HCEO
hceconomics.org

Human Capital and Economic Opportunity Global Working Group

Working Paper Series

Working Paper No. 2015-003

Good Things Come to Those Who (Are Taught How to) Wait: Results from a Randomized Educational Intervention on Time Preference

Sule Alan
Seda Ertac

February, 2015

Human Capital and Economic Opportunity Global Working Group
Economics Research Center
University of Chicago
1126 E. 59th Street
Chicago IL 60637
www.hceconomics.org

Good Things Come to Those Who (Are Taught How to) Wait: Results from a Randomized Educational Intervention on Time Preference*

Sule Alan
University of Essex

Seda Ertac
Koc University

September 2014

Abstract

We report results from the impact evaluation of a randomized educational intervention targeted at elementary school children. The program uses case studies, stories and classroom activities to improve the ability to imagine future selves, and emphasizes forward-looking behavior. We find that treated students make more patient intertemporal choices in incentivized experimental tasks. The effect is stronger for students who are identified as present-biased in the baseline. Furthermore, using official administrative records, we find that treated children are significantly less likely to receive a low "behavioral grade". These results are persistent one year after the intervention, replicate well in a different sample, and are robust across different experimental elicitation methods.

Keywords: *intertemporal choice; randomized field experiments; non-cognitive skills*
JEL Codes: *C93, D91, I28*

Sule Alan, Department of Economics, University of Essex. Email: salan@essex.ac.uk

Seda Ertac, College of Administrative Sciences and Economics, Koc University, Rumeli Feneri Yolu, Sariyer, Istanbul 34450 Turkey. Email: sertac@ku.edu.tr

*This paper is part of a field project mainly funded by the Turkish division of the ING Bank. Other funders are Turkish Capital Markets Board Intermediaries Association, TUBITAK (Career Grant 111K444), Koc University, and the TUBA-GEBIP Program, whom we would like to thank for generous financial support. We are grateful to conference participants at the 2013 workshop on "Self-control, Self-regulation and Education" at Aarhus University, the 2014 North American Meeting of the Econometric Society, 2014 FUR Conference, and seminar participants at George Mason University, Erasmus University, Goethe University, Sabanci University, Norwegian School of Economics and University of Essex for helpful comments. We would also like to thank Elif Kubilay, Nergis Zaim, Banu Donmez, Enes Duysak, Emre Karabulutoglu and Aslihan Tutuncu, as well as numerous other students who provided excellent research assistance. We also gratefully acknowledge financial support from the European Investment Bank Institute through its EIBURS initiative. The findings, interpretations and conclusions presented in this article are entirely those of the authors and should not be attributed in any manner to the European Investment Bank or its Institute. All errors are our own.

1 Introduction

A growing body of research shows that certain attitudes and personality traits, also referred to as "non-cognitive skills", are strongly associated with achievement in various economic and social domains. Among these traits, patience and self-control attract particular attention, as empirical studies show their predictive power on outcomes including educational attainment, occupational success, and a range of health outcomes such as obesity and substance abuse.[1] Furthermore, these particular traits are likely to be associated with rational financial decision making and the ability or willingness to acquire and use financial knowledge, that is, lack of them might be partly responsible for observed insufficient savings, overindebtedness and even poverty.[2]

An important strand of the literature on preferences, attitudes and traits studies children, and shows that the childhood period is important for the formation and development of a crucial set of non-cognitive skills, which appear to be at least as important as cognitive skills in determining social and physical well-being in adolescence and in adulthood (see Heckman et al. (2006)). In particular, among children and adolescents, impatience has been found to be associated with a higher likelihood of using alcohol and cigarettes, a higher body mass index, a lower propensity to save, lower grades, and more disciplinary conduct violations at school (Castillo et al. (2011), Sutter et al. (2013)). These correlations actually tend to persist into adulthood: studies in both economics and psychology have shown that individuals who displayed patience and self-control as a child or adolescent have better outcomes in terms of health, performance in school, labor market success, social competence, and lifetime income (e.g. Golsteyn et al. (2013), Moffitt et al. (2011), Mischel et al. (1989)).

While the literature cited above aims to answer the question of what factors are associated with, for example, being more patient or more able to delay gratification, little is known about whether it is possible to *make* an individual more patient or more able to cope with self-control issues, that is, whether these traits are malleable. The evaluation literature provides valuable insights in this regard. Early childhood interventions, which are typically directed at bolstering cognitive skills in order to close achievement gaps, seem to generate favorable outcomes in various key domains in social life, especially in disadvantaged children. However the main channel seems to be enhanced non-cognitive skills rather than cognition (see Heckman and Kautz (2014) and the references therein for an extensive review of these interventions).

Further evidence on the potential malleability of non-cognitive skills comes from a separate but related literature, which shows that life experiences and exposure to exogenous shocks such as conflict or natural disasters can alter preferences and personality traits in general, and time preference in particular (e.g. Perez-Arce et al. (2011), Voors et al. (2012)). Coupled with the evidence that personality traits are influenced by the childhood environment (Borghans et al. (2008)), these find-

[1]See Fuchs (1982), Laibson (1994), Laibson et al (1998), Bickel et al (1999), Della Vigna and Paserman (2005), Heckman et al (2006), Knudsen et al (2006), Ameriks et al (2007), Meier and Sprenger (2010), Jamison et al (2012), Finke and Huston (2013).

[2]Recent evidence indeed shows that individuals who discount the future less are more likely to acquire financial literacy (Meier and Sprenger (2013)).

ings provide reason to expect that a targeted educational program on forward looking behavior in children can make an impact on time preferences. Supporting this on the theory side, Becker and Mulligan (1997) suggest that forward looking behavior and the ability to focus on the future when making intertemporal consumption choices can be learned as part of general human capital accumulation. This model in fact can provide a useful backdrop for fleshing out a potential mechanism through with an educational intervention on forward-looking behavior might work.

Motivated by the aforementioned correlational studies and the literature on interventions, we design and evaluate an educational program in collaboration with an interdisciplinary team, targeting 3rd and 4th graders in elementary schools in Turkey. The intervention specifically aims to improve the ability to be "forward-looking" and to exercise self-control and delay gratification in intertemporal choices, through a set of carefully-designed education materials (case studies, stories, in-class games) that are conveyed by trained teachers. The core idea behind the program is to impart the ability or habit of imagining the future consequences of alternative courses of action in order to make future utilities vivid and less remote, i.e. teach forward-looking behavior.

The program is implemented in a randomized-controlled fashion at the school level. We use a version of a "phase-in" design, where subgroups of participating schools implement the program in different periods of time. The initial phase of the program, in which the teachers in a random subset of the participating schools received the training material, was implemented in Spring 2013. The next phase was implemented on another subgroup of treatment schools in Fall 2013, practically replicating the intervention. Students that are in the remaining schools, which constitute our "pure control" group, never received training and have moved on to middle school at the end of the experimental period.

We compare the intertemporal choices of these three treatment groups in three different measurement phases (end of Spring 2013, end of Fall 2013, end of Spring 2014). This, combined with the phase-in nature of the design, allows us to both study the temporal nature of the treatment effect at different intervals after the conclusion of the program, and to establish robustness of results across different samples and measurement methods. The effectiveness of the program is measured in terms of students' behavior in incentivized experimental tasks as well as their actual behavioral conduct, as graded and reported to the school administration. In addition, rich information is collected on student, teacher and family characteristics via baseline surveys, and assessment reports by the teacher for each individual child.

We find that treated students make significantly more patient intertemporal choices in incentivized time preference elicitation tasks. More specifically, we find that treated students demand about 22 to 32% less gifts to accept to wait for one week. This result is (1) robust to the use of alternative elicitation tasks, (2) persists one year following the intervention, and (3) persists under anonymity. We find that, controlling for baseline patience, the program predominantly benefits students who were identified to be present-biased in the baseline: the treated among initially present-biased students end up demanding 55% less gifts than the rest of the students of the same type. A quite striking finding that emerges from our data is that one year after the intervention, treated students are about 9 percentage points less likely to receive a low "behavioral grade", based on official school administrative records. This finding is consistent with the recent evidence on the relationship between time preference and disciplinary conduct at school (Sutter et al. (2013), Castillo et al. (2011)). Our results, both quantitatively and qualitatively, apply to both treatment groups in the study, that is, the estimated treatment effects on all our outcomes coming from the initial treatment group are robustly replicated in the second treatment group in the phase-in design.

The paper makes several contributions to the literature. To our knowledge, this is the first study to date that i) directly targets specific non-cognitive skills in a classroom environment at such young ages, ii) evaluates causal impact by measuring changes in experimental and real-life behavior and outcomes within a randomized-control framework, and iii) quantifies short- and longer-term impact that is robust to using alternative measurement methods. The paper also offers useful insights for guiding education policy. There has been widespread concern over dismal school outcomes and widening achievement gaps in most countries, despite genuine efforts by governments and sizeable resources devoted to both increasing school attendance and the quality of teaching. Our study can be viewed as a cost-effective supplement to these efforts. Embedding character education into the general elementary school curriculum via fun activities and games tailored to core objectives takes little effort and drains much less resources than most intervention programs, and can produce significant improvements in key non-cognitive skills such as the ability to delay gratification.

The rest of the paper is organized as follows: Section 2 provides the background of the program, Section 3 discusses potential mechanisms of behavioral change, Section 4 explains the evaluation design and measurement procedures, Section 5 presents the data and discusses the results, and Section 6 concludes.

2 Background

As of 2012, twelve years of compulsory education is divided into 3 stages of 4 years of schooling in Turkey. Transition to middle school is required after grade 4 and the pupil receives the next 4 years of compulsory education in middle school, and the last 4 in high school. A single teacher usually teaches all grades at the elementary level (for 4 years).

The program we evaluate is one of the arms of a set of randomized-controlled interventions designed to improve certain non-cognitive skills in elementary school children in the classroom environment. The program is implemented in a large number of state elementary schools in Istanbul, Turkey with the permission and oversight of the Ministry of Education.[3] While a few decades ago only the very rich would send their children to private schools, the enlarging middle class of Turkey now mainly prefers private schools to public schools for their children. Therefore, the program mainly reaches students from lower socio-economic backgrounds, although there still is considerable

[3]The Ministry of National Education is centrally responsible for developing and monitoring curricula, allocating teachers, building schools and communicating with the private sector and NGOs for extra-curricular projects. For the tasks used in measurement, we obtained ethics approval from the local IRB, which included parental consent requirements. Consent was denied by the parents only in case of five students.

variation in socio-economic status across students in our sample, which we exploit in our analyses.

The program we help develop and evaluate was offered as part of a corporate social responsibility project of a major international bank's Turkish division. The Ministry of Education encourages schools and teachers to participate in socially useful projects offered by the private sector, NGOs, government and international institutions. These projects, upon careful examination and endorsement by the Ministry, are made available to interested schools. The ministry allows up to 5 lecture hours per week for project-related activities. Participation in these projects is at the discretion of teachers, however teachers and school administrations (headmasters) receive a type of "soft incentive" involving merit points based on the projects they are involved in. Many of these projects involve issues such as environment, art, foreign languages, health, dental care etc. In the absence of any projects, students use the free hours as unstructured playtime, so these projects do not crowd-out any core teaching.

After obtaining permission from the Ministry, contacted teachers were informed that upon participation, they would be invited to seminars on topics around financial awareness and savings, and that they would be asked to cover materials prepared for the students within an eight-week period for at least 2 hours a week. They were not informed about the nature and content of the materials until they arrived at the seminars and they were never informed about any measurement procedure throughout the project. The training seminars were conducted by education consultants and overseen by the Ministry, where teachers were given extensive training on the use of the materials and on the importance of these non-cognitive skills for achievement in many domains. The education materials, including many supplementary classroom activities related to the core topics were delivered to the teachers after all baseline data were collected and teacher training seminars were conducted. In the implementation phase of the intervention, the materials were conveyed to students within the weekly free hours allocated to teachers.

While the target concepts of the materials were determined and conveyed by the authors, the contents were shaped by an interdisciplinary team of education psychologists, a group of elementary school teachers, and children's story writers, according to the age and cognitive capacity of the students.[4] The core material involves 8 mini case studies with topics including imagining the future-self (forward-looking behavior), self-control against temptation goods, smart shopping, games to make future utilities vivid and close-by, saving for a target, viewing and evaluating alternative future outcomes, and developing coping mechanisms against temptation to meet a savings target. For example, in the first week children cover a case study titled "Zeynep's Time Machine", telling the story of Zeynep, a girl who wants a bike for which she has to save, but is also faced with alluring short term consumption possibilities. The time machine allows Zeynep to go to the two alternative future states (having saved for the bike or not), and observe the consequences of her decisions. Students are asked to imagine themselves in similar situations, and these types of games are played in class every week. It is important to note that special care was taken to use the stories to start discussions about alternative courses of action and to encourage forward-looking behavior rather

[4] All training materials are property of the ING Bank of Turkey and can be viewed at www.turuncudamla.com.tr.

than giving children direct, unquestioned "advice" to act patiently.

3 Mechanisms of Behavioral Change

The core idea behind our intervention is that rather than being fixed traits that stay the same over the life cycle, individuals' time preferences are malleable and in particular can be influenced within the school environment. While genetics as well as parental characteristics and investment in the pre-school period may shape preferences initially, these preferences can evolve further during the educational phase, and in particular, may be influenced through targeted programs. From a theoretical standpoint, there can be different mechanisms through which an educational intervention like the current one might affect intertemporal choices. One framework that fits very well with the ideas put forth in our educational content is the one proposed by Becker and Mulligan (1997), who argue that foresight can be improved, i.e., investment in the ability to judge the benefits of the future in exchange for today's pleasures is possible through "individual effort". The idea is that individuals can spend mental resources and learn to make future situations less remote. This can help them evaluate future utilities correctly and learn to reason with the present-self better, lessening the pressure of immediate gratification. A simple intertemporal model that captures this reasoning is as follows:

Consider two periods: pre-program (1) and post-program (0). Suppose child i is endowed with discount factor δ_0 pre-training, as a function of

$$\delta_0 = \varphi(X^P, I^P)$$

where X^P is a vector of background characteristics and skills (possibly genetic) and I^P is parental investment for child's time preference where φ is a concave function such that $\frac{\partial\varphi}{\partial I^P} > 0$ and there is diminishing returns. Now, suppose that while individuals' time preferences are initially shaped by parental characteristics and investments in the pre-school period, they may evolve further in the school environment through programs specifically designed to influence them. To formalize this, we postulate that the post-training discount factor δ_1 for child i is given by:

$$\delta_1 = \psi(e(I^T), \delta_0(.))$$

where e is the effort child i exerts to make distant pleasures less remote and to put more weight on future utilities. The function ψ is such that

$$\frac{\partial\psi}{\partial I^T} = \frac{\partial\psi}{\partial e}\frac{\partial e}{\partial I^T} \geq 0$$

And, $\frac{\partial\psi}{\partial e}>0$, $\frac{\partial e}{\partial I^T} \geq 0$. An educational investment implemented by the teacher (I^T), which specifically targets forward looking behavior may promote the effort the child puts into visualizing and judging (weighting) future utilities, by either teaching the benefits or reducing the costs of such effort. In either case, the expected behavioral change would be for the discount factor to increase and the child to make more patient choices. This type of investment is typically expected to be made by parents, however, we argue that a well-equipped teacher can also contribute greatly to the development of a child's time preferences. Note also that these two types of investments could be substitutes, or complements, such that in the case of perfect complementarity, the teacher's investment would have no effect on the time preferences of the child without parental investment.

While a model that emphasizes the ability to evaluate future utilities fits well with our educational intervention, other mechanisms could also be in effect to lead to similar changes in intertemporal choice. One such alternative is self-signalling. Several papers in the literature have put forward models where individuals care about their self-image in their own eyes, and can change their behavior to manage their own impression of themselves (e.g. Bodner and Prelec (2003)). If the education works to instil patience or self-control as a "valued personality trait" in children's minds and makes children aspire to being a patient person, this could generate an extra utility component from being able to delay gratification and thereby lead to more patient behavior. Such a model would give rise to the same hypothesis as above, which is that children who are exposed to the program will make more patient choices when faced with an intertemporal decision problem.

It is important to stress here that the point of our paper is not to test any particular theory, nor is it an attempt to differentiate amongst some competing theories. Our focus in the paper is confined to putting forward novel empirical evidence that time preferences may be indeed malleable in the short and the longer-term via targeted educational programs.

4 Design and Measurement

4.1 Evaluation Design

Our experimental protocol follows a version of a "phase-in" design, where randomly selected schools (unit of randomization) receive training at different points, in order to measure causal impact and assess the ability of the program in generating robust treatment effects across different phases and samples.

After the official documents were sent to all elementary schools in designated districts of Istanbul by the Istanbul Directorate of Education, 3rd grade teachers in these schools were contacted in random sequence and were offered to participate in the program. They were informed that upon participation they would be assigned to different training phases within the coming two academic years. Once a teacher stated a willingness to participate, we assigned her/his school in to one of the three arms of the intervention: A willing teacher had a 40 percent chance of being in the initial treatment group (IT henceforth), a 30 percent chance of being in the control-then-treatment group (CT henceforth), and a 30 percent chance of being in the pure control group (PC henceforth).

The randomization was performed among the schools in which at least one teacher stated willingness to participate in the program. Therefore, the estimated impact of the program is the average treatment effect on the treated and is not readily generalizable to the population. Approximately

60% of the contacted teachers accepted our offer and the most common reason for non-participation was being "busy with other projects, although happy to participate in this program at a later date" (about 20%). The rest of the non-participation was due to "impending transfer to a school in another city, with a willingness to participate if the program is implemented there" (about 5%), and being "not in a position to participate due to private circumstances" (about 10%).[5] It should also be noted that random assignment was done at the school level and not at the classroom level, since the physical proximity of classrooms and teachers would be likely to generate significant spillover effects.[6]

We performed the random assignment immediately after a positively ended phone call, as we had to inform the teachers for the upcoming training seminars in case they were assigned to the initial treatment group. We offered two alternative dates to make sure they could come to the seminars, which were conducted by education psychologists and took an entire day. After we reached the number of schools for which teacher seminars and data collection were feasible, we stopped the calls. With this procedure, we ended up with 15 schools in IT, 10 in CT and 12 in PC, totaling 37 clusters (schools).

All involved teachers were promised to eventually receive all training materials and to participate in training seminars, but they were not told when within the next two academic years they would receive the treatment, until the random assignment was completed. The promise of the training offer was made to the teacher and not to current students, that is, while children in the pure control group never received the training as they moved on to middle school after year 4, their teachers will, albeit at a later time.

Table 1 shows the design of the implementation and measurement phases. After the randomized sample was obtained and before launching training seminars for teachers in initial treatment group (IT), data on a large set of variables were collected from the entire sample through student, teacher and parent surveys, to obtain baseline measurements (Phase 0). The first phase of the program was implemented in Spring 2013 by the teachers in the IT group and the second phase in Fall 2013 by the teachers in the CT group. We collected the experimental outcome measures in 3 phases by physically visiting all classrooms.[7]

In Spring 2013, after the completion of training by the IT group we collected initial experimental data from all treatment arms (see Section 4.2 for the details of the measurements). Note that at this point, neither the teachers nor the students in the CT and PC groups had received training. Therefore data on this phase allow us to assess the first, short-term impact of the training. At the end of Fall 2013, we collected another set of experimental data, again from all groups. Note that at this time our CT group had completed the program. Data from this phase allow us to assess 1) the impact of the training on a new sample, i.e. establish robustness across samples, 2) whether

[5]The remaining 5% promised to call back with their response but never did.

[6]Istanbul is a very large city and our schools are geographically spread out. We therefore do not expect any spillover problems across schools.

[7]Incentivized experiments were run by the authors and a select group of student assistants extensively trained by the authors. We were also helped by a set of experienced interviewers contracted by a field company and trained by us, in conducting surveys.

the impact measured in the first phase persists after eight months following the initial treatment, i.e. establish temporal robustness. Finally, in Spring 2014 we collected the last set of experimental data from all treatment arms. The data from this final phase allow us to assess longer-term impacts on experimental outcomes and, as we also collected end-of- 4th-year administrative records for each pupil before they proceed to middle school, whether there is any impact of the program on a crucial real outcome: behavioral conduct.

An important concern with respect to the implementation of the program is teacher heterogeneity. Teachers differ in terms of their style of teaching, their dedication and their overall attitude toward any given subject. Despite its well-structured form, the program we offer is expected to be implemented in different ways and styles by different teachers, as is true for any other subject taught in elementary schools, including core subjects like mathematics and science. Indeed, an anonymous survey we conducted at the end of the whole study in June 2014 reveals that about 24% of the treated teachers implemented the training material in a highly intense way, 73% implemented it at a moderate intensity, and the rest said they did not have a chance to implement it (see Figure 1).[8] Given this heterogeneity, our treatment variable should be thought of as "program offer" rather than the program itself, since we can never estimate the impact of the program itself. However, it should be noted that this is not a shortcoming of the design and the effects we estimate are in fact the relevant inputs for policy, since if and when a program such as the current one is scaled up, teacher heterogeneity in implementation is inevitable.

4.2 Experimental Tasks and Procedures

The core component of our methodology to evaluate the impact of the program is incentivized experiments, which elicit time preference, time inconsistency and risk tolerance using real stakes. Measurement of these traits is now quite standard through widely-used experimental tasks (e.g. Gneezy and Potters (1997), Holt and Laury (2002), Harrison et al. (2005), Andreoni and Sprenger (2012)).[9] We also evaluate the impact of the program on the end-of-year behavioral conduct grades given to each pupil. These grades are part of the official administrative records kept for each pupil in the school's database.

We use two tasks for eliciting time preferences: (1) a "multiple price list" (MPL) task, (2) a "convex time budget" (CTB) task. These tasks are used in different measurement phases, as explained in Table 1. For the 1st phase of measurement, we use the MPL task. In this task, subjects make a series of choices between a fixed amount to be received today, and increasingly larger amounts to be received in the future.[10] The minimum larger-later amount that induces the individual to be willing to wait is a measure of impatience. That is, more impatient individuals

[8] The question was "how intensely did you apply the training material?" and teachers could choose three options: "I implemented it at a high intensity, "I implemented it at a moderate intensity", "I did not quite have a chance to implement it"

[9] Using the terminology of Harrison and List (2004), our experiments can be classified as artefactual field experiments.

[10] See Andersen et al. (2006) for a methodological discussion. The task has also been used in children and adolescents, by Bettinger and Slonim (2007), Castillo et al. (2011) and Sutter et al. (2013), among others.

require a larger premium to be willing to sacrifice current consumption and wait for the future reward. In our experiments, we fix the earlier reward to be 2 gifts out of a gift box that contains toys, stationary, hair bands etc., whereas the larger reward ranges between 2 and 10 gifts (see Appendix). We give children two multiple price list sheets that include 9 decisions each, between: (1) two gifts today versus more gifts to be received one week from today, (2) two gifts to be received in one week from today versus more gifts to be received in two weeks from today. At the end of the experiment, one decision out of one of the two lists is randomly selected, and rewards are given according to subjects' choices in the selected decision problem. By keeping the delay length the same and varying the delay to the earlier reward, these two sets of decisions allow us to both measure patience and identify time inconsistency. As shown in Table 1, we use the MPL task to compare three groups: the IT group, the CT group, and a random subset of the PC group. Because of the phase-in design and the measurement timetable, only the IT group had received the treatment when the MPL task was implemented, and the other two groups serve as control in this task.

The second type of task we use is a Convex Time Budget (CTB) task, adapted from Andreoni and Sprenger (2012).[11] In this task, children are asked to allocate 5 tokens between an earlier and a later option, where waiting pays according to an interest rate. We keep the timing of the early reward and the delay consistent with the MPL task (today vs. one week later and one week vs. two weeks). For each of these time profiles, subjects make one decision with an interest rate of $r=0.25$ and one with $r=0.5$. In total, children make 4 decisions, one of which is selected randomly and implemented (see Appendix). While the MPL task was implemented when only the IT group had received treatment (Phase 1), the the CT group had also received the training when we implemented the CTB task (Phase 2). That is, the treatment group in the analyses using CTB involves both the IT and CT groups, whereas PC serves as control.

The differences in the tasks and the timing of measurements allow us to study the across sample and temporal robustness of program effects as well as robustness to alternative measurement methods. In the 3rd and final measurement phase, the CTB was used again, however, the procedures were somewhat different. Specifically, the children now made the intertemporal allocation choice anonymously, and everyone in the class received the same early and late reward, based on the random selection of one (anonymous) decision. This third measurement phase was intended to give the training the least chance to work, in the sense that (1) for the IT group more than one year had passed, and for the CT group about 6 months had passed since the implementation of the program, (2) with anonymity, there could be no motive to please the experimenters and/or the teacher or classmates (see Section 5.6 for a more extensive discussion of this concern).

In addition to intertemporal choices, we also have access to the choices of the children under risk. For risk preference elicitation, we use a version of the task in Gneezy and Potters (1997), where children have 5 tokens to allocate between a riskless option and a risky option. With 50% chance, the tokens invested into the risky option are tripled, and with 50% chance they are lost,

[11]The reasons for using a different elicitation task was to prevent children from making the same choice as before to appear consistent, and to test the robustness of the treatment effect with another one of the major tasks used in the experimental literature on time preference.

depending on the color of a ball that the child draws from an opaque urn that contains one yellow and one purple ball. Tokens invested in the riskless option are safe. We use the tokens invested in the risky option as a measure of risk tolerance.

All experiments were run in-class, with pencil and paper. The gift bag contained sufficient numbers of all types of attractive items, and children were assured that the same types of items would be available to choose from in future rewards.[12] We took great care to ensure that teachers were not present when we explained the procedures and children made their choices. We also arranged the timing of the visits to different classrooms within one school such that children from classes would not have a chance to talk to each other about the measurements.

5 Results

5.1 Data

We have data from 3rd and 4th grade students in 73 classes in 37 schools, who are 9 to 10 years of age. Our experimental outcome variables are obtained from students who were present in class on the day of the experiments, for others experimental outcomes are missing (this amounts to 1921 observations in the MPL task and 1846 observations in the CTB task). We first assess whether the key covariates we use that are expected to be highly predictive of our outcome measures are balanced across treatments, that is whether the randomization was successful. For this, we estimate the probability of being assigned to a treatment arm conditional on these covariates. These covariates include gender, teacher-assessed academic success, teacher-assessed behavioral conduct, socio economic status as well as cognitive function measured by Raven's Progressive Matrices (Raven et al. (1986)). They also include class size, teacher gender, teacher age and the self-reported "patience" of the teacher. Table 2 presents the results of logit regressions of treatment status on these variables.[13] Recall that we have three treatment cells, including control. We compare each treatment cell with one another and present them in different columns. For example, the first column presents the coefficients of a logit regression of initial treatment status (IT) without pure control (PC), so as to assess the balance across treatments IT and CT. Inspecting the simulated p-values, we confirm that the randomization was successful and, ex-ante, none of our treatments is correlated with observable characteristics. With respect to individual coefficients, we observe a p-value of less than 10 percent only on one occasion, which can be attributed to random chance.

[12] Because incentives are of paramount importance, we extensively surveyed several toy shops as well as children and mothers outside our sample to identify items that are extremely desirable for children of this age in recent days. We also made sure to use different sets of toys and gifts in different measurement phases, to avoid satiation and maintain comparability.

[13] The simulated p-value is obtained by randomly re-assigning treatment status to schools and re-estimating the same logit model and obtaining the associated chi-square value. Under the null hypothesis that these covariates are jointly independent of the treatment status, these p-values presented in the last row give the probability of obtaining chi-square values as big as the one obtained from the actual sample.

5.2 Baseline Correlations with Real Outcomes

We start by noting that our measures have strong links to children characteristics and actual outcomes. This suggests that, as expected based on the literature, the experimental measures are truly eliciting a personality trait and they have predictive power over actual decisions and outcomes. In order to report these correlations in the absence of any intervention, we use the control subset of our data. Given our measurement timetable, this subset is composed of the CT and a subset of PC in the MPL task, and the PC in the CTB task. Table 3 shows that impatience, as measured by the average number of early choices in the MPL task or the average early allocation in the CTB task, is negatively correlated with school grades and cognitive function (Table 3, columns 1 through 4). Children that are assessed by the teacher to be well-behaved make more patient choices in both tasks, whereas children assessed by the teacher to be easily distracted are more impatient (Table 3, columns 5 through 8).[14]

5.3 Treatment Effect on Experimental Measures

Based on the null hypothesis that "the program made no impact on the experimental outcome y^E , we postulate our main empirical model as follows:

$$y_{ij}^E = \alpha_0 + \alpha_1 T_j + \epsilon_{ij}$$

where y_{ij}^E denotes the experimental outcome, that is, the number of early choices in the MPL task or the early allocation in the CTB task by student i in school j, T_j is a binary variable that indicates the treatment status of school j, and by design $T \perp \epsilon$.[15] We also estimate the average treatment effect by conditioning on baseline covariates:

$$y_{ij}^E = \alpha_0 + \alpha_1 T_j + X_{ij}'\gamma + \varepsilon_{ij}$$

where X is a vector of observables that are potentially predictive of the outcome measures we use. These include student gender, teacher-assessed academic success, behavior assessment, socio-economic status as well as cognitive function. The estimated $\hat{\alpha}_1$ is the average treatment effect on the treated in both specifications. Since the covariates contained in X are also orthogonal to

[14] The table presents results on math grades but the results do not change if we use other grades such as Turkish or science. The results are also robust to using the today-one week choices rather than taking the average of all decisions.

[15] Note that the latter condition fails to hold if the program also affects some other traits that are correlated with the outcomes we measure, in which case the estimated effects can be considered as "total" effects of the program. Our study design can partially address this concern as we also measured some key personality attributes using the Early Adolescent Temperament Questionnaire"(Capaldi and Rothbart (1992), Ellis and Rothbart (2001)) in the first measurement phase. This survey measures seven personality traits: inhibitory control, activation control, surgency, fear, frustration, and aggression, and we also have a measure of grit. These traits are measured post-program for the IT group and pre-program for the CT and PC groups. Regressions (available upon request) show that the program has made no impact on any of these traits.

treatment status, including them in the regression does not affect the consistency of $\hat{\alpha}_1$, nor does their exclusion lead to omitted variable bias. Estimating the first equation implies $\epsilon_{ij} = X_{ij}^{'}\gamma + \varepsilon_{ij}$, so that estimating the second is expected to reduce the variability of regression errors, leading to more precise estimates of treatment effects.

5.3.1 Treatment Effect in Measurement Phase 1

We first evaluate the initial phase of the program using our first experimental outcome measure, collected shortly after the end of the program for the initial treatment group. Figure 2 shows the average number of early choices in the today-one week and one week-two weeks decisions sheets in the MPL task, for each treatment status.[16] We observe here that in both cases, children in the treatment group (IT) make fewer early choices. Since the 9 decisions in an MPL sheet each have 1-gift increments for the future reward, one less early choice corresponds to the child demanding one less gift to accept to wait for a week. The figure shows that in the treatment group, children demand on average 0.86 less gifts to wait for a week from today, and 0.84 less gifts to wait for a week from next week.

Table 4 statistically formalizes these effects. While the first column presents the average number of early choices for the control group (CT+PC), rows 1 and 3 report the differences in means and, rows 2 and 4 present the estimated treatment effects by controlling for a set of covariates. In all models, the estimated treatment effect is highly significant and its size ranges between 0.84 and 0.95 gifts. Put in relative terms, we find that the program reduced the average number of early choices measured by MPL task by a statistically significant 30 percent approximately (since the mean of the control group is about 3 early choices). [17]

These values constitute our first, short term impact estimates which compare the students in the initial treatment group with those in the control group. In what follows, we examine the second phase and show that the effects we estimate in the medium term for the IT group are both qualitatively and quantitatively similar to these short term estimates. Moreover, we show that the program produces quantitatively similar results in another sample (the CT group), although behavior is measured by a different task.

[16] There is a portion of our subjects (11.3%) who make choices with multiple switchpoints in one or both of the MPL sheets, or profess to not understanding the task. We exclude these observations from our analyses. It is important to note that the incidence of such errors is balanced across the three treatment groups. In addition, the incidence is, as expected, strongly negatively correlated with cognitive ability.

[17] A common concern in intertemporal experiments where the early reward comes today is a potential lack of trust in receiving future gifts. In terms of estimating the treatment effect, this should not be an issue because of the randomized-controlled nature of our framework. As long as the issue of trust is not different across treatment status, even if it is present, it would not bias our estimates. Still, we can check for such a possibility using elicited risk preferences, and we observe that risk preferences do not affect early choices differently in treatment and control. Incidentally, we find that the elicited risk preferences are independent of treatment status i.e., are not affected by the program (unreported regressions available upon request).

5.3.2 Treatment Effect in Measurement Phase 2

We next look at the impact of the program in the 2nd phase, where the treated group now includes both IT and CT, and the control group is PC. The outcomes measured in this phase for the IT group can be considered as medium term outcomes, as the measurement was carried out eight months following the implementation of the program, and they can be considered short-term outcomes for the CT group. In this phase we use the CTB task, with 4 decisions that vary in (1) the delay to the early reward, (2) the interest rate. Figure 3 shows the average number of tokens allocated to the early date for the treatment and control groups when the decision is between today and one week later, while Figure 4 presents decisions made between one week later and two weeks later.

In both figures, the first observation of note is that children allocate fewer tokens to the early date when the return to waiting is larger. That is, children rationally respond to incentives to save.[18] The figures also show that as in the MPL task, children who received the education (IT and CT groups combined) allocate fewer tokens to the early date, in all 4 decisions.

Table 5 presents the estimated effect sizes and associated p-values for r=0.25 and r=0.5, when the decision is between today and one week later. The first rows of each panel in the table confirm that there is a significant treatment effect at the 1% level on early allocations when we look at all treated students combined (IT & CT) vs. pure control (PC). Treated children allocate 0.43 fewer tokens to the early date with the low interest rate and 0.41 fewer tokens with the higher interest rate. Considering that the average numbers of early allocations in the control group are 2.02 and 1.76 tokens with the lower and higher interest rates respectively, these effects amount to statistically significant 21 and 23 percent reductions in the tokens allocated to "today" in the treatment group. The effects are robust to the addition of the set of correlates pertaining to the child and the teacher.

Table 6 reports similar results for the one week-two week decisions. The estimated treatment effects are 0.46 and 0.45 for the low and high interest rates, respectively, indicating that treated students allocate approximately 23 and 24 percent fewer tokens to the early date relative to the control group (control means are 2.04 and 1.86 tokens, respectively). Note that these estimated effects are quite similar to those obtained from the MPL task in percentage terms.

A natural question here is whether the effect of the program is different between the IT and CT groups, since they received the training at different times but were measured at the same time, with the data on the CTB task collected immediately after the treatment for the CT group and about 8 months after for the IT group.[19]Rows 3 to 6 in each panel of Table 5 show significant treatment effects for both IT and CT groups estimated against the pure control group. What is striking is that the estimated effect sizes are statistically equal across IT and CT, suggesting that

[18]The average differences between early choices in the full dataset with r=0.25 and r=0.5 are significant at the 1% level in t-tests, both for today-one week and one week-two weeks.

[19]Educational interventions given within this project are not confined to the treatment that is subject of this paper. The IT group in fact received additional materials targeting some other non-cognitive skills, which may or may not have affected the 2nd and 3rd phase results we present here for that group. Given the statistically equal effect sizes we estimate in both the 2nd and the 3rd phase for IT and CT, who did not receive any additional materials, there is strong evidence that additional material did not affect the IT group's experimental outcomes over and above the first material.

the program generates a persistent impact on these experimental outcomes. Formally, we do not reject the equality of treatment effects in the IT and CT groups against the PC group (Wald tests, p=0.94 and p=0.69 for the low and high interest rates, respectively). Considering the low interest rate case and the today vs. one week later decision for example, students in the IT and CT groups allocate 22% and 21% fewer tokens to the early date respectively (the mean value of the tokens allocated to the early date in PC group is 2.02 and the estimated treatment effects are -0.44 and -0.43 respectively, see Table 5). Similar results hold for the one week-two week decisions (Table 6, rows 3 to 6 in each panel): statistically significant treatment effects are estimated for both treatment groups IT and CT against the control, and the estimates are not statistically different from each other in size (Wald tests, p=0.57 and p=0.62 for the low and high interest rates, respectively).

These results not only suggest that the program generates a significant impact on children's intertemporal choices measured by incentivized tasks, replicated in two different samples, but also signal that the impact of the program does not fade after a delay of 8 months. In the next section, we explore the temporal nature of the treatment effect further.

5.3.3 Treatment Effect in Measurement Phase 3: Longer-Term Impact

The final measurement phase in our design uses a CTB task with a single decision (the tradeoff between today versus one week later, with r=0.25), conducted in an anonymous fashion to avoid any potential motives to please the experimenter or classmates. The measurements took place about a year after the program was completed in the IT group, and about six months after in the CT group. Data from this last phase can therefore help us explore the longer-term robustness of the treatment effect, up to a period of one year.

Figure 5 shows the mean early allocations in this phase across the three groups: IT, CT and PC. While control subjects allocate 2.23 tokens to today, CT and IT groups allocate 0.64 and 0.72 fewer tokens, indicating treatment effects of 30% and 32%, respectively. Both combined (IT and CT) and individual (IT or CT only) treatment effects are significant at the 1% level (see Table 7), and we again do not reject the hypothesis that the effect sizes are equal between the IT and CT groups (Wald test, p=0.34).

These results suggest that a significant treatment effect of a similar size to the original measurements still stands after up to a year from the implementation of the program. Given these results on experimental outcomes, the immediate question that follows is: has the program made any impact on actual behavior outside of experiments?

5.4 Treatment Effect on Behavioral Conduct

Given the convincing evidence from the experimental literature that behavior in experimental discounting tasks are correlated with behavioral conduct of children and adolescents (which is also borne out in our data in the control group), a natural question is whether our intervention had any impact on the behavioral conduct of our students. To answer this question, we analyze official disciplinary grades, given by the class teacher and submitted to the Ministry of Education's database by the school administration for the academic year 2013-2014. These grades are based on a 1-to-3 scale, which admittedly yields limited variation, but nevertheless are useful to estimate treatment effects. We define "low behavior grade" as a behavioral grade of 1 or 2, and as can be seen in Figure 6, about 23 percent of the students in the control group (PC) and 14 percent in both IT and CT groups received low grades at the end of the school year.[20]

Table 8 presents the marginal effects from logit regressions where the dependent variable is a dummy for the low behavioral grade. The first column gives the marginal effects with and without covariates and the last column gives associated p-values. It is quite clear from the table that the program has lowered the probability of receiving low behavioral grades by about 6 to 9 percentage points. These estimates are significant when exogenous correlates are included in the regressions, and do not reach significance when they are not.[21]

Note also that these treatment effects can be considered as "longer-term" because the behavioral grades were collected at the end of the academic year 2014, about one year after program implementation for the IT group and about 6 months after for the CT group. We do not reject the equality of the effect sizes across IT and CT (p-value=0.87).

5.5 Time Inconsistency and Heterogeneity in Treatment Effects

We now turn to the question: is there a type of student that benefits from this program more than the rest? Identifying the nature of such heterogeneity, if any, is important for both understanding the mechanisms through which training might work, and also for policy implications for populations of different composition. In order to explore this question, we first look at the two important variables we use as covariates in the treatment effect regressions: academic success and family wealth, which could have theoretical bases for creating differential effects. We do not find any treatment effect heterogeneity across students with different levels of academic success or family wealth.

Given the nature of our training, we then ask whether the program has a larger (or smaller) effect on children with existing self-control issues. In order to explore this, we use a feature of our design that allows us to elicit impatient choices before and after treatment, and the ability of our tasks to identify dynamic inconsistencies. Specifically, we compare the behavior of the CT group and a random subsample of the PC group in the CTB task, taking advantage of the fact that the CT group went through the CTB task after treatment and the MPL task before treatment, whereas the PC group did both tasks without any treatment. Classifying children in terms of their pre-treatment biases in the MPL task, and looking at their post-treatment choices in the CTB task, we are able to measure the effect of treatment on children that are, for example, present-biased.

In order to measure dynamic inconsistency in the MPL task, we compare the early choices

[20] Looking at the distribution of behavioral grades and also talking to school administrators, it is understood that given the large proportion of the highest score, any grade less than 3 indicates some behavioral issues for sure, with most teachers content with giving a 2 in these cases. This is why we combine the lower two categories.

[21] This is expected since the inclusion of correlates that are orthogonal to the treatment dummy and highly predictive of the outcome should not affect the estimated sizes but they are expected to lower the variance of the regression error and consequently the standard errors of the point estimates.

made in the today-one week and one week-two weeks decision problems. According to the standard (exponential) discounting model, individuals would be expected to choose the same switch-point in the two multiple price lists that involve the same delay of one week, regardless of when the early reward comes. Such behavior indicates dynamic consistency. We classify our subjects as "increasingly patient" (or "hyperbolic", in line with the literature) if the premium they require to wait is higher when choosing between today and one week later, than when choosing between one week and two weeks later. Such individuals are more patient in future trade-offs that do not involve the present. Finally, we classify subjects as "increasingly impatient" (or "hypobolic") if they are more patient in decisions that contain trade-offs between today and one week, than in decisions between one week and two weeks. [22]

In order to understand treatment heterogeneity based on dynamic consistency, we estimate the average treatment effects on the number of tokens allocated to the earlier date in the CTB task, conditionally on the three types of dynamic consistency/ preference reversal measured by the MPL task: exponential, hyperbolic and hypobolic. Of course, we cannot use the treatment arm of IT for this analysis, as both MPL and CTB were performed post-treatment for this treatment arm (see Table 1).

Figure 7 shows the control means and average treatment effects for the three types of children, whereas Table 9 presents the estimated average treatment effects and p-values for the three groups. The regressions reported in the table also control for pre-treatment patience measured by the MPL task. The strikingly large treatment effect comes from the students who were ex-ante measured to be present-biased (hyperbolic). Specifically, among these students, those who received the treatment allocate on average 1.6 fewer tokens to the earlier date, and this effect is statistically significant at the1 percent level. This estimate implies that treated students who were measured to be present-biased in the MPL task allocated 55 percent less gifts to the earlier date in the CTB task (baseline value for this group of students is 2.88 tokens). The direction of the treatment effect is the same for the students who were ex-ante measured to be exponential or hypobolic, but the effect sizes are much smaller (0.44 and 0.73 tokens, respectively) and point estimates do not reach statistical significance for the former. These results suggest that treatment effects come predominantly from the group of children who were initially present-biased and potentially exhibiting self-control issues, controlling for baseline patience.[23][24]

Overall, these results are very promising. A particular strength of our design is worth revisiting

[22]We use the terms "hyperbolic" and "increasingly patient" as well as "hypobolic" and "increasingly impatient" interchangeably in the paper, but do not make a claim that the preferences we identify necessarily fit these functional forms.

[23]If we look at the overall distribution of time inconsistency across treatment and control, we get the result that treated children are significantly more likely to be exponential in the MPL task (68.6% vs. 56.9%). However, this result does not hold in the CTB task (dynamic consistency is not significantly different in treatment vs. control). We therefore prefer not to make a claim about treatment effects on dynamic consistency, because of the lack of robustness across tasks.

[24]Unfortunately data limitations prevent us from carrying out the same analysis for the behavior grades, that is, analyzing whether the treatment effect on these grades come from initially present biased students. The reason for this is the combination of two problems: 1) the grades are coded on a 1 to 3 scale permitting very little variation for us to exploit and, 2) the sample with which we can do the analysis is too small.

at this point. For the two treatment groups IT and CT, we have measurements at different intervals after the program was implemented, and with different elicitation tasks. This type of design gives us very little chance of generating favorable results if a true and robust impact is not there. We not only estimate the impact of a program, but also further ask whether the treatment effects are temporally persistent and replicable across different samples in size and in statistical significance. The robustness of the treatment effects to (1) the passage of time, (2) different measurement methods, as well as the effect on a crucial real outcome, makes us confident that the observed effects are genuine and that the results would be replicable in different samples, which is important from a policy perspective.

5.6 Addressing Potential Demand Effects

One potential issue in a study like the current one is potential "demand effects", reflecting the concern that the revealed preference to delay consumption in our experiments may be coming from the desire to please or conform to the expectations of the teacher or the experimenter. If this is the case, experimental results would not be coming from actual preference shifts and would possibly be short-lived. The fact that we have a treatment effect on an important actual outcome provides strong evidence that the training has led to "real" behavioral shifts in the treated children. However, we still address potential demand concerns in terms of our experimental outcomes below.

Several measures were taken at the design stage, to address and/or reduce the extent of potential demand effects: (1) giving powerful incentives: the gifts were chosen to be extremely attractive, especially for this sample of children who come from a lower socioeconomic background, (2) the teacher was not present while the experiment was conducted, and because other measures such as risk tolerance and extensive surveys were collected, our visits were not directly linked to the training program in children's minds, (3) the 3rd measurement was conducted anonymously, making it impossible for anyone to know a particular child's decisions, and was well past the period in which these ideas were discussed in class, (4) we collected baseline data from the teachers about the propensity of each child to desire to please and conform to/obey the teacher, as well as self-reported data from the child about how important it is for her to comply with things that she feels are expected of her.

Looking at the data, the first observation to make in this regard is that with anonymity, treatment effects are still very strong and significant (see Table 7). In addition, neither teacher-assessed conformity nor self-reported compliance with expectations has any interactions with treatment effects in any task.[25] These results, coupled with the improvements in real behavioral conduct, make us quite confident that the experimental findings we document are not due to demand effects and reflect real changes in behavior.

[25] Regression results are available upon request

6 Summary and Conclusion

In this paper we report results from an educational intervention designed to improve forward-looking behavior and the ability to delay gratification in intertemporal choices in children. The program is implemented in a randomized controlled fashion, using a phase-in design that allows us to estimate causal impacts at different time intervals and across samples. The effectiveness of the program is measured in terms of children's behavior in incentivized experimental tasks as well as their official end-of-year behavioral grades.

We find that treated students make significantly more patient intertemporal choices in incentivized time preference elicitation tasks. This effect is particularly strong for the students who were identified as present-biased in the baseline. We also find that treated students are significantly less likely to receive a low behavioral grade. The estimated effects are persistent and robust across implementation phases and elicitation methods.

While our study design allows us to estimate the overall impact of the program, as in many program evaluation studies it is not possible to precisely identify which aspect(s) of this particular program generated the results we report, without assuming a structural model. In this paper, we choose not to subscribe to any particular theoretical model to guide our empirical work or to interpret our findings. We rather focus on the question of whether it is possible to achieve measureable changes in behavior through emphasizing specific concepts in the classroom environment, to which we give an affirmative answer. It is, of course, possible to design future interventions similar to ours by embedding features or experiments that would allow one to differentiate between competing mechanisms of change. We leave this for future research.

A final caveat applies to the generalizability of our results. Our estimated program effects are based on a sample of willing teachers, therefore they inform us about the impact on the students whose teachers were willing to participate in the program, and not necessarily the impact on the population as a whole. However, considering that two major reasons for non-participation was teachers' involvement in other projects and impending relocation (constraints rather than unwillingness), and that willingness to participate reaches about 85% if this is taken into account, our estimates of treatment effects on the treated are likely to be not too far from average treatment effects.

While much is now known in terms of which skills are crucial to complement academic success and behavioral conduct of students, more research is needed to explore educational channels to enhance these skills. There is no doubt that the family plays a crucial role in this respect. However, in parts of the society where families are not in the position of transferring these skills to their children for reasons typically related to poverty and illiteracy, our paper provides evidence that all is not lost, and that skills such as forward looking behavior can be developed and good behavioral habits can be made stick in classrooms by teachers who adopt and convey these ideas.

Tables

Table 1: Program Implementation and Measurement Phases

	Initial Treatment (IT)	Control-then-Treatment (CT)	Pure Control (PC
Phase 0			
Program Implementation	-	-	-
Measurement	March 2013	March 2013	March 2013
	(Baseline Data)	(Baseline Data)	(Baseline Data)
Phase 1			
Program Implementation	March-April 2013	-	-
Measurement	May 2013	May 2013	May 2013
	(Multiple price list)	(Multiple price list)	(Multiple price list, random subset)
Phase 2			
Program Implementation	-	October-November 2013	-
Measurement	December 2013	December 2013	December 2013
	(Convex time budget)	(Convex time budget)	(Convex time budget)
Phase 3			
Program Implementation	-	-	-
Measurement	May 2014	May 2014	May 2014
	Convex time budget (anonymous)	Convex time budget (anonymous)	Convex time budget (anonymous)

Table 2: Randomization Balance

Variables	IT vs CT	IT vs PC	CT vs PC
Male	-0.05	-0.07	0.04
	(0.08)	(0.11)	(0.09)
Low behavioral Score	-0.05	-0.39	0.09
	(0.25)	(0.34)	(0.38)
Academic Success	-0.07	-0.03	-0.02
	(0.10)	(0.16)	(0.15)
Family Wealth	-0.12	-0.16	0.06
	(0.17)	(0.20)	(0.22)
Cognitive Score	0.00	-0.01	-0.02
	(0.04)	(0.05)	(0.05)
Classroom Size	0.05	0.08	0.01
	(0.05)	(0.07)	(0.04)
Male Teacher	0.08	-0.31	0.14
	(0.75)	(0.71)	(0.87)
Teacher Age	0.00	0.07*	0.13*
	(0.04)	(0.04)	(0.08)
Teacher Patience	0.06	0.01	-0.09
	(0.07)	(0.05)	(0.08)
χ^2 (Simulated p-value)	6.31 (0.92)	14.4 (0.64)	6.29 (0.99)

Note: Clustered standard errors (at school level) are in parentheses. Simulated p-values are calculated via 10,000 repetitions. *: significant at 10%.

Table 3: Baseline Correlations

	avgtime	avgctb	avgtime	avgctb	avgtime	avgctb	avgtime	avgctb
Math score	-0.263**	-0.330***						
	(0.112)	(0.0704)						
Raven			-0.334**	-0.348***				
			(0.134)	(0.0609)				
Easily distracted					0.651**	0.372**		
					(0.280)	(0.149)		
Good behavior							-0.394	-0.589***
							(0.242)	(0.152)
Constant	4.129***	3.180***	3.016***	1.915***	3.018***	1.776***	3.333***	2.123***
	(0.482)	(0.306)	(0.131)	(0.0633)	(0.130)	(0.0836)	(0.192)	(0.111)
R-Squared	0.012	0.071	0.015	0.064	0.010	0.015	0.005	0.045
Number of Obs	448	290	394	481	518	411	490	324

Note: Standard errors are in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 4: Treatment Effect on Number of Early Choices (MPL)

	Control Mean	ATE	Correlates	P-value
Today vs. One Week	3.04	-0.86	NO	0.002
Today vs. One Week		-0.95	YES	0.001
One Week vs. Two Weeks	3.29	-0.84	NO	0.008
One Week vs. Two Weeks		-0.93	YES	0.009
Number of Observations	1,703			

Note: Standard errors are clustered at the school level. Correlates are teacher assessed success, family wealth, cognitive score, student gender, teacher age, teacher gender and classroom size.

Table 5: Treatment Effect on Number of Early Choices (CTB)

	Control Mean (PC)	ATE	Correlates	P-value
Today vs. One Week (r=0.25), (IT+CT vs PC)	2.02	-0.43	NO	0.005
Today vs. One Week (r=0.25), (IT+CT vs PC)		-0.38	YES	0.008
Today vs. One Week (r=0.25), (IT vs PC)		-0.44	NO	0.006
Today vs. One Week (r=0.25), (IT vs PC)		-0.40	YES	0.006
Today vs. One Week (r=0.25), (CT vs PC)		-0.43	NO	0.030
Today vs. One Week (r=0.25), (CT vs PC)		-0.33	YES	0.072
Today vs. One Week (r=0.50), (IT+CT vs PC)	1.76	-0.41	NO	0.007
Today vs. One Week (r=0.50), (IT+CT vs PC)		-0.30	YES	0.014
Today vs. One Week (r=0.50), (IT vs PC)		-0.39	NO	0.012
Today vs. One Week (r=0.50), (IT vs PC)		-0.28	YES	0.017
Today vs. One Week (r=0.50), (CT vs PC)		-0.46	NO	0.020
Today vs. One Week (r=0.50), (CT vs PC)		-0.34	YES	0.037
Number of Observations	1,846			

Note: Standard errors are clustered at the school level. Correlates are teacher assessed success, family wealth, cognitive score, student gender, teacher age, teacher gender and classroom size.

Table 6: Treatment Effect on Number of Early Choices (CTB)

	Control Mean (PC)	ATE	Correlates	P-value
One Week vs. Two Weeks (r=0.25), (IT+CT vs PC)	2.04	-0.46	NO	0.001
One Week vs. Two Weeks (r=0.25), (IT+CT vs PC)		-0.41	YES	0.006
One Week vs. Two Weeks (r=0.25), (IT vs PC)		-0.49	NO	0.001
One Week vs. Two Weeks (r=0.25), (IT vs PC)		-0.44	YES	0.002
One Week vs. Two Weeks (r=0.25), (CT vs PC)		-0.39	NO	0.046
One Week vs. Two Weeks (r=0.25), (CT vs PC)		-0.31	YES	0.098
One Week vs. Two Weeks (r=0.50), (IT+CT vs PC)	1.86	-0.45	NO	0.011
One Week vs. Two Weeks (r=0.50), (IT+CT vs PC)		-0.37	YES	0.033
One Week vs. Two Weeks (r=0.50), (IT vs PC)		-0.43	NO	0.018
One Week vs. Two Weeks (r=0.50), (IT vs PC)		-0.34	YES	0.051
One Week vs. Two Weeks (r=0.50), (CT vs PC)		-0.52	NO	0.019
One Week vs. Two Weeks (r=0.50), (CT vs PC)		-0.44	YES	0.042
Number of Observations	1,846			

Note: Standard errors are clustered at the school level. Correlates are teacher assessed success, family wealth, cognitive score, student gender, teacher age, teacher gender and classroom size.

Table 7: Treatment Effect on Number of Early Choices (CTB, anonymous)

	Control Mean (PC)	ATE	P-value
Today vs. One Week (r=0.25), (IT+CT vs PC)	2.23	-0.70	0.001
Today vs. One Week (r=0.25), (IT vs PC)		-0.72	0.001
Today vs. One Week (r=0.25), (CT vs PC)		-0.64	0.003
Number of Observations	1,826		

Note: Standard errors are clustered at the school level.

Table 8: Treatment Effect on End-of-Year Behavior Grade

	Control (PC)	Marginal Effect	Correlates	P-value
Treatment IT+CT	0.227	-0.087	NO	0.174
		-0.088**	YES	0.022
Treatment IT		-0.082	NO	0.174
		-0.079**	YES	0.017
Treatment CT		-0.068	NO	0.138
		-0.056**	YES	0.027
Number of Observations	1,740			

Note: The table presents marginal effects from logit regressions where the dependent variable is a dummy that captures a "low behavioral grade" recorded by the school administration at the end of the school year. The first column gives the baseline value, which is the proportion of students who received bad behavior grade in the control group. Standard errors are clustered at the school level.

Table 9: Treatment Effect Heterogeneity

Variables	Control (PC)	ATE	P-value
Exponential	1.86	-0.44	0.077
Hyperbolic	2.88	**-1.59**	**0.004**
Hypobolic	1.98	-0.73	0.020
Test $ATE_{hyperbolic} < ATE_{Exponential}$	**p-value=0.002**		
Test $ATE_{hyperbolic} < ATE_{Hypobolic}$	**p-value=0.004**		
Test $ATE_{hypobolic} < ATE_{Exponential}$	p-value=0.225		

Note: This table presents the heterogeneous treatment effects on the number of tokens allocated to the early date in the CTB task across types of dynamic consistency/inconsistency measured by the MPL task. Average treatment effects (ATE) are estimated by running regressions of the average number of early choices in the CTB task on treatment dummy and the number of early choices in the MPL task for each type (exponential, hyperbolic, hypobolic) seperately. Standard errors are clustered at the school level. The treatment arm IT is excluded, as the MPL task for this arm was conducted post-treatment. Given the number of tests=3, p-value for decisions is taken to be 0.05/3=**0.017.**

Figures

Figure 1

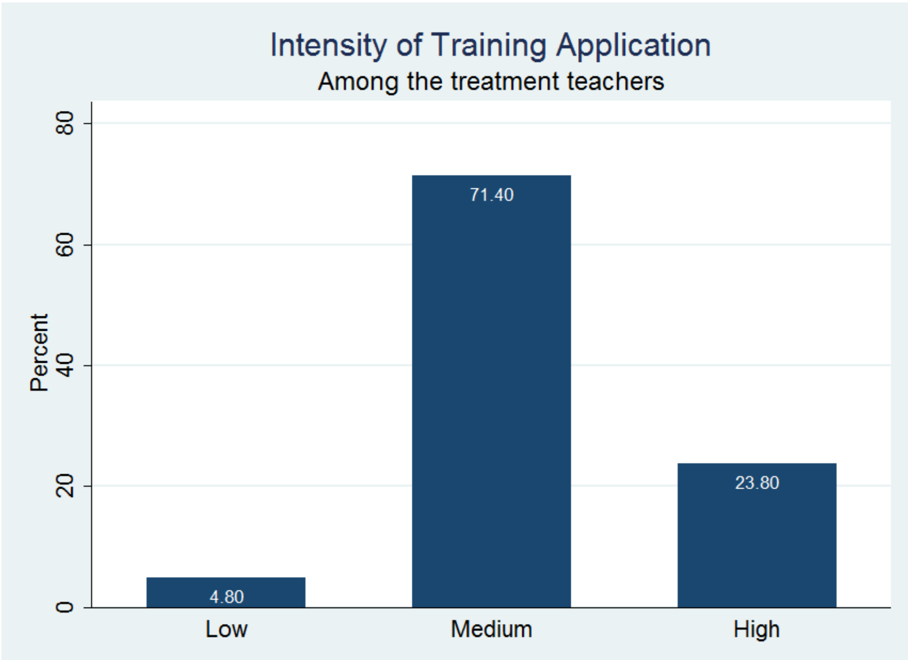


Figure 2

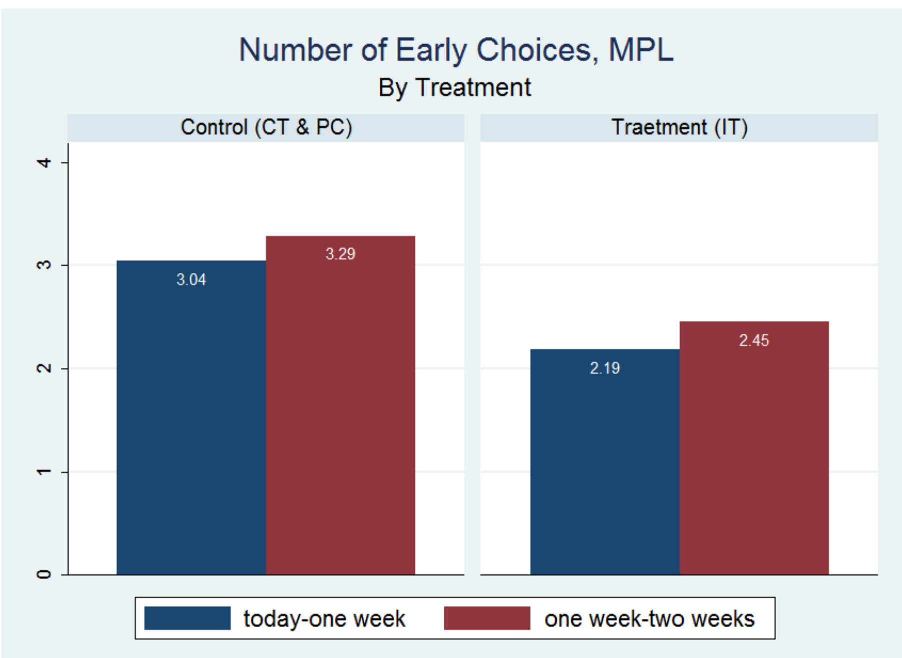


Figure 3

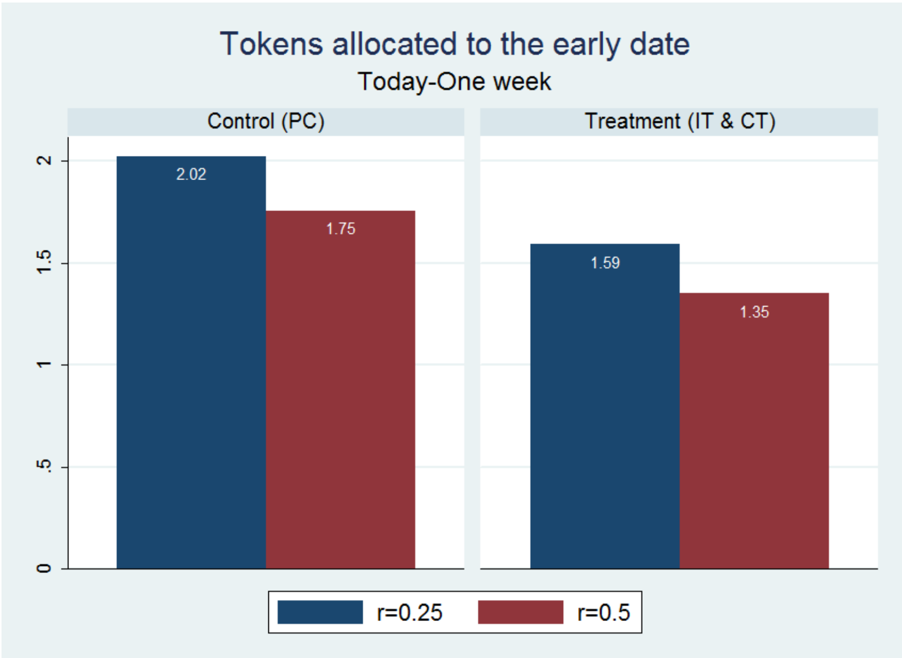


Figure 4

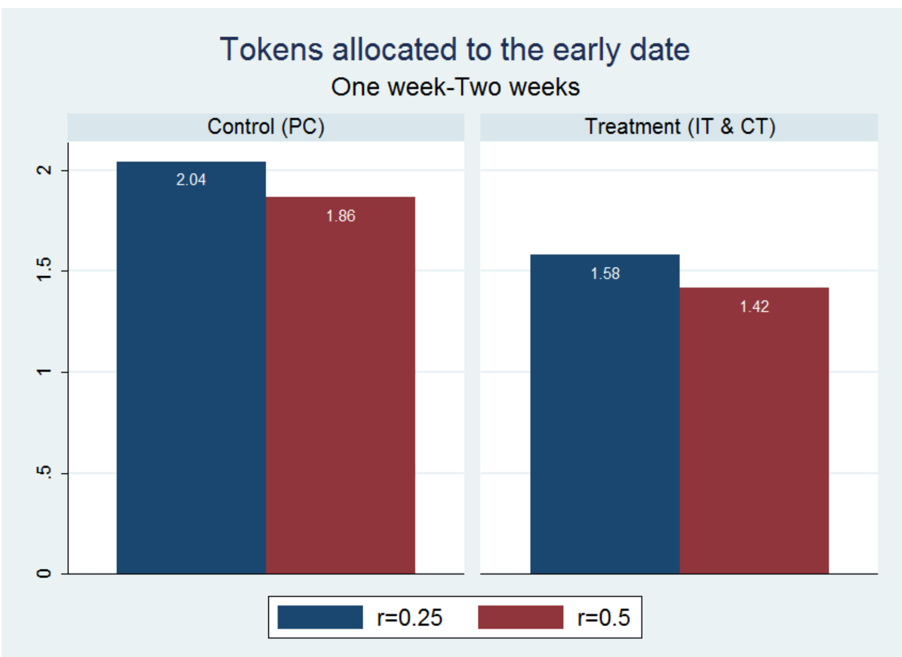


Figure 5

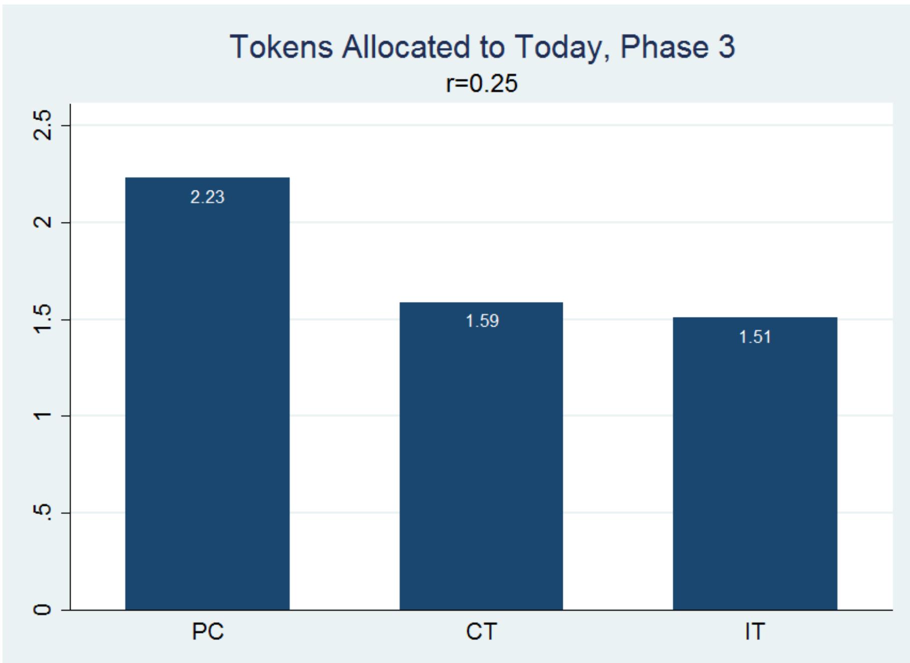


Figure 6

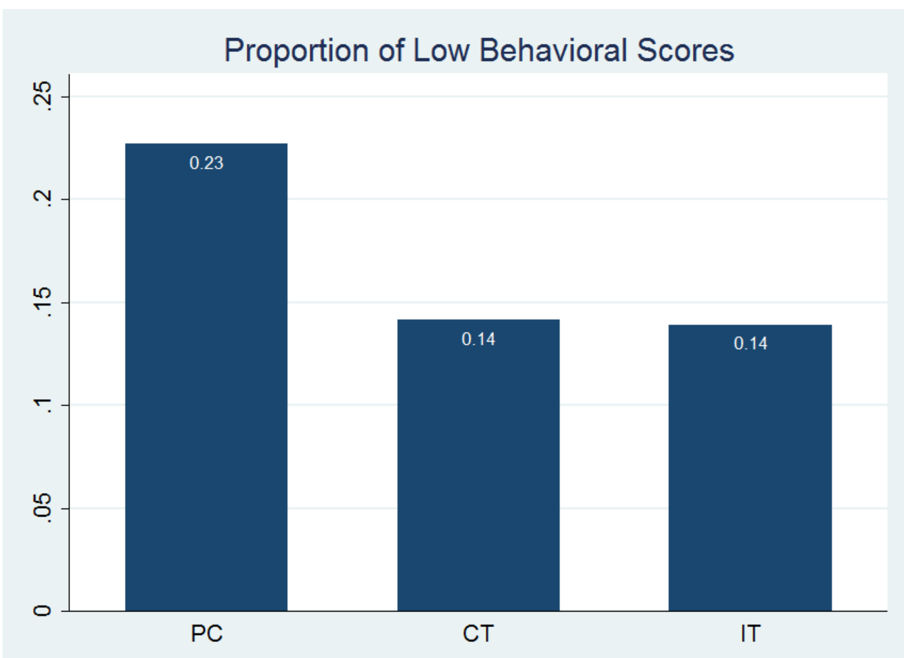
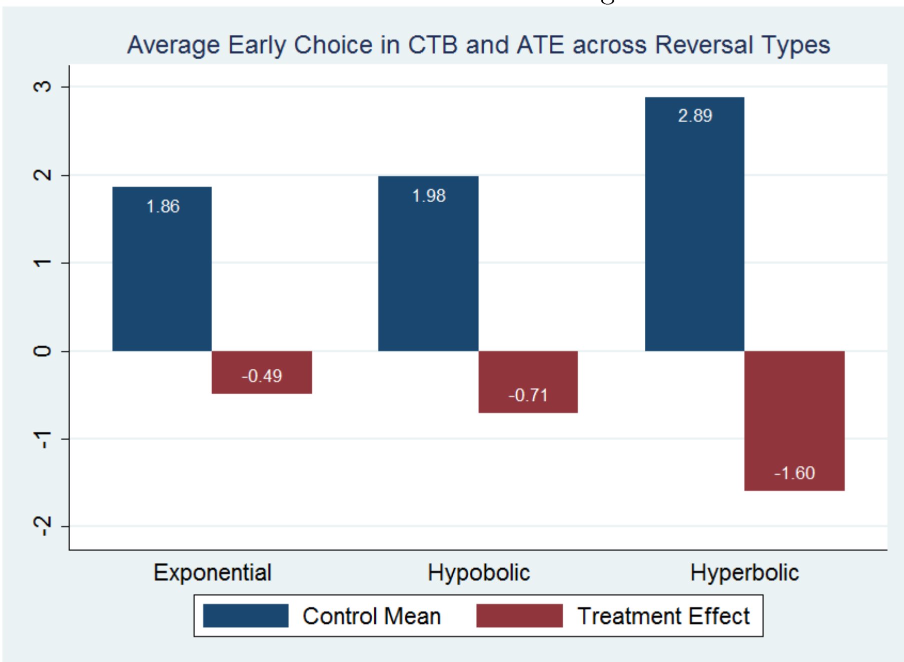


Figure 7


Average Early Choice in CTB and ATE across Reversal Types
3
2
1
0
-1
-2
1.86
-0.49
1.98
-0.71
2.89
-1.60
Exponential
Hypobolic
Hyperbolic
Control Mean
Treatment Effect

References

[1] Ameriks, J., Caplin, A., Leahy, J., & Tyler, T. , 2007. Measuring self-control problems. The American Economic Review, 966-972.

[2] Andersen, S., Harrison, G. W., Lau, M. I., & Rutström, E. E., 2006. Elicitation using multiple price list formats. Experimental Economics 9(4), 383-405.

[3] Andersen, S. G. W. Harrison, M. I. Lau, E. E. Rutstrom, 2008. Eliciting Risk and Time Preferences, Econometrica 76(3), 583-618.

[4] Andreoni, J., Sprenger, C., 2012. Estimating Time Preferences from Convex Budgets. American Economic Review 102 (7), 3333–56.

[5] Becker, G. S. and Mulligan, C. B. 1997. The endogenous determination of time preference. Quarterly Journal of Economics 112 (3), 729-758.

[6] Bettinger, E., Slonim, R., 2007. Patience among children. Journal of Public Economics 91 (1), 343–363.

[7] Bickel, W., Odum, A., Madden, G., 1999. Impulsivity and Cigarette Smoking: delay discounting in current, never and exsmokers. Psychopharmacology 146 (4), 447–454.

[8] Bodner, R., and Prelec, D. 2003. "Self-signaling and Diagnostic Utility in Everyday Decision Making," in The Psychology of Economic Decisions, vol. Rationality and Well-being , ed. by I. Brocas, and J. Carillo, Chapter 6. Oxford University Press.

[9] Borghans, L., Duckworth, A., Heckman, J.J., Ter Weel, B., 2008. The economics and psychology of personality traits. Journal of Human Resources 43 (4), 972–1059.

[10] Capaldi, D. M., Rothbart, M. K., 1992. Development and validation of an early adolescent temperament measure. Journal of Early Adolescence 12 (2), 153–173.

[11] Castillo, M., Ferraro, P. J., Jordan, J. L., Petrie, R., 2011. The today and tomorrow of kids: Time preferences and educational outcomes of children. Journal of Public Economics 95 (11), 1377–1385.

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

[13] Ellis, L. K., Rothbart, M. K., 2001. Revision of the Early Adolescent Temperament Questionnaire. Poster presented at the biennial meeting of the Society for Research in Child Development, Minneapolis, MN.

[14] Finke, M.S., Huston, S.J, 2013. Time preference and the importance of saving for retirement. Journal of Economic Behavior and Organization 89, 23–34.

[15] Fuchs, V. R., 1982. Time Preferences and Health: An Exploratory Study. Economic Aspects of Health, 93–120.

[16] Golsteyn, B. H., Grönqvist, H., Lindahl, L., 2013. Adolescent Time Preferences Predict Lifetime Outcomes. The Economic Journal.

[17] Harrison, G. W., List, J. A., 2004. Field experiments. Journal of Economic Literature, 1009-1055.

[18] Harrison, G., Lau, M., Rutstrom, E, Sullivan, M. 2005. Eliciting Risk and Time Preferences Using Field Experiments: Some Methodological Issues. In Field Experiments in Economics, Research in Experimental Economics, Vol. 10, ed. by J. Carpenter, G. Harrison, and J. List. Greenwich, CT: JAI Press, 125–218.

[19] Holt, C. A., Laury, S. K., 2002. Risk aversion and incentive effects. American Economic Review 92(5), 1644-1655.

[20] Heckman, J., Stixrud, J., Urzua, S., 2006. The Effects of Cognitive and Noncognitive Abilities on Labor and Social Behavior. Journal of Labor Economics 24 (3), 411–482.

[21] Heckman, J., Kautz, T., 2014. Fostering and Measuring Skills: Interventions That Improve Character and Cognition, in Heckman, J. J., Humphries, J. E., & Kautz, T. (Eds.). The Myth of Achievement Tests: The GED and the Role of Character in American Life., University of Chicago Press., 341-431.

[22] Jamison, J., Karlan, D., Zinman, J., 2012. Measuring Risk and Time Preferences and Their Connections with Behavior. Handbook of Experimental Economics 2.

[23] Knudsen, E. I., Heckman, J. J., Cameron, J., Shonkoff J. P., 2006. Economic, neurobiological, and behavioral perspectives on building America's future workforce. Proceedings of the National Academy of Sciences 103 (27), 10155–10162.

[24] Laibson, D. I., Repetto, A., Tobacman, J., Hall, R. E., Gale, W. G., Akerlof, G. A., 1998. Self-control and saving for retirement. Brookings Papers on Economic Activity, 91–196.

[25] Meier, S., Spenger, C., 2010. Present-Biased Preferences and Credit Card Borrowing. AEJ: Applied Economics 2, 193–210.

[26] Meier, S., Sprenger, C., 2013. Discounting Financial Literacy: Time Preferences and Participation in Financial Education Programs, Journal of Economic Behavior and Organization 95, 159-174.

[27] Mischel, W., Shoda, Y. and Rodriguez, M., 1989. Delay of gratification in children. Science, 244(4907), 281-302.

[28] Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R.J., Harrington, H., Houts, R., Poulton, R.,Roberts, B. W., Ross, S., Sears, N. R., Thomsom, W. M., Caspi, A., 2011. A gradient of childhood self-control predicts health, wealth, and public safety. Proceedings of the National Academy of Sciences 108 (7), 2693–98.

[29] Perez-Arce, F., 2011. The effect of education on time preferences. Rand Labor and Population Series, WR-844.

[30] Raven, J.C., Court, J.H., and Raven, J., 1986. Raven's Progressive Matrices and Raven's Coloured Matrices. London: H.K. Lewis.

[31] Sutter, M., Kocher, M.G., Ruetzler, D., Trautmann, S.T., 2013. Impatience and uncertainty: Experimental decisions predict adolescents' field behavior. American Economic Review 103 (1), 510–531.

[32] Voors, M. J., Nillesen, E. E. M., Bulte, E. H., Lensink, B. W., Verwimp, P. and Soest, D. P. van (2012). 'Violent conflict and behavior: a field experiment in Burundi', American Economic Review, vol. 102(2), pp. 941-64.

Appendix

Multiple Price List Task

Children are asked to make a decision for each row in each list. Lists are presented separately.

Today versus One Week

◯ 2 Gifts Today	◯ 2 Gifts in One Week
◯ 2 Gifts Today	◯ 3 Gifts in One Week
◯ 2 Gifts Today	◯ 4 Gifts in One Week
◯ 2 Gifts Today	◯ 5 Gifts in One Week
◯ 2 Gifts Today	◯ 6 Gifts in One Week
◯ 2 Gifts Today	◯ 7 Gifts in One Week
◯ 2 Gifts Today	◯ 8 Gifts in One Week
◯ 2 Gifts Today	◯ 9 Gifts in One Week
◯ 2 Gifts Today	◯ 10 Gifts in One Week

One Week versus Two Weeks

◯ 2 Gifts in One Week	◯ 2 Gifts in Two Weeks
◯ 2 Gifts in One Week	◯ 3 Gifts in Two Weeks
◯ 2 Gifts in One Week	◯ 4 Gifts in Two Weeks
◯ 2 Gifts in One Week	◯ 5 Gifts in Two Weeks
◯ 2 Gifts in One Week	◯ 6 Gifts in Two Weeks
◯ 2 Gifts in One Week	◯ 7 Gifts in Two Weeks
◯ 2 Gifts in One Week	◯ 8 Gifts in Two Weeks
◯ 2 Gifts in One Week	◯ 9 Gifts in Two Weeks
◯ 2 Gifts in One Week	◯ 10 Gifts in Two Weeks

Convex Time Budget Task

Children are asked to choose one of the 6 decisions, in each column (presented as separate decision sheets).

$r = 0.25$,Today vs. One Week	$r = 0.25$,One Week vs. Two Weeks
◯ 5 Gifts Today and 0 Gifts in One Week	◯ 5 Gifts in One Week and 0 Gifts in Two Weeks
◯ 4 Gifts Today and 1.25 Gifts in One Week	◯ 4 Gifts in One Week and 1.25 Gifts in Two Weeks
◯ 3 Gifts Today and 2.5 Gifts in One Week	◯ 3 Gifts in One Week and 2.5 Gifts in Two Weeks
◯ 2 Gifts Today and 3.75 Gifts in One Week	◯ 2 Gifts in One Week and 3.75 Gifts in Two Weeks
◯ 1 Gifts Today and 5 Gifts in One Week	◯ 1 Gifts in One Week and 5 Gifts in Two Weeks
◯ 0 Gifts Today and 6.25 Gifts in One Week	◯ 0 Gifts in One Week and 6.25 Gifts in Two Weeks

$r = 0.50$,Today vs. One Week	$r = 0.50$,One Week vs. Two Weeks
◯ 5 Gifts Today and 0 Gifts in One Week	◯ 5 Gifts in One Week and 0 Gifts in Two Weeks
◯ 4 Gifts Today and 1.5 Gifts in One Week	◯ 4 Gifts in One Week and 1.5 Gifts in Two Weeks
◯ 3 Gifts Today and 3 Gifts in One Week	◯ 3 Gifts in One Week and 3 Gifts in Two Weeks
◯ 2 Gifts Today and 4.5 Gifts in One Week	◯ 2 Gifts in One Week and 4.5 Gifts in Two Weeks
◯ 1 Gifts Today and 6 Gifts in One Week	◯ 1 Gifts in One Week and 6 Gifts in Two Weeks
◯ 0 Gifts Today and 7.5 Gifts in One Week	◯ 0 Gifts in One Week and 7.5 Gifts in Two Weeks