father's skills			
Below median	0.011	−0.007	0.096**
	(0.059)	(0.062)	(0.048)
N	1,173	1,409	1,131
Above median	0.173***	0.102*	0.078
	(0.053)	(0.055)	(0.049)
N	1,331	1,065	1,376
Equality (*p*-value)	0.072	0.042	0.743

NOTES: Own calculations based on the estimation sample. The table shows the treatment effect estimated separately for binary indicators whether the child's pre-treatment / his or her parent's skill stated in the panel heading is above the median. Specifications are similar to Table 6, but estimations were run separately for above- and below-median skills. *N* indicates the number of observations that enter the regressions. The bottom row of each panel reports the *p*-value of a Wald test for the equality of the coefficients. Standard errors are clustered at school level. Significance at $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

TABLE A11: Teachers' assessment of LQ-specific aspects by grade

	(1)	(2)	(3)	(4)	(5)
		Grade effect (relative to grade 2)			
Dependent variable	Mean	Grade 3	Grade 4	Grade 5	F-test p-value
Popularity of LQ lessons	5.411	−0.206	−0.030	−0.058	0.635
		(0.163)	(0.111)	(0.133)	
Average duration of a LQ lesson	31.888	−0.722	−0.064	−0.796	0.372
		(0.647)	(0.516)	(0.570)	
Problems with implementation	0.573	0.088**	0.051	0.048	0.449
		(0.038)	(0.036)	(0.038)	
Identification with program	5.303	−0.076	0.095	0.246**	0.001
		(0.090)	(0.103)	(0.116)	

NOTES: Own calculations based on the teacher data, 241 observations at teacher–grade level. This table compares teacher-answered LQ-specific questions (see Table B6) by grade. All questions refer to all LQ classes a teacher taught. Because the grades in which a teacher taught LQ classes vary across teachers, we observe variation across grades. Column (1) gives the overall mean. To compare aspects by grade, we regress the aspect stated on the left on grade indicators in columns (2) to (4) (grade 2 is the reference category) and strata fixed effects. Column (5) gives the p-value of an F-test of equality of the grade indicator coefficients reported in columns (2) to (4). Standard errors (in parentheses) are clustered at school level. Clustering standard errors on the teacher level (as this tables treats teacher–grade observations as independent albeit questions refer to all LQ classes) does not change the patterns of significant effects indicated by school-clustered standard errors. Significance at $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

TABLE A12: Overall and grade-specific treatment effects on children's behaviors

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
			Grade-specific treatment effects				F-test
Dependent variable	Mean	Overall TE	Grade 2	Grade 3	Grade 4	Grade 5	p-value
Studying: 2–3 hrs or more a day	0.500	0.018	0.038	0.009	−0.003	0.007	0.832
		(0.019)	(0.032)	(0.041)	(0.036)	(0.036)	
Working	0.279	−0.021	−0.028	−0.017	−0.042	−0.008	0.900
		(0.021)	(0.030)	(0.034)	(0.040)	(0.036)	
Helping in the household	0.792	−0.003	−0.015	0.014	−0.026	0.006	0.718
		(0.017)	(0.034)	(0.031)	(0.026)	(0.026)	
Missed school: due to work	0.180	−0.001	−0.136	−0.065	0.119	0.126	0.579
		(0.081)	(0.147)	(0.096)	(0.134)	(0.266)	
Missed school: due to illness	1.255	−0.200*	−0.151	−0.322*	−0.062	−0.287	0.610
		(0.103)	(0.163)	(0.170)	(0.173)	(0.238)	

NOTES: Own calculations based on the estimation sample, number of observations: 3,069 (hours spent studying), 3,114 (working, helping in household), 2,751 (days missed school due to work), 2,848 (days missed school due to illness). The table reports the treatment effects on child behaviors as outcome variables. Definitions of the child behavior variables are reported in Table B10. Column (1) states the mean value of the outcome variable across all children. Column (2) displays the overall treatment effect on child behaviors, columns (3) to (6) the grade-specific treatment effects. The first three dependent variables are binary indicator variables. To estimate the grade-specific treatment effects on child behaviors, we use the same specification as for sensitive periods, with the outcome variables stated on the left of this table. Column (7) gives the p-value of an F-test on equality of the grade-specific treatment effects. All regressions include strata fixed effects. Standard errors (in parentheses) are clustered at school level. Significance at $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

TABLE A13: Overall and grade-specific treatment effects on parenting styles

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
			Grade-specific treatment effects				F-test
Dependent variable	Mean	Overall TE	Grade 2	Grade 3	Grade 4	Grade 5	p-value
Emotional warmth	3.403	−0.060*	−0.080	−0.053	−0.029	−0.073	0.854
		(0.031)	(0.050)	(0.049)	(0.049)	(0.048)	
Inconsistent parenting	2.595	−0.046	−0.027	−0.081*	−0.032	−0.046	0.840
		(0.031)	(0.053)	(0.045)	(0.048)	(0.048)	
Monitoring	2.989	−0.044	−0.088**	−0.031	0.001	−0.050	0.393
		(0.030)	(0.044)	(0.048)	(0.042)	(0.046)	
Negative communication	2.584	−0.014	−0.038	0.023	−0.032	−0.005	0.784
		(0.035)	(0.055)	(0.054)	(0.047)	(0.049)	
Psychological control	1.904	0.010	0.003	−0.001	0.008	0.030	0.950
		(0.026)	(0.041)	(0.041)	(0.043)	(0.044)	
Strict control	2.653	−0.061*	−0.107**	−0.045	0.034	−0.112**	0.091
		(0.033)	(0.053)	(0.056)	(0.052)	(0.056)	

NOTES: Own calculations based on the estimation sample, 3,196 observations. The table reports the treatment effects on parenting style indicators as outcome variables. Definitions of parenting styles are reported in Table B11. Column (1) states the mean value of the outcome variable across all children. Column (2) displays the overall treatment effect on each parenting style, columns (3) to (6) the grade-specific treatment effects. To estimate the grade-specific treatment effects on parenting style, we use the same specification as employed for sensitive periods, except that the outcome variables are the variables stated on the left of this table. Column (7) gives the p-value of an F-test on equality of the grade-specific treatment effects. All regressions include strata fixed effects. Standard errors (in parentheses) are clustered at school level. Significance at $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

TABLE A14: Overall and grade-specific treatment effects on teaching style and teacher attitudes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
			Grade-specific treatment effects				F-test
Dependent variable	Mean	Overall TE	Grade 2	Grade 3	Grade 4	Grade 5	p-value
Multiple teaching concepts	0.374	0.092	0.038	0.100*	0.115**	0.103*	0.239
		(0.056)	(0.065)	(0.058)	(0.057)	(0.059)	
Visual aids in teaching	0.894	0.006	0.006	0.006	0.007	0.007	0.826
		(0.029)	(0.037)	(0.027)	(0.027)	(0.028)	
Days of absence	13.343	3.442*	2.471	3.717**	3.140*	4.043**	0.204
		(1.952)	(2.565)	(1.883)	(1.838)	(1.920)	
Enjoy being a teacher	0.538	0.036	0.005	0.062	0.035	0.032	0.182
		(0.053)	(0.059)	(0.055)	(0.054)	(0.055)	

NOTES: Own calculations based on the teacher data, 959 observations at teacher–grade level. The table reports the treatment effects on teaching style and teacher attitudes as outcome variables (see Table B5). Column (1) states the mean value of the outcome variable across all children. Column (2) displays the overall treatment effect on teaching style and teacher attitudes, columns (3) to (6) the grade-specific treatment effects. To estimate the grade-specific treatment effects on teaching style and teacher attitudes, we use the same specification as employed for sensitive periods, except that the outcome variables are the variables stated on the left of this table. Column (7) gives the p-value of an F-test on equality of the grade-specific treatment effects. All regressions include strata fixed effects. Standard errors (in parentheses) are clustered at school level. Significance at $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

TABLE A15: Treatment effects on socio-emotional skills (overall and by grade) when controlling for child involvement, parent and teacher responses

	(1)	(2)	(3)	(4)	(5)
		Grade-specific treatment effects			
Dependent variable	Overall TE	Grade 2	Grade 3	Grade 4	Grade 5
Self-control index	0.113***	0.207***	0.134**	0.068	0.024
	(0.035)	(0.060)	(0.066)	(0.065)	(0.064)
Patience index	0.058	0.188***	−0.027	0.021	0.021
	(0.040)	(0.066)	(0.075)	(0.073)	(0.074)
Prosociality index	0.070**	0.113*	−0.033	0.087	0.116*
	(0.033)	(0.062)	(0.070)	(0.063)	(0.066)

NOTES: Own calculations based on the estimation sample, numbers of observations as in Tables 6 and 7. The table displays overall (column (1)) and grade-specific (columns (2) to (5)) treatment effects on children's socio-emotional skills, when controlling for 15 variables measuring potential responses to treatment of children, parents, and teachers: binary indicators for the child studying at least 2 hours per day, working and helping in household, time missed school due to work or illness (see Table B10 for definitions); six indicators for parenting style (see Table B11 for definitions); indicators for the teacher using multiple teaching concepts, visual aids, and enjoys being a teacher, as well as the teacher's days of absence (see Table B5 for definitions). Missing information on the control variables is imputed with 0 and we add an indicator variable for the imputation to the set of control variables. Standard errors (in parentheses) are clustered at school level. Significance at $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

TABLE A16: 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		0.008	0.025	0.043	0.003	0.026
Grade 2	0.001	0.000	0.002	0.006	0.000	0.001
Grade 3	0.081	0.148	0.198	0.387	0.176	0.802
Grade 4	0.517	0.873	0.883	0.884	0.873	1.000
Grade 5	0.665	0.568	0.608	0.800	0.603	0.685
Patience						
Pooled		0.422	0.524	0.428	0.261	0.164
Grade 2	0.028	0.029	0.058	0.051	0.002	0.020
Grade 3	0.547	0.616	0.663	0.637	0.474	0.802
Grade 4	0.929	0.683	0.742	0.884	0.857	1.000
Grade 5	0.837	0.610	0.648	0.800	0.603	0.685
Prosociality						
Pooled		0.020	0.044	0.058	0.008	0.026
Grade 2	0.048	0.016	0.022	0.051	0.002	0.016
Grade 3	0.663	0.398	0.466	0.637	0.432	0.802
Grade 4	0.086	0.361	0.450	0.709	0.675	1.000
Grade 5	0.077	0.028	0.053	0.102	0.005	0.091

NOTES: Own calculations based on the estimation sample. **Columns (1) and (2)** state the p-values of the treatment effects on the outcome stated in the panel using conventional critical values of the t-test. In each panel, the first row displays the p-value that corresponds to the overall treatment effect 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 3. 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 7 or Figure 3). 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. **Column (3)** 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 3.2. 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)** displays 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 A17: Treatment effects on socio-emotional skills when consecutively adding control variables

	(1)	(2)
	Treatment effects when controlling for	
	pre-treat. outcomes	pre-treat. outcomes and child and family characteristics
Self-control		
Overall	0.104***	0.101***
	(0.038)	(0.037)
Grade 2	0.196***	0.185***
	(0.060)	(0.063)
Grade 3	0.129*	0.132**
	(0.067)	(0.066)
Grade 4	0.050	0.049
	(0.069)	(0.068)
Grade 5	0.029	0.028
	(0.063)	(0.062)
Patience		
Overall	0.032	0.033
	(0.041)	(0.041)
Grade 2	0.142**	0.140**
	(0.066)	(0.067)
Grade 3	−0.046	−0.047
	(0.077)	(0.077)
Grade 4	0.007	0.019
	(0.074)	(0.073)
Grade 5	0.012	0.005
	(0.076)	(0.075)
Prosociality		
Overall	0.085**	0.084**
	(0.036)	(0.035)
Grade 2	0.121**	0.123**
	(0.062)	(0.061)
Grade 3	−0.030	−0.034
	(0.073)	(0.074)
Grade 4	0.106*	0.107*
	(0.063)	(0.063)
Grade 5	0.121*	0.134**
	(0.067)	(0.067)

NOTES: Own calculations based on the estimation sample. Each cell gives the treatment effect on the socio-emotional skill in the grade stated on the left of the table. 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. The regressions corresponding to the treatment coefficients in column (1) additionally include the pre-treatment outcome variable. If the pre-treatment outcome variable is missing, it is imputed by zero. The regression also includes an imputation indicator. The regressions underlying column (2) additional include the child and family characteristics used in Table A6 and defined in Table B9. Again, missing family and child covariates variables are imputed and we control for the imputation. School-clustered standard errors in parentheses. Significance at $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

TABLE A18: Share of children reaching the highest value of a measurement (ceiling effects)

	(1)	(2)	(3)	(4)	(5)
		By school grade			
	Overall	grade 2	grade 3	grade 4	grade 5
Self-control					
Exp: Time-consistency when patient (dummy)	0.227	0.283	0.225	0.185	0.206
Svy: ISC (rev.)	0.061	0.063	0.060	0.059	0.064
Patience					
Exp: Patience	0.109	0.142	0.107	0.086	0.096
Svy: Patience survey item	0.336	0.342	0.304	0.308	0.387
Prosociality					
Exp: Altruism in dictator games	0.078	0.076	0.067	0.075	0.092
Svy: Prosocial behavior scale	0.199	0.161	0.204	0.205	0.233

NOTES: Own calculations based on the estimation sample. This table shows the share of children whose score reaches the ceiling (i.e., the highest possible value) of the respective measurement. For time-consistency, this corresponds to being time-consistent as opposed to being time-inconsistent.

TABLE A19: Treatment effects on socio-emotional skills under different sample restrictions

	(1)	(2)	(3)	(4)
	Outcome indices restricted to...			
	both components are non-missing		other indices are non-missing	
DEPENDENT VARIABLE	Number of observations	Treatment effect	Number of observations	Treatment effect
Self-control index	3,075	0.102***	3,213	0.107***
		(0.039)		(0.039)
Patience index	3,207	0.035	3,213	0.033
		(0.042)		(0.042)
Prosociality index	3,109	0.070*	3,213	0.074**
		(0.037)		(0.036)

NOTES: Own calculations based on the estimation sample. This table shows the overall treatment effects under alternative sample restrictions. First, the sample is restricted to non-missing observations in both components of a given outcome index (instead of using only one component to calculate the outcome index when the other component is missing). Column (1) displays the resulting number of observations and column (2) the corresponding treatment effects. Second, the sample is restricted to observations with non-missings in all three outcome indices. Column (3) shows the number of observations and column (4) the corresponding treatment effects. Standard errors are clustered at school level. Significance at $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

B Scales and Questionnaires

B.1 Psychometric scales

TABLE B1: 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: Own representation. Impulsivity Scale for Children (ISC) taken from Tsukayama et al. (2013).

TABLE B2: Items of the Strengths and Difficulties Questionnaire's 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: Own representation. 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 that are used to construct the prosociality scale.

B.2 School Questionnaire

Table B3: Questions about school characteristics

Variable	Definition
Number teachers	Total number of teachers in the school.
Female teachers	Number of female teachers in the school.
Number classrooms	Number of classrooms in the school.
Separate rooms	=1 if every grade is taught in a separate classroom.
Sanitary	=1 if sanitary facilities in the school are good on a good/medium/poor scale.
Furniture	=1 if school provide desks, chairs, and tables.
Overhead projector	=1 if the school has a working overhead projector.

Notes: Own representation. Questions are answered by the head teacher of each school.

B.3 Teacher Questionnaires

Table B4: Questions on pre-determined teacher characteristics

Variable	Definition
Female	=1 if teacher is female, 0 else.
Age	Teacher's age in years.
Lives in village of school	=1 if teacher lives in the village of the school in which s/he teaches, 0 else.
Completed teacher training	=1 if teacher completed a teacher training, 0 else.
Work experience	Teacher's years of work experience (current school and previous schools combined).
Reason: love of job	=1 if teacher answers the question "Why did you become a teacher? (main reason only)" with "I love teaching and I always wanted to become a teacher", 0 else (reasons "job security" and "lack of other opportunities").

Notes: Own representation. Teachers answer these questions in school (e.g., in a free period). In treatment schools, all LQ teachers were interviewed; in control schools, two randomly chosen teachers were interviewed (who were available when our research assistants conducted the interviews).

Table B5: Questions on teaching style and teacher attitudes

Variable	Definition
Multiple teaching concepts	=1 if teacher answers the question "What are your preferred teaching practices?" with "use multiple representation of concepts", 0 else (answers "rely mainly on textbook " and "use real world examples").
Visual aids in teaching	=1 if teacher uses visual aids such as maps, pictures, diagrams, charts in lessons, 0 else.
Days of absence	Number of days of work missed in the last 12 months.
Enjoy being a teacher	=1 if teacher answers the question "How much do you enjoy being a teacher?" with "very much" on a seven-point scale from 0 ("not at all") to 7 ("very much"), 0 else.

Notes: Own representation. Teachers answer these questions in school (e.g., in a free period). In treatment schools, all LQ teachers were interviewed; in control schools, two randomly chosen teachers were interviewed (who were available when our research assistants conducted the interviews).

TABLE B6: Teacher-answered LQ-specific questions

VARIABLE	DEFINITION
Popularity of LQ lessons	Teacher's answer to the question "How do you think students liked the [LQ] lessons?" on a seven-point scale from 1 ("not at all") to ("very much").
Average duration of a LQ lesson	Average duration of a LQ lesson in minutes, separately assessed for each teacher instructing the LQ program.
Problems with implementation	=1 if teacher reports that the LQ material was confusing or that she/he did not feel well prepared to teach the LQ program, 0 else.
Identification with program	Teacher's answer to the question "How much do you identify with the LQ program?" on a seven-point scale from 1 ("not at all") to ("completely identify").

NOTES: Own representation. Teachers answer these questions in school (e.g., in a free period). As these questions only apply to teachers who taught the LQ program, only these teachers are asked the questions. We interviewed all LQ teachers in treatment schools. Teachers could report multiple problems with the program implementation. The average duration of an LQ lesson is assessed separately for each grade. The other questions refer to all LQ classes a teacher taught.

B.4 Measurement of parents' socio-emotional skills

The socio-emotional skills of the parents of the sampled children are assessed in a similar manner as their children's skills. In particular, self-control, patience, and prosociality are measured by indices that combine incentivized lab-in-the-field experiments with psychometrically validated survey scales. Table B7 below presents the choice list used to elicit parents' time-consistency and patience. The former is a binary indicator that is 1 if all six decisions in choice sets 1 and 2 are identical and at least one of the 12 decisions is patient; and 0 else. The latter is the number of patient choices (ranging from 0 to 18). Experimental altruism is measured using the same dictator games as for the children (see Table 3), but the exchange rate between stars (the experimental currency) and Taka is adjusted for parents. The survey measures are the 13-item Brief Self-Control Scale (Tangney et al., 2004), a one-item patience scale (Falk et al., 2018), as well as a three-item Big Five Agreeableness scale to gauge parents' prosociality (John et al., 1991; Schupp and Gerlitz, 2008), see Table B8 for the items.

TABLE B7: Time preferences experiment for adults

	Option A (pays amount below)	**Option B** (pays amount below)	Annual interest rate (in %)
Choice set 1	tomorrow	in 3 months	
Choice set 2	in 1 month	in 4 months	
Choice set 3	in 1 year	in 1 year 3 months	
Choice 1	100	105	20
Choice 2	100	110	40
Choice 3	100	120	80
Choice 4	100	125	100
Choice 5	100	150	200
Choice 6	100	200	400

NOTES: Own representation.

Table B8: Survey measures for parents' socio-emotional skills

Brief Self-Control Scale	
Using the scale provided, please indicate how much each of the following statements reflects how you typically are (1 = "not at all" to 7 = "very much"):	
1	I am good at resisting temptation.
2	I have a hard time breaking bad habits. (reversed item)
3	I am lazy. (reversed item)
4	I say inappropriate things. (reversed item)
5	I do certain things that are bad for me, if they are fun. (reversed item)
6	I refuse things that are bad for me.
7	I wish I had more self-discipline. (reversed item)
8	People would say I have iron self-discipline.
9	Pleasure and fun sometimes keep me from getting work done. (reversed item)
10	I have trouble concentrating. (reversed item)
11	I am able to work effectively towards long-term goals.
12	Sometimes, I cannot stop myself from doing something, even if I know it is wrong. (reversed item)
13	I often act without thinking through all the alternatives. (reversed item)
Patience survey item	
We now ask for your willingness to act in a certain way [...] Please again indicate your answer on a scale from 1 to 7, where 1 means you are "completely unwilling to do so" and 7 means you are "very willing to do so". You can also use any numbers between 1 and 7 to indicate where you fall on the scale like 1, 2, 3, 4, 5, 6, 7.	
	How willing are you to give up something that is beneficial for you today in order to benefit more from that in the future?
Big Five Agreeableness	
Using the scale provided, please indicate how much each of the following statements reflects how you typically are (1 = "does not apply to me at all" to 7 = "applies to me perfectly"): I see myself as someone who ...	
1	... is sometimes somewhat rude to others. (reversed item)
2	... has a forgiving nature.
3	... is considerate and kind to others.

Notes: Own representation. The Brief Self-Control Scale is taken from Tangney et al. (2004). The patience item is the same as the one in Falk et al. (2018). This version of the Big Five Inventory has been validated by Schupp and Gerlitz (2008) and is based on John et al. (1991).

B.5 Household and Mother Questionnaire

TABLE B9: Questions on child and family characteristics

VARIABLE	Definition
Female	=1 if C is female, 0 else.
Age	C's age in years.
Number of siblings	Number of siblings (whether they live in the HH or not).
Muslim	=1 if the family is Muslim, 0 else (usually Hindu).
Grade 2	=1 if C is in grade 2 during intervention, 0 else.
Grade 3	=1 if C is in grade 3 during intervention, 0 else.
Grade 4	=1 if C is in grade 4 during intervention, 0 else.
Grade 5	=1 if C is in grade 5 during intervention, 0 else.
HH income (in 1,000 Taka)	Total HH income in 1,000 Taka (1,000 Taka ≈ $11,80 in 2019), including wages, salaries, in-kind benefits, net value of agricultural products (can be negative), and cash transfers.
Dwelling: tin wall	=1 if the main construction material of the dwelling's wall is tin, 0 else (e.g., brick, wood, bamboo).
Dwelling: tin roof	=1 if the main construction material of the dwelling's roof is tin, 0 else (e.g., brick, wood, bamboo).
Dwelling: mud floor	=1 if the main construction material of the dwelling's floor is mud, 0 else (e.g., brick).
Dwelling: area	Area family's dwelling (unit: decimal/dismil; 1 dismil ≈ 40 square meters).
Dwelling: electricity	=1 if family's dwelling is connected to the national power grid, 0 else.
M: age	M's age in years.
M: missed work due to illness	=1 if M missed work due to any chronic illness/disability in the last 30 days.
M: literate	=1 if M can read and write, 0 else.
M: primary educ. or less	=1 if M has received 3 or less years of education, 0 else.
M: secondary educ.	=1 if M has received 4–10 years of education, 0 else.
M: post-sec. educ.	=1 if M has received more than secondary education, 0 else.
M: high-status job	=1 if M's main occupation is wholesale trader, labor contractor, (service-sector) employee, doctor, advocate, tutor, Imam, or receives rent; 0 else.
M: routine job	=1 if M's main occupation is in a routine job (manual labor, industry job, transport), 0 else.
M: agricultural work	=1 if M's main or secondary occupation is agricultural (including looking after live stock and poultry on the family farm) and the main occupation is not a high-status or routine job (as defined above), 0 else.
M: not working	=1 if M is not working outside the household, also not on the family farm or in a family business (unemployed, disabled, retired), 0 else.
M: no interview	=1 if M did not answer questionnaire, 0 else.

NOTES: Own representation. C = child, M = mother, HH = household. Variables for fathers are coded analogously to variables refering to mothers. Education and occupation variables are mutually exclusive indicators.

Table B10: Items regarding children's activities

Variable	Definition
Studying: 2–3 hrs or more a day	=1 if the child spends at least 2 hours studying after school on a normal day, 0 else.
Working	=1 if the child works (either paid or unpaid) on the family's farm or in the family business or works for money for an employer other than family, 0 else.
Helping in the household	=1 if the child regularly helps in the household, 0 else.
Missed school: work	Number of days of school a child missed within the last 12 months (past year) because s/he has been working.
Missed school: health	Number of days of school a child missed within the last 30 days (last month) due to illness.

Notes: Own representation. For children aged 6–11 these questions are answered by the mother of the child. Older children answer the questions themselves.

Table B11: Six scales for parenting style

Item
Mothers assess the following statements on the scale 1 "never", 2 "seldom", and 3 "sometimes", 4 "frequently", and 5 "very frequently":
Emotional warmth
1 I use words and gestures to show my child that I love him/her.
8 I comfort my child when he/she feels sad.
13 I praise my child.
Inconsistent parenting
5 I threaten my child with punishment, but don't actually follow through with it.
16 I reduce punishments or lift them ahead of time.
18 It is hard for me to be consistent in my childrearing.
Monitoring
3 I talk to my child about things he/she has done, seen, or experienced when he/she was out.
6 When my child is outside the home, I know exactly where he/she is.
15 I try to actively influence my child's circle of friends.
Negative communication
2 I criticize my child.
9 I shout at my child when he/she did something wrong.
14 I scold my child when I am angry at him/her.
Psychological control
10 I feel that my child is ungrateful because he/she disobeys.
11 I stop talking to my child for a while when he/she did something wrong.
17 I am disappointed and sad when my child misbehaves.
Strict control
4 I punish my child when he/she was disobedient.
7 I tend to be strict with my child.
12 I make it clear to my child that he/she should not oppose orders and decisions.

Notes: Own representation. These items have been taken from the Panel Analysis of Intimate Relationships and Family Dynamics project's parenting questionnaire (Thönnissen et al., 2019) and are, for example, also used in the German Socio-Economic Panel Study (Richter et al., 2013). The numbers in the left column indicate the order in which the items are included in the questionnaire.

References

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.

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

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

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.

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.

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

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

John, O. P., E. M. Donahue, and R. L. Kentle (1991). *The Big Five Inventory—versions 4a and 5*. University of California, Berkeley, Institute of Personality and Social Research.

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.

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

Richter, D., M. Metzing, M. Weinhardt, and J. Schupp (2013). SOEP Scales Manual. SOEP Survey Paper (Technical Report) 138, German Institute for Economic Research (DIW Berlin), Socio-Economic Panel Study.

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.

Schupp, J. and J.-Y. Gerlitz (2008). Big Five Inventory-SOEP (BFI-S). Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS).

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.

Thönnissen, C., B. Wilhelm, P. Alt, H. Greischel, and S. Walper (2019). Scales and Instruments Manual, Waves 1 to 10; Release 10.0. Technical Report, Panel Analysis of Intimate Relationships and Family Dynamics.

Todd, P. E. and K. I. Wolpin (2003). On the Specification and Estimation of the Production Function for Cognitive Achievement. *The Economic Journal 113*(485), F3–F33.

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.