The Effects of Chess Instruction on Academic and Non-Cognitive Outcomes: Field Experimental Evidence from a Developing Country*

Asad Islam†
Monash University

Wang-Sheng Lee

Deakin University and IZA, Bonn

Aaron Nicholas
Deakin University

This version: August 2019

Project implemented by Global Development And Research Initiative (GDRI)

Abstract

We conduct a randomized field experiment to investigate the benefits of an intensive chess training program undertaken by primary school students in a developing country context. We examine the effects on academic outcomes, and a number of non-cognitive outcomes: risk preferences, patience, creativity and attention/focus. Our main finding is that chess training reduces the level of risk aversion almost a year after the intervention ended. We also find that chess training improves math scores, reduces the incidence of time inconsistency and the incidence of non-monotonic time preferences. However, these (non-risk preference) results are less conclusive once we account for multiple hypothesis testing. We do not find any evidence of significant effects of chess training on other academic outcomes, creativity, and attention/focus.

JEL codes: C93, D80, I21

Keywords: chess training; math; non-cognitive outcomes; risk; randomized experiment

* We thank participants at the Labour Econometrics Workshop 2017 in Auckland, the 2017 Asian and Australasian Society of Labour Economics (AASLE) conference in Canberra, the 17th IZA-SOLE Transatlantic Meeting of Labor Economists (2018) in Buch/Ammersee, and workshops at Deakin University, Monash University and Singapore Management University for their comments. Foez Mojumder provided excellent research support. This research would also have not been possible without cooperation from the Department of Primary Education (DPE) in Bangladesh, local school teachers, NGO partner Global Development Research Initiative (GDRI) and our chess coaches. We are grateful to Deakin University and Monash University for their generous research funding.
† Corresponding author. Department of Economics, Monash University, PO Box 197, Caulfield East, Victoria 3145, Australia; Tel: +61 3 99032783; Fax: +61 3 99031128; Email: Asadul.Islam@monash.edu

1. Introduction

Chess as an education tool is becoming increasingly common in many countries. Its popularity is at least in part attributable to its perceived effect on cognitive skills in general, and math ability in particular. In recent years, chess coaching for children has become increasingly popular in developed countries.[1] The European Parliament has expressed a favorable opinion on using chess courses in schools as an educational tool (Binev et al., 2011). In 2014, School Library Journal's best education pick of the year was a chess-related product called Yamie Chess, which is backed by Harvard and MIT academics.[2] The benefits of playing chess regularly have been suggested in a documentary that focuses on an inner-city school in New York, and two European countries – Armenia and Poland – have even made chess instruction compulsory in their primary-school curricula.[3] More recently, the city of Bremen in Germany has decided to introduce one hour of chess per week as a subject in primary schools in 2020, an issue covered widely in the German press.[4]

Parents and teachers generally view chess as a highly regarded extracurricular activity in primary school. However, to date, there is hardly any study rigorously examining the effects of chess instruction. An exception is Jerrim et al. (2018), who report results from a randomized controlled trial (RCT) conducted in the UK to evaluate the impact of teaching children chess on academic outcomes. Contrary to popular belief, they found no evidence that teaching children chess improved their math ability. There were also no impacts on reading and science.

In this paper, we conduct an RCT to examine the effects of intensive chess lessons among grade five students in a developing country. We follow the curriculum approved by the World Chess Federation. We differ from Jerrim et al. (2018) and the literature on the impact of chess training on two counts. First, we study the link between chess and non-cognitive outcomes such as risk preferences, patience, creativity, attention and focus. Second, we examine the effects of chess learning in a developing country context. Children in our

[1] For example, in the US, the Chess Club and Scholastic Center of St. Louis (a 6,000-square-foot, state-of-the-art chess center widely recognized as the premier chess facility in the country and one of the best in the world) helps provide chess-coaching services to many elementary and middle schools in St. Louis, Missouri. For the list of schools, see: https://saintlouischessclub.org/education/partners-education (accessed March 27, 2017).

[2] Yamie Chess features an interactive coloring math comic book written by experienced math teachers for K-8 supplemental math learning.

[3] The documentary film Brooklyn Castle (2012) highlights the after-school chess program in an inner-city public school in Brooklyn, New York and how they became the first middle-school team to win the U.S. Chess Federation's national high school championship. For compulsory chess instruction in Armenia, see: https://www.theguardian.com/world/2011/nov/15/armenia-chess-compulsory-schools (accessed March 27, 2017). For Poland, see: http://cis.fide.com/en/chess-news/325-poland-chess-in-all-schools (accessed March 27, 2017).

[4] See https://en.chessbase.com/post/chess-makes-smart-scholastic-tournament-in-bremen-2019 (accessed 2 July, 2019).

experiment come from rural primary schools in Bangladesh who do not have previous experience playing chess. Our setting is particularly well-suited to test the benefits of a chess training program because unlike children in urban areas in a developed country, most children in rural areas in a developing country will never have been exposed to the game of chess before, much less any other cognitively demanding games.[5]

We first examine the effects of chess training on test scores. Our primary outcomes for test scores come from a standardized, compulsory public exam that all fifth-grade students in Bangladesh must take – the Primary School Certificate (PSC) exam – which took place 9-10 months after the completion of chess training. While we are particularly interested in examining the effects on math test scores because of the perceived math benefits from playing chess, we also examine the results for students' first language and science.[6]

Chess is often regarded as a game reflecting real life and teaching children how to play chess in a prescribed systematic fashion might also help in their development of important non-cognitive outcomes. Therefore, we pay particular attention to the collection of extensive data on non-cognitive outcomes to examine the effects of chess training. In particular, we measure risk preferences, patience, creativity and attention/focus.

Chess, through the formation of strategies, can be useful for the conceptualization and calculation of risks.[7] For example, chess players often sacrifice pawns, bishops, knights, rooks, or queens if it helps checkmate the opponent's king and win the game. Such sacrifices are inherently risky because if one's calculations are faulty, the sacrifice could prove to be fatal, eventually leading to a quick loss of the game. Gambits and sacrifices can be made during any of the three phases of a chess game – opening, middlegame, or endgame. Such an association between risk taking and chess playing is, for example, utilized to study the link between risk preferences and attractiveness (Dreber et al., 2013) through behaviour in chess.[8] Thus, learning

---

[5] Jerrim et al. (2016, p. 46) report in their study that chess playing activity at their baseline was 48% in treatment schools and 45% in control schools. Such levels are not surprising given that their study was based in an urban developed country setting.

[6] Studies in the education literature (e.g., Scholz et al. 2008; Trinchero and Sala 2016) also suggest that chess improves children's math skills because the game has some elements in common with the mathematical domain and because it promotes suitable habits of mind.

[7] Risk aversion is a trait typically associated with welfare-relevant, later life outcomes. Hence, its detection (and potential manipulation) from an early age may be of particular policy interest. Davis and Eppler-Wolff (2009) argue that parents need to understand the significance of risk-taking as a teaching experience for children. Higher risk aversion has been shown to be detrimental to key household decisions, such as choice of occupation, portfolio selection and moving decisions (Guiso and Paiella, 2008). On the other hand, higher risk aversion has also been linked to less disciplinary referrals and a higher probability of high school completion (Castillo, Jordan and Petrie, 2018).

[8] There, risk taking in chess is measured by exploiting a standardized classification of opening moves and expert assessments. As chess players in our setting are beginners who are unlikely to have a well thought out opening

how to play chess and gaining an appreciation of basic chess strategy can help in the development and articulation of risk preferences in children. Of course, being able to calculate and appreciate risks may either increase or decrease risk aversion: the risk hypothesis we test is therefore two-sided. In addition, chess playing may decrease risk aversion through increased exposure to competition. Experimental studies by Eriksen and Kvaløy (2017) and Spadoni and Potters (2018), for example, provide evidence that an increase in competitive pressure decreases risk aversion.

Furthermore, chess might help teach children to be more patient, more focused, and have more self-control.[9] It can potentially motivate children to become willing problem-solvers, able to spend hours quietly immersed in logical thinking. Chess can also be a useful tool to teach the importance of forward-looking behavior. An important element in chess is the evaluation process, i.e., one needs to look a few steps ahead during a chess game and consider and evaluate alternative scenarios. Chess can teach children how to focus and visualize by imagining a sequence of events before it happens. The schematic thinking approach in chess resembles trees and branches in sequential-decision analysis and might also be useful and possibly transferable to math skills, as has been emphasized previously (Scholz et al., 2008; Trinchero and Sala, 2016).

In addition to children's risk preferences and time preferences, we also investigate whether undertaking intensive chess lessons can affect children's creativity and attention/focus. Although there is some debate over whether creativity is an aspect of intelligence or a personality trait, several studies have shown that creativity can be experimentally manipulated (see Runco and Sakamoto, 1999, for a review). The ability to focus on a task at hand is also a useful non-cognitive outcome that chess might be able to nurture. Attention is considered to be a major part of working memory, responsible for the control of flow of information, switching between tasks and selection of relevant stimuli and inhibition of irrelevant ones (Travis, 1998). The study of the development of attention occupies a central place in cognitive developmental psychology, and we use frequently used tests for focus/attention in our evaluation.

---

repertoire (a regular set of openings they use to start the game), it is not possible to adopt such an approach to measure risk preferences.

[9] Becker and Mulligan (1997) suggest that observed differences in time preferences are not innate and that the evolution of these preferences may be endogenous. This implies that children could be taught to be more forward thinking. If patience and other time preference-related characteristics of children vary across gender or demographic groups, different educational paths and career outcomes may occur. For example, Castillo et al. (2011) find that boys are more impatient than girls, and that impatience has a direct correlation with disciplinary referrals – behavior that has been shown to be predictive of economic success.

This paper is relevant to several sub-fields of economics. First, there has been much recent interest in the development of non-cognitive skills in children and their importance in later life outcomes in the economics literature. Non-cognitive skills have been shown to be very important for a host of outcomes, including schooling, social behaviors, drugs, smoking, truancy, teenage pregnancy, involvement in crime, and labor market success (Heckman et al., 2006; Carneiro et al., 2007). In addition, although a large literature in experimental economics has focused on the role of risk preferences in explaining life outcomes (e.g. Dohmen et al. 2011; Sutter et al. 2013), surprisingly little is known about differences in risk preferences at an early age and how these preferences are developed, or how they may alter the life paths of students (Andreoni et al., 2019). Chess may be of particular interest to policymakers who are interested in identifying programs that can provide early stimulation and help develop such important "soft" life skills in children during their formative years. Second, in the program evaluation literature, there is increasing interest in evaluating interventions that have the potential to be scaled up (Banerjee et al., 2017). Given resource and institutional constraints, the effectiveness of scalable interventions that can be deployed which can form the basis of public policy is to date not well explored. As introducing chess as a subject in school will not be very costly, the educational intervention we examine in this paper most certainly has the potential to be scaled up if smaller proof-of-concept studies such as this paper show positive results. Indeed, some countries like Armenia and Poland and cities like Bremen in Germany have already made the decision to scale up despite scant rigorous experimental evidence on the effects of chess instruction.

Overall, the main finding in our paper is that chess training has a significant effect on reducing the level of risk aversion almost a year later. Based on conventional p-values and wild bootstrap p-values, we also find that chess training has a positive impact on math scores in the national exam and reduces the incidence of both time inconsistency and monotonic time preferences. However, the results are less conclusive once we account for multiple hypothesis testing using the false discovery rate (FDR). Effects of chess training on the other academic outcomes, creativity, and attention/focus were not statistically significant.

The paper is structured as follows. Section 2 briefly discusses how chess can translate to learning outcomes. Section 3 provides information on the intervention. Section 4 describes the data and the academic and non-cognitive outcomes measured in this study. Section 5 presents the results of the intervention. Section 6 concludes.

2. Chess and Learning Outcomes

Transfer of learning occurs when a set of skills acquired in one domain generalizes to other domains or improves general cognitive abilities. Little is known about the extent to which chess skills transfer to other domains of learning. Although near transfer (i.e., transfer that occurs between closely related domains, such as math and physics) might be possible, several studies have shown that chess players' skills tend to be context-bound, suggesting that it is difficult to achieve far transfer from chess to other domains. For example, it has been found that memory for chess positions fails to transfer from chess to digits both in adults and children (Schneider et al., 1993), and that chess players' perceptual skills do not transfer to visual memory of shapes (Waters et al., 2002). In the Tower of London task, a well-known test for executive functioning in which participants solve 16 four-, five-, and six-move problems each, chess planning skills did not improve the ability of chess players to solve these tasks (Unterrainer et al., 2011). Levitt et al. (2011) find that that the ability to transfer backward induction prowess from the chess board to experimental games is quite sensitive to the particulars of the game in question.

We are not aware of any studies that have explored in depth the link between chess skills and non-cognitive outcomes, although some previous work has focused on the effects of chess on focused attention and metacognition (Scholz et al., 2008), despite an observation made more than two centuries ago from a notable chess enthusiast. The renowned inventor and U.S. founding father Benjamin Franklin wrote the following in a magazine essay, "The Morals of Chess" (1786):

> "The game of chess is not merely an idle amusement. Several very valuable qualities of the mind, useful in the course of human life, are to be acquired or strengthened by it, so as to become habits, ready on all occasions. For life is a kind of chess, in which we have often points to gain, and competitors or adversaries to contend with, and in which there is a vast variety of good and ill events, that are, in some degree, the effects of prudence or the want of it."

Franklin goes on to suggest in his essay that by playing chess, one may learn foresight (considering consequences before taking action, i.e., planning chess moves), circumspection (seeing the big picture, i.e., surveying the whole chess board, the relations among pieces and situations, and the dangers the pieces are exposed to) and caution (not to make moves too hastily and to abide by all the consequences of one's rashness). Circumspection implies that a person thinks carefully before doing or saying anything, a quality that is expected to be correlated with patience. Combining foresight and caution implies a person will learn to take calculated risks, thereby linking chess playing style and skill with risk preferences.

3. The Program and the Data

3.1 The Chess Intervention

The intervention took place in primary schools in rural communities in two districts-Khulna and Satkhira—in southwest Bangladesh in January-February 2016. Our chess experiment is a clustered randomized controlled trial with randomization at the school level involving fifth grade students (10 years old on average) in 2016 in 16 primary schools.[10] These schools were chosen randomly from a set of more than 200 schools in those regions. The sampling frame included all schools in the sub-districts where both treatment and control schools were located.[11] The location of the 16 treatment and control schools can be seen in Figure 1. In general, the treatment schools and control schools were geographically spread out such that no two schools (either treatment or control) are close to each other, with each of them at least 5 km apart. In the context of rural Bangladesh where walking is the predominant mode of transport and where children tend to play with their neighbors, such distance between schools effectively means that program spillovers to control schools is very unlikely.

The schools were randomly divided into two groups: eight in the treatment group and eight in the control group.[12] Students in the treatment schools received 12 days of chess training (spread over three weeks). A pre-program baseline test of chess knowledge suggests that most children in our analysis sample did not know how to play chess. The chess knowledge test comprised a series of four questions. The first question asked: "Do you know how to play chess?" Children who responded "Yes" or "A little bit" were further probed with further specific questions about "which is the most powerful piece on the chess board" and how chess pieces move and capture in two chess positions that were provided in diagrams. Only one child answered all three questions on basic knowledge of the chess rules correctly, and 4.22% in the control group and 2.75% in the treatment group answered at least two out of the three questions correctly. This latter difference was not statistically significant (p-value = 0.514). Training sessions were conducted separately at each school at the beginning of the academic year in January-February of 2016. The chess instruction involved teaching the rules of chess and basic chess strategy.

---

[10] Computer randomization of schools was implemented using a pre-specified seed.

[11] One of the co-authors (Islam) spent his childhood and attended primary and secondary school in that area. The schools are fairly typical of many parts of rural Bangladesh. The area was chosen because of the author's local knowledge and contacts at the schools and among district-level administrators, who helped facilitate logistics for implementing the intervention.

[12] During the study's design phase, while randomization at the class level was considered and deemed preferable, it was ruled out for several reasons. First, there is the possibility of contamination between treatment and control group classes. For instance, when one class is receiving the intervention, students from other classes might want to join in. Second, most schools in rural Bangladesh only have one class of students for each grade.

The lesson plan was based on free instructional chess materials available from the Chess in Schools Commission of the World Chess Federation (FIDE) (see Appendix 1 for the syllabus used for the chess lessons). This lesson plan was developed by chess experts specifically for use as course material in primary schools. We hired two instructors to deliver the entire chess program to the eight treatment schools.[13] Both instructors are qualified chess coaches and have extensive experience teaching chess to children. One is a FIDE master and former national champion of Bangladesh, and the other is a seven-time divisional champion and a chess coach by profession. They both also have formally been appointed as trainers by the National Chess Federation in Bangladesh.

The 12-day training program for students in all the treatment schools was spread over three weeks and conducted during regular school hours. The program was first implemented in four treatment schools during three weeks in January 2016, with a further four treatment schools getting exposure to the program in the subsequent three weeks. In the first week of training (three days of training), each instructor conducted one session per day at 8:00 am in the morning. In the second week of training (five days of training), each instructor conducted two sessions per day with the first session at 8:00 am in the morning and the second session at 12:00 pm in the afternoon. In the third week of training (four days of training), each instructor continued to conduct two sessions per day with the first session at 8:00 am in the morning and the second session at 12:00 pm in the afternoon.

After the two-hour chess lesson for each day was completed, students were allowed to practice chess by playing against each other for an additional 30 minutes. To carry out the practice sessions, each instructor was supported by several field staff who are amateur chess enthusiasts. During the training sessions each pair of students received a chess set to use in class.[14] The intervention involved providing a total of 24 hours of chess instruction (daily two-hour lessons spread over 12 days) and about six hours of supervised chess practice playing against an opponent, which allowed the students to apply any new skills they had just learned. Thus, the students received approximately 30 hours of chess training – above the 25 hours Sala and Gobet (2016) report as the threshold above which chess instruction produces substantial effects.

[13] One of the co-authors of the paper (Lee) is also a national master in chess and helped ensure the suitability of the syllabus for the intervention.
[14] Some pictures of the field setting can be found in Appendix 2, in which normal classrooms have been used to conduct the chess lessons. Some schools have double shifts, where fifth-grade students start classes in the afternoon. We scheduled chess lessons to start later in these schools.

In general, there was little or no disruption to normal academic activities in both the treatment and control schools due to either the program or our elicitation of outcomes from the survey instruments. This was possible due to several factors. First, the school curriculum during the start of the school year (January and February) is relatively light, as contact time with students at the beginning and at the end of the school year is usually dominated by administrative and non-teaching activities. This includes organizing the demanding logistics of registering students, receiving and distributing teaching materials, and understanding new government policies or programs. Throughout January, as part of the annual National Education Week (a government information campaign designed to encourage parents to enroll their children in school), teachers are expected to recruit students by making visits to homes, markets, and other public places to meet parents.

Second, unlike primary schools in developed countries or in urban settings, effective instructional time in rural primary schools in Bangladesh is relatively short (Tietjen, Rahman, and Spaulding, 2004; Islam 2019). There are several contributing factors: (i) Teacher absenteeism is a major issue in rural Bangladesh[15]; (ii) Instructional time at rural schools is further reduced by the effective hours of operation. Even if teachers at rural schools are present, they were more likely to arrive late for school or depart before the official end of the school day than their urban counterparts because of domestic chores (predominately female) and income-generating activities (all males). As a result, Tietjen, Rahman and Spaulding (2004) found that teaching or "instruction" occupied on average 63 percent of the class time in the classes they observed.

Further, given the frequent later than official school start times in rural primary schools, the scheduling of our classes before the start of school day minimized the displacement of day-to-day academic studies. Hence, to the extent that any displacement occurs, the chess training program is most likely displacing idle class time or unstructured play activities that the students in the control group were observed playing, such as Ekka-dokka (hopscotch), Gulikhela (game of marbles), Ha-du-du (game of tag), and Kanamachi (a game where a blindfolded participant tries to catch other players).

Student feedback on the chess lessons was very positive. Of the 248 students (out of 294) respondents in the treatment group who provided feedback on the chess lessons, *all* of them said they liked playing chess, and 99.2% said they would like more chess lessons. In

[15] For example, Chaudhury et al. (2006) find that 16 percent of teachers are absent on a given school day, and 23.5 percent were absent once out of two visits in a school.

addition, 94.5% of the children said that during Week 1, they played or discussed chess with at least one classmate outside the chess program; the percentage remained high in Week 2 (87.5%). The chess sets used in the training program were donated to each respective school at the end of the three-week training program so that the children could continue playing and practising chess after lessons had ended. The students' interest in chess does not appear to be transitory. When we checked to see whether treatment-group members were still playing chess 9-10 months later, we found that 94.3% of them had played chess with a classmate during the previous week, and 87.5% of them had played chess with other friends or relatives during the previous week.

Before the chess training program launched, a household survey was carried out in November and December 2015 to collect some basic household information, including demographic profiles of the children and their parents. The respondents were parents of the children participating in the chess experiment. We also tested their pre-program math skills and chess knowledge. At the end of the chess training program, we conducted tests on risk preferences, time preferences, creativity, and math skills. The risk and time preference tests were incentivized as per standard practice in experimental economics.

Figure 2 describes the project's key timelines. Short-run outcomes (Wave 1) were measured at the end of the three-week chess training program (the day after), and longer-term outcomes (Wave 2) were measured about 9-10 months after training ended – at the end of October 2016. We also assessed whether the program had an impact on academic performance based on results from a national exam that fifth-grade students had to take during November 20-27, 2016.

3.2 Sample and Baseline Balance

Based on the name list of students provided by the treatment and control schools, 704 families were approached in November and December 2015 in order to collect baseline data for the experiment. The response rate to the parent questionnaire was 594/703 = 84.4%, and a complete set of non-missing covariates were obtained for 281 treatment group members and 288 control group members (n = 569) after accounting for item non-response.

Table 1 presents the differences in means of parental and household characteristics for the treatment and control groups. There are no significant differences between treatment and control groups except for the variable indicating whether the mother is a housewife. The results suggest that the randomization process was well implemented.

The children in our sample are mostly underprivileged, with parents from relatively low socio-economic backgrounds. Approximately a third of parents did not complete primary school. In more than 86% of families, no members of the household have an education higher than 10th grade. About 64% of fathers are engaged in agriculture or day labor, another 29% work in small business activities, and 6% work in services. Almost all the mothers are housewives. The average household size is 4.4, and the monthly income is less than 8,500 takas (about US $110).

The sample sizes in our regression adjusted impacts for Wave 1 presented in Tables 2, 4 and 5 are smaller than the baseline sample in Table 1. For example, the sample size for the risk preferences using Wave 1 when we regression adjust controlling for parental and household characteristics is 450/569, which is 79.1% of the grade 4 sample. The main reason for the reduction in sample from baseline to Wave 1 is students dropping out between grades 4 and 5. Note that data from the parent questionnaire was collected at the end of academic year when the students were in grade 4. However, the experiment was conducted when students progressed to the next grade at the start of the following year. Many of these students dropped out from school or could not progress to grade 5. Hence, there was some attrition from our initial baseline sample which happened before our experiment actually started.[16] In addition, a discrepancy in sample size arises when we do and do not use regression adjustment to control for parental and household characteristics as the former requires information from the parent questionnaire, which is not available for all families.[17]

High student absenteeism from schools is a big problem in Bangladesh, with more than a quarter of children aged 7-14 years missing at least one day of school in a six-day school week in the rural areas of Bangladesh (Kumar and Saqib, 2017). Tietjen, Rahman and Spaulding (2004) found based on surprise visits to government primary schools in Bangladesh that the actual percentage of students enrolled who were in attendance on the day of the visit ranged from 43 percent to 67 percent. This explains the variation in sample sizes for the various outcomes we examine.[18] As many outcomes were collected on different school days, whether an outcome was measured largely depended on whether a student attended school that day. In

[16] Ahmed at al. (2007, p.12) report using administrative data that promotion rates in primary schools in Bangladesh have been largely stable over time, and were between 75-83% for promotion from grade 4 to 5 in 1998-2004. Students need to sit for the PSC exam at the end of grade 5, and the pass rate in this exam is used to evaluate the teachers' performance. Hence, teachers try to not promote students whom they think might fail the PSC exam.

[17] This is why the regression unadjusted sample is larger than the regression adjusted sample in Tables 2 to 5.

[18] Student absenteeism is common in many developing countries – Banerjee et al., (2007) in India for the Balsakhi Program administered by Pratham, and Duflo et al. (2011) on the tracking of students in Kenya found nearly 20% of children were absent on test days. The absenteeism rate in our sample is similar to Islam (2019) who studied schools in the same region as the present study.

general, however, this attrition did not pose a problem for the integrity of the experimental design. As Tables C.1 and C.2 in Appendix 3 show, there were no significant differences in characteristics between the treatment and control groups in any of the samples examined. This is also true when comparing the sample for which we have data on outcomes with those from the initial sample that we have lost to attrition. This suggests that attrition in our sample is not systematically related to any particular set of characteristics and is likely to be unrelated to the process of randomization.

4. Outcomes

4.1 Academic Outcomes

We use exam marks from the Primary School Certificate (PSC), administered nationwide annually in Bangladesh to all fifth-grade students as the primary outcome for cognitive abilities. The PSC is a written exam, administered face-to-face and delivered through paper-and-pencil tests at the end of fifth grade. This exam took place in November 2016, approximately 9-10 months after the conclusion of the chess program. The PSC comprises six mandatory subjects: Bengali, English, science, social science, math, and religion. In the experiment, we focus on examining their results for mathematics, students' first language and science (as in Jerrim et al., 2016, 2018).

The test items consist of multiple-choice questions with three or more response options, open-ended questions requiring short constructed responses, and essay writing. Student performance is reported by percentage of points scored out of the maximum possible score. The maximum possible score is 600 points (100 points for each subject). The minimum requirement to meet the national standard is 33%.[19]

As we had a particular interest in the potential links between chess and math, two separate math tests were developed to measure students' math skills before and after the chess training sessions. The tests intended to assess problem-solving capacities in math, requiring students to use application and reasoning skills. Both tests included 11 questions to be completed in one hour. The tests contained two types of items: multiple-choice questions and constructed responses (demonstrating computing ability by solving word problems). To develop the tests, the local math textbook for fourth-grade students in Bangladesh was

[19] Due to privacy reasons, we were unable to access the numerical scores awarded to every student for each of the exams taken. However, we were able to obtain the letter grades awarded to every student for each of the six subjects, as well as an overall grade point average (GPA) score. The conversion from letter grades to scores used in Bangladesh primary schools is as follows: A+ = 5 points; A = 4 points; A- = 3.5 points; B = 3 points; C = 2 points; D = 1 point; and F = 0 points.

consulted, as were local school teachers and educators to help develop the test. The tests were conducted to assess students' content and cognitive domains. Content domains include addition, subtraction, multiplication, division (including money and product transactions), fractions, geometric skills, and reading, comparing and interpreting graphical representations of data. As our analysis sample comprised students from rural areas, with students generally coming from poorer socio-economic backgrounds with lower academic knowledge bases than their urban counterparts, we factored in students' backgrounds when designing the tests.

4.2 Risk Preferences

Risk preferences were elicited in both waves of the study. Given our sample of young children in a rural environment, the Gneezy and Potters (1997) allocation task was utilized. The single-decision allocation task is also sufficient for our purposes since we are interested in the treatment effects of chess, and not in the estimation of parameters of the utility function.[20] The first-wave task was incentivized by awarding the students stationary items based on their decisions. Different stationary items (e.g. pens, rulers, erasers – see Appendix 4 for the precise items) were awarded to reduce diminishing returns in utility associated with receiving multiple instances of the same item. The task involves choosing from one of five alternatives. The outcome of each alternative is determined by a coin flip. Thus, each alternative constitutes a lottery. The first alternative is completely risk-free, rewarding four items to a person regardless of the result from the coin flip. The alternatives grow progressively riskier, with the final alternative rewarding 12 items for a "heads" and no items for a "tails." In choosing this final alternative, students are choosing to "invest" all four items with a 50% chance of them tripling and a 50% chance of losing the investment. The expected value of the alternatives (in terms of items) increases with the level of risk. Thus, a risk-neutral or risk-loving person always chooses the final alternative, while a risk-averse individual will choose between the first and fourth alternatives, depending on the extent of their risk aversion. The description of the task is found in Appendix 4.

To ensure that students do not discuss or see the choices made by other students during implementation of the task, each student was called up one at a time, then taken to a separate room. A control question was included prior to students making their actual choices to ensure that each student understood the consequences of their decisions. Following their decisions, a coin was flipped in front of them to decide how many stationary items they would receive.

[20] For a review of risk-elicitation tasks, see Charness et al. (2013).

In the second wave, conducted in late October 2016, the same task was used, with two changes. First, to control for potential order effects in the various tasks (e.g. from students' success in one task influencing their behavior in another), we switched the orders of the risk and time preference elicitation tasks (the risk preference task was done first in the first wave). Second, to further reduce diminishing returns in utility associated with receiving multiple instances of the same item, we rewarded students with tokens that could be used to purchase several new attractive items (see the second part of Appendix 4). In this implementation of the risk preference task, the safest alternative gave students a guaranteed five tokens, while the riskiest alternative gave students the possibility of obtaining 15 tokens ("heads") or no tokens ("tails"). Hence, the rate of return on investment remains the same as in the first wave. Details of the task undertaken in Wave 2 can also be found in Appendix 4.

4.3 Time Preferences

Time preferences were elicited in both waves and at the same time as risk preferences, with the order of the two tasks reversed across waves. In the first wave (January-February 2016), we used a multiple-price-list format popularized by Coller and Williams (1999). Unlike risk preferences, it is less common to find single-decision implementations of elicitation tasks for time preferences.[21] Additionally, it is common for the multiple price list format to be implemented on children.[22]

In this task, students make five decisions. For each decision, they choose between receiving four pieces of candy tomorrow ("earlier"), vs. receiving x pieces of candy in eight days ("later"), where $x \in \{4, 6, 8, 10, 12\}$. This is close to the design adopted by Alan and Ertac (2018), in which the choice was between two gifts today vs. y gifts one week later, where $y \in \{2, 4, 6, 8, 10\}$. We chose candy to differentiate it from the incentives presented in the risk preference elicitation tasks in hopes of reducing any diminishing marginal utility associated with potentially obtaining too many stationary items. Candy was also used to incentivize children's time preference elicitation in Andreoni et al. (2017). The design adopts the "front-end delay" found in Harrison et al. (2002) and Castillo et al. (2011), whereby no rewards are

[21] The exception to this is Angerer et al. (2015) who effectively implement the time-preference equivalent for the Gneezy and Potters (1997) task. They find that both the multiple price list and simpler single decision task are highly correlated. However, the latter lacks the ability to identify inconsistent behaviour (which they find cannot be attributable to mere misunderstanding).

[22] For example, Bettinger and Slonim's (2007) study involved children ages 5-16 in the US; Castillo et al.'s (2011) analysis involved children ages 13-14 in the US; Sutter et al.'s (2013) study involved children ages 10-18 in Austria; Alan and Ertac's (2018) study involved children ages 9-13 in Turkey.

presented on the same day the task is performed. In doing so, the aim is to minimize any apparent impatience arising from a lack of trust in the experimenters, or any psychological discontinuities that may arise from imagining payment in the future vs. an immediate "now" that may generate a higher level of time inconsistency in the form of present bias.

Following previous studies on time preferences, we attempt to test for time inconsistency by presenting students with an additional five decisions. Here, the candy received between the "earlier" and "later" alternatives remains identical, with the only difference being that the earlier alternative was paid out in eight days, and the later alternative, in 15 days. This delay resembles the seven-day (earlier) and 14-day (later) implementation that Alan and Ertac (2018) used. Time inconsistency is particularly relevant to our implementation because it often has been tied to self-control, commitment problems, and procrastination (e.g. Frederick et al., 2002). It is unclear *a priori* whether the effect of chess training will be stronger on patience or on the incidence of time consistency.

The students were paid for only one of the 10 decisions they made for the time preference task. This was determined by having an experimenter (randomly) draw one of 10 numbered pieces of paper from a jar in front of the students (the detailed instructions are provided in Appendix 5).

Our Wave 2 time preference task was refined and chosen after observing the results from Wave 1: The students were extremely patient in Wave 1, with 85% of them choosing the "later" option at an effective interest rate of 50%. Hence, in an attempt to increase the granularity and variation in the information elicited from student choices, we adopted the convex time-budget task of Andreoni and Sprenger (2012). This also is done in Alan and Ertac (2018) in their follow-up wave. This task differs from the Wave 1 task in the following dimensions: (i) There are only three, rather than five, decisions (choice sets), and each choice set now contains five (instead of only two) alternatives; (ii) There is no more front-end delay since this may be making students overly patient in the first wave; and (iii) We rewarded students with tokens that could be used to purchase several new attractive items.

Specifically, in this task, students have to make three decisions. For each decision, the student chooses from five alternatives where each alternative results in receiving some tokens today (early) and some other tokens in seven days (later). For each decision, the most impatient alternatives result in receiving 12 tokens earlier and no tokens later, while the most patient alternatives result in receiving no tokens earlier and $z = 12 \times (1 + r)$ tokens later, where $r \in \{0, 0.33, 0.66\}$ is the interest rate. In addition, we continued to test for time inconsistency by including three more decisions that differed only in having the "earlier" outcome in seven days and the "later" outcome in 14 days. Only one of the six decisions was paid out; this was determined using the same method as in Wave 1. The equivalent interest rates in Alan and Ertac (2018) were 0.25 and 0.50. We included $r = 0$ as an indicator of the concavity of the utility function since any choice to delay receiving tokens in this case can be attributed purely to the diminishing returns to utility of receiving tokens. Since the students could effectively receive everything early and delay their own actual consumption, one can also view choosing to receive tokens later at $r = 0$ as a demand for a commitment device. The tokens earned in this task, together with the tokens earned in the risk task in Wave 2, could then be exchanged for several different attractive items (see Appendix 4). Instructions for the convex time-budget task are provided in Appendix 5.

4.4 Creativity and Attention/Focus

We also investigate whether undertaking intensive chess lessons can affect children's creativity and attention/focus. For assessing creativity, we use the Torrance Tests of Creative Thinking (Torrance, 1966) and Guilford's (1967) alternative uses test. For attention and focus, we employ two frequently used tests for the assessment of attention: the digit-cancellation test (Diller et al., 1974) and the digit-symbol test (Wechsler, 1991). These tests are described in more detail in Appendix A and Appendix B.

5. Empirical Approach

With randomization, the identification strategy used is straightforward. The benchmark model used to estimate the intention to treat effects (ITT) – the average treatment effect for children in fifth grade in schools that were randomly assigned to receive chess training – is the following OLS regression:

$$Y_{i,s} = \alpha + \delta treat_s + \beta X_{i,s} + \varepsilon_{i,s} \tag{1}$$

$Y_{i,s}$ denotes outcomes for individual *i* in school *s*, and $treat_s$ is whether a school was assigned to treatment group or not. Randomization was done at the school level, and all students in fifth grade in 2016 in the treatment schools were invited to participate in the chess training program.[23] We regression-adjust our results using a set of baseline covariates, $X_{i,s}$ which

[23] Unfortunately, student attendance on each day of the chess training was not recorded, thereby not allowing us to measure treatment receipt.

includes individual and household characteristics of the student to increase the precision of our results. Standard errors are clustered at the school level.

As an alternative way of performing statistical inference due to the clustered nature of the data, *p*-values using the wild bootstrap proposed by Cameron et al. (2008) are also computed. As many outcomes have been examined, this raises the issue of multiple hypothesis testing. To control for the false discovery rate (FDR), we provide sharpened *q*-values (Benjamini et al., 2006) using the procedure implemented in Stata by Anderson (2008). The interpretation of q-values is analogous to interpreting p-values – the q-values presented denote the lowest critical level at which a null hypothesis is rejected when controlling for the false discovery rate. Families of related p-values are typically used to estimate q-values. In our study, we take a conservative approach and use all outcomes tested rather than grouping the tests into families based on the domain tested.

6. Results

We present two sets of program impacts – unadjusted and regression adjusted – for the various cognitive and non-cognitive outcomes examined in Tables 2 to 5. The sample sizes for unadjusted and regression adjusted results vary and depend on whether both baseline data on characteristics and data on the outcome were measured. As data were collected on different days, the variation in sample sizes across outcomes partly reflects the fact that on any given day, student absenteeism is high in primary schools in rural Bangladesh.

Three alternative sets of *p*-values are presented. First, in the columns for unadjusted and regression adjusted impacts, we present conventional standard errors in parentheses and the associated *p*-values (using asterisks) from a regression model based on clustered standard errors. Second, *p*-values using the wild bootstrap (1,000 replications) proposed by Cameron et al. (2008) are reported in square brackets. Third, we compute FDR sharpened *q*-values (Benjamini et al., 2006) using the procedure in Anderson (2008). These *q*-values are presented using curly brackets.

6.1 Academic Results

We consider two types of test scores to measure cognitive ability. The first involves the use of a project-administered math test. The treatment group scored slightly better in the pre-program math test relative to the control group, but the difference was not statistically significant (providing further supporting evidence that the randomization was well-implemented). The gap between the treatment and control groups widened in the post-program test conducted shortly after the intensive chess training had ended. However, the difference was again not statistically significant (see Table 2).

The second measurement of academic achievement involved the use of the PSC exam which took place 9-10 months after the training. The results of the PSC exam are provided in Table 3. We find a significant positive effect from our intensive chess-instruction program on math grades in the PSC exam using both conventional p-values and the wild cluster bootstrap (*p*-value = 0.030 using the wild cluster bootstrap).[24] The treatment-control difference of 0.71 points is approximately equivalent to between half and a full letter math grade. However, the false discovery rate (FDR) sharpened q-values that account for multiple hypothesis testing suggest that this difference is not significant (q-value = 0.161). Likewise, although the impact on overall GPA (0.41) is statistically significant using conventional clustered standard errors and the wild bootstrap (*p*-value = 0.086), the FDR sharpened q-values suggest it is not significant.

6.2 Risk Preference Results

The average value of the alternative chosen in the risk-elicitation task was used for assessing a treatment effect on risk preferences, in which a higher value indicates a riskier choice. The values range from 1-5 in Wave 1, and 1-6 in Wave 2.[25] Results are depicted in Table 4. In Wave 1, treated students invested, on average, 0.3 more items into the risky "asset" (*p*-value = 0.144). In Wave 2, treated students invested, on average, 1.75 more tokens into the risky asset (*p*-value =0.002). Hence, although we find no significant effect on risk preferences in Wave 1, a strong effect (both in terms of size and significance) emerges in Wave 2 – chess training decreases risk-aversion.[26] Importantly, this impact remains statistically significant using the FDR q-values and in both the regression adjusted and non-regression adjusted samples.[27]

Figure 3 breaks down the treatment effects according to each available alternative and highlights the changes between Waves 1 and 2. For both waves, we can see that the largest difference emerges for alternative 1 – the safest alternative. In addition, there is a strong effect

[24] When the number of bootstrap replications is increased from 1000 to 5000, the p-value from the wild cluster bootstrap is very similar (=0.034).

[25] There was one additional alternative in Wave 2 because of the higher granularity of the rewards.

[26] When the number of bootstrap replications is increased from 1000 to 5000, the p-values from the wild cluster bootstrap for the non-regression-adjusted and regression-adjusted impacts are still highly significant (equal 0.0008 and 0.0004 respectively).

[27] The results remain statistically significant when we include pre-program project-administered math test scores as an additional control variable (which we do not use in our general set of controls as it will reduce our sample size).

in Wave 2 on alternative 6 – the riskiest alternative – suggesting that chess training may have resulted in a significant number of students switching from being risk-averse to either risk-neutral or risk-loving over time.[28] The fact that the effects on risk-preferences are detected only 9-10 months after the initial program was launched suggests that these effects are possibly linked to changes in habitual and long-term behaviour rather than the purely cognitive aspect of having been instructed on how to play chess. The results are consistent with our finding that nine out of ten students were still playing chess 9-10 months after the intervention ended, which allows students enough time to develop a deeper understanding of strategy and risk in the game through playing hundreds of games, while also giving them prolonged exposure to interactions involving bilateral and non-physical competition.

6.3 Time Preference Results

In Wave 1 of the time-elicitation task, students were given five choice sets and indicated in each instance whether they would take the patient alternative ("later") or impatient alternative ("earlier"). For each individual, we assign a count of impatient alternatives chosen. Their sum was used to assess average treatment effects, with higher values indicating more impatience. We also did this for the five choice sets with one week of delay. The results are depicted in the first two rows of the top panel of Table 5. The results for both the standard and delayed choice sets are highly statistically insignificant (p-values = 0.716 and 0.540, respectively), as well as small in magnitude.

For Wave 2, students had two choice sets, with each set containing five alternatives[29] Each alternative is assigned a score 1-5, with a higher score indicating greater impatience. For each student, we summed the scores across the two choice sets. The results (the first two rows of the bottom panel of Table 5), with and without delay, remain highly statistically insignificant (p-values = 0.898 and 0.634, respectively).

Given that time preferences were elicited using a multiple price-list method, we can conduct two additional tests. The first involves a test for time inconsistency. In both waves, we

[28] Figure 2 also suggests that the control group is more risk-averse in Wave 2. This is consistent with evidence summarised in Schildberg-Hörisch (2018), suggesting that both children and adults grow more risk-averse over time. It is also possible that the small differences in the risk tasks across waves (the Wave 1 incentive involved stationary items; the Wave 2 incentive included other items such as food) and the switch in the order of the time and risk preference tasks may partly explain changes in control group preferences. However, there is no a priori reason to expect such effects to be isolated to the control group alone.

[29] A third choice set involving $r = 0$ was included to elicit the presence of diminishing returns in utility. If the marginal utility of receiving tokens at any given period of time is non-diminishing, students should choose alternative 1. In our results, only 26% of students chose alternative 1, suggesting that diminishing returns in utility plays a non-trivial role in decisions.

had students make decisions over an original and one-week-delayed set that differ only in having payoffs in the latter realized seven days later than the original. We consider two possible variables for a test of time inconsistency: (i) a continuous variable that scores a “1” for each decision that fails to match across both the original and the corresponding one-week-delayed decision, and (ii) a binary variable that takes on a value of “1” if *at least one* decision in the original decisions fails to match their corresponding one-week-delayed decision.

For time inconsistency, there is some evidence that students in the treatment group are less likely to make time inconsistent decisions in Waves 1 and 2 using conventional p-values. The FDR q-values remain significant for time inconsistency in Wave 1, but only for the smaller regression adjusted sample and not for the larger non-regression adjusted sample.

The second additional test we perform on the time preference data involves checking for non-monotonicity of time preferences. Well-defined, monotonic time preferences require that a choice at some interest rate r must be at least as patient as some other interest rate $r' < r$ (e.g. see Harrison et al., 2002). In Wave 1, this translates to students switching from the “earlier” to “later” option at most once. In Wave 2, it requires that a choice at some interest rate r must be of a value at least as high as the choice at some other interest rate $r' < r$. We construct a binary variable that takes the value “1” if such a monotonicity requirement is violated. The results are presented in the last row of each panel in Table 5. Both conventional p-values and wild cluster bootstrap p-values suggest that students in the treatment group are less likely to violate the monotonicity requirement in Waves 1 and 2. However, the insignificance of the FDR q-values suggests that this result might not be robust.

6.4 Results for Creativity and Attention/Focus

Our results do not suggest that there are any short term effects of chess instruction on creativity, or medium term effects on focus and attention. Discussion of these additional non-cognitive outcomes are provided in Appendix A and Appendix B.

6.5 Heterogeneous Treatment Effects

We estimate subgroup effects by gender, pre-program math ability, and household income. With an individual belonging to one of two possible subgroups (S = 1 or S = 0) and an indicator created for each subgroup type, we use the following estimating equation:

$$Y_{i,s} = \alpha + \delta^1 (S = 1 \times treat_s) + \delta^2 (S = 0 \times treat_s) + \delta^3 (S = 1) + \beta X_{i,s} + \varepsilon_{i,s} \qquad (3)$$

The two interaction terms involving the treatment dummy can be interpreted as the impact of the treatment for each subgroup type. Specifically, δ^1 is interpreted as the ITT for individuals in the first subgroup (e.g. males), and δ^2 is interpreted as the ITT for individuals in the second subgroup (e.g. females). Table 6 reports the subgroup impact results. Overall, we fail to detect any subgroup differences by gender, baseline math ability, or baseline household income.

7. Summary and Conclusions

This paper evaluates the effects of learning chess using a randomized experiment on grade five students in rural Bangladesh. The intervention comprised of a 30-hour training program based on a curriculum approved by the World Chess Federation. By employing a field experiment and collecting a range of academic and non-academic outcomes, we have provided credible estimates of the benefits chess instruction can have for children's cognitive and non-cognitive outcomes. In terms of academic outcomes, we use high-stakes, age-appropriate, and externally marked academic tests for schools to measure the effectiveness of the intervention, meaning our results are unlikely to be influenced by limitations surrounding the outcome test. We examine both short-term effects based on assessments made shortly after the conclusion of the program, as well as medium-term effects based on assessments conducted 9-10 months after the program ended, allowing us to examine whether there is a lasting effect.

One novel contribution of this paper is a focus on the link between chess and non-cognitive outcomes relevant to the labor market: risk, time preferences, patience, creativity, attention, and focus. The previous literature has emphasized potential links between chess and academic outcomes.

Our main finding is that chess training reduces the treatment group's level of risk aversion almost a year after the intervention ended. This finding is robust to correction for multiple hypothesis testing. While our impact estimates based on conventional p-values and wild bootstrap p-values provide some indication of effects on math scores, time inconsistency and non-monotonic time preferences, there is less conclusive evidence after controlling for multiple hypothesis testing using the false discovery rate.

It is often said that chess is an easy game to learn but difficult to master. Our intervention helped to introduce the game of chess to students who had, in general, previously not been exposed to the game. Beyond the rules of how pieces move and how the game is won, strategy and tactics in various phases of the game were also introduced. It was ascertained that approximately nine out of ten students continued to play and practice chess when they were asked 9-10 months after the intensive three-week chess course ended. It is plausible that this repeated playing and honing of their skills could have contributed to a better appreciation for the concept of risk-taking, leading to a reduction in risk aversion. Our findings are consistent with evidence showing that risk preferences share an association with aspects of cognitive ability (Frederick 2005; Dohmen et al. 2010; Benjamin et al. 2013 and Andreoni et al. 2019). It also highlights that the cognitive development potentially offered by chess instruction need not be realised primarily through traditionally recognised cognitive outcomes such as math scores.

In light of Jerrim et al's (2018) recent experimental evidence that chess training had no impact on academic outcomes for students in an urban setting, and our experimental finding that there is at best only weak evidence of an impact on math for students in a rural developing country setting, the common and stereotypical link that is made implying a transfer of learning between chess and math should be reconsidered. Although math was the main hypothesis and focus of the experiment, with non-cognitive outcomes considered as secondary outcomes of interest in our research design, uncovering a link between chess and risk preferences is intriguing. Further work will need to be done to better understand more precisely the mechanisms underlying how chess can affect the development of risk preferences.

As some of the outcomes examined in this study are new to this literature, further field experiments can help determine the robustness of our findings. Our intervention is based on data from rural areas of a developing country, and the results obtained do not necessarily have external validity. Nonetheless, by focusing the intervention on a group of children who essentially had no prior experience playing chess and who did not have access to many contemporary toys and games common in developed countries (e.g. board games, computer games, mobile devices, Lego, etc.) that provide mental stimulation, we potentially allow for a fuller impact of chess lessons (if any) to emerge and be realized.

References

Ahmed, M., K. Ahmed, N. Khan and R. Ahmed. (2007). Access to education in Bangladesh - country analytic review of primary and secondary education. Institute of Educational Development, BRAC University.

Alan, S. and S. Ertac. (2018). Fostering patience in the classroom: results from randomized educational intervention. *Journal of Political Economy*, 126: 1865-1911.

Anderson, M. (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: 1481–1495.

Andreoni, J., A. Girolamo, J. List, C. Mackevicius and A. Samek. (2019). Risk Preferences of Children and Adolescents in Relation to Gender, Cognitive Skills, Soft Skills, and Executive Functions, National Bureau of Economic Research Working Paper No. 25723.

Andreoni, J., M. Kuhn, J. List, A. Samek and C. Sprenger. (2017). Field experiments on the development of time preferences (No. 00615). The Field Experiments Website.

Andreoni, J. and C. Sprenger. (2012). Estimating time preferences from convex budgets. *The American Economic Review*, 102: 3333–3356.

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: 158–169.

Banerjee, A., S. Cole, E. Duflo, and L.Linden. (2007). Remedying education: Evidence from two randomized experiments in India. *Quarterly Journal of Economics*, 122: 1235–1264.

Banerjee, A., R. Banerji, J. Berry, E. Duflo, H. Kannan, S. Mukerji, M. Shotland and M. Walton. (2017). From proof of concept to scalable policies: challenges and solutions, with an application. *Journal of Economic Perspectives*, 31: 73–102.

Becker, G. and C. Mulligan. (1997). The endogenous determinants of time preferences. *Quarterly Journal of Economics*, 112: 729–758.

Benjamin, D., S. Brown and J. Shapiro. (2013). Who is ‘behavioral’? Cognitive ability and anomalous preferences. *Journal of the European Economic Association*, 11: 1231–1255.

Benjamini, Y., A. Krieger and D. Yekutieli. (2006), Adaptive linear step-up procedures that control the false discovery rate. *Biometrika*, 93: 491–507.

Bettinger, E. and R. Slonim. (2007). Patience among children. *Journal of Public Economics*, 91: 343–363.

Binev, S., J. Attard-Montalto, N. Deva, M. Mauro and H. Takkula. (2011). Declaration of the European Parliament, 0050/2011.

Burks, S., J. Carpenter, L. Goette and A. Rustichini. (2009). Cognitive skills affect economic preferences, strategic behavior, and job attachment. *Proceedings of the National Academy of Sciences*, 106: 7745–7750.

Carneiro, P., C. Crawford and A. Goodman (2007). The impact of early cognitive and noncognitive skills on later outcomes, Centre for the Economics of Education Discussion Paper No. 92, UK.

Cameron, A., J. Gelbach and D. Miller. (2008). Bootstrap-based improvements for inference with clustered errors. *Review of Economics and Statistics*, 90: 414–427.

Castillo, M., J. Jordan, and R. Petrie. (2018). Discount rates of children and high school graduation. *The Economic Journal*, 129: 1153–1181.

Castillo, M., J. Jordan and R. Petrie. (2018). Children's rationality, risk attitudes and field behavior. *European Economic Review*, 102: 62–81.

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

Charness, G., U. Gneezy and A. Imas. (2013). Experimental methods: Eliciting risk preferences. *Journal of Economic Behavior & Organization*, 87: 43–51.

Chaudhury, N., J. Hammer, M. Kremer, K. Muralidharan and F. Rogers. 2006. Missing in action: teacher and health worker absence in developing countries. *Journal of Economic Perspectives*, 20: 91–116.

Coller, M. and M. Williams. (1999). Eliciting individual discount rates. *Experimental Economics*, 2: 107–127.

Cramond, B. (1993). The Torrance Tests of Creative Thinking: From design through establishment of predictive validity. In R. F. Subotnik & K. D. Arnold (Eds.), *Beyond Terman: Contemporary Longitudinal Studies of Giftedness and Talent* (pp. 229–254). Norwood, NJ: Ablex.

Csikszentmihalyi, M. (1990). The domain of creativity, in M. A. Runco and R. S. Albert, eds, *Theories of Creativity*, Sage Press, Newbury Park, CA, pp. 190–212.

Cunha, F. and J. Heckman. (2008). Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *Journal of Human Resources*, 43: 738–782.

Davis, S. and N. Eppler-Wolff. (2009). *Raising Children Who Soar*, Teachers College Press: New York.

Diller, L., Y. Ben-Yishay, L. Gerstman, R. Goodkin, W. Gordon and J. Weinberg. (1974). Studies in cognition and rehabilitation in hemiplegia. [Rehabilitation Monograph No. 50]. New York: New York University Medical Center Institute of Rehabilitation Medicine.

Dohmen, T., A. Falk, D. Huffman and U. Sunde. (2010). Are risk aversion and impatience related to cognitive ability? *American Economic Review*, 100: 1238–1260.

Dohmen, T., A. Falk, D. Huffman, U. Sunde, J. Schupp and G. Wagner. (2011). Individual risk attitudes: Measurement, determinants, and behavioural consequences. *Journal of the European Economic Association*, 9: 522–550.

Dreber, A., C. Gerdes and P. Gränsmark. (2013). Beauty queens and battling knights: Risk taking and attractiveness in chess. *Journal of Economic Behavior and Organization*, 90: 1–18.

Duflo, E., P. Dupas and M. Kremer. (2011). Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya. *American Economic Review*, 101: 1739–1774.

Eriksen, K. W. and O. Kvaløy. (2017). No guts, no glory: An experiment on excessive risk-taking. *Review of Finance*, 21: 1327-1351.

Franklin, B. (1786). The morals of chess, *Columbian Magazine*, December issue: 159–161.

Frederick, S. (2005). On the ball: cognitive reflection and decision-making. *Journal of Economic Perspectives*, 19: 25–42.

Frederick, S., Loewenstein, G. and O'Donoghue, T., (2002). Time discounting and time preference: A critical review. *Journal of Economic Literature*, 40: 351–401.

Gneezy, U. and Potters, J., (1997). An experiment on risk taking and evaluation periods. *Quarterly Journal of Economics*, 112: 631–645.

Golsteyn, B. H. Grönqvist and L. Lindahl. (2014). Adolescent time preferences predict lifetime outcomes. *Economic Journal*, 124: F739–F761.

Guilford, J. P. (1967). *The nature of human intelligence*. New York: McGraw-Hill.

Guiso, L. and M. Paiella. (2008). Risk aversion, wealth, and background risk. *Journal of the European Economic association*, 6: 1109–1150.

Harrison, G., M. Lau and M. Williams. (2002). Estimating individual discount rates in Denmark: A field experiment. *American Economic Review*, 92:1606–1617.

Heckman, J. and T. Kautz. (2012). Hard evidence on soft skills. *Labour Economics*, 19:451–464.

Heckman, J., J. Stixrud and S. Urzua. (2006). The effects of cognitive and noncognitive abilities on labour market outcomes and social behaviour, *Journal of Labor Economics*, 24: 411–482.

Heckman, J. and Y. Rubinstein. (2001). The importance of noncognitive skills: Lessons from the GED testing program. *American Economic Review*, 91:145–149.

Islam, A. (2019). Parent-Teacher Meetings and Student Outcomes: Evidence from a Developing Country. *European Economic Review*, 111: 273–304.

Jerrim, J. L. Macmillan, J. Micklewright, M. Sawtell and M. Wiggins. (2016). Chess in Schools: evaluation report and executive summary. University College London: UK.

Jerrim, J. L. Macmillan, J. Micklewright, M. Sawtell and M. Wiggins. (2018). Does teaching children how to play cognitively demanding games improve their educational attainment? Evidence from a randomised controlled trial of chess instruction in England. *Journal of Human Resources*, 53: 993–1021.

Kumar, A. and N. Saqib. (2017). School absenteeism and child labor in rural Bangladesh. *Journal of Developing Areas*, 51: 299–316.

Levitt, S., J. List and S. Sadoff. (2011). Checkmate: exploring backward induction among chess players. *American Economic Review*, 101: 975–990.

Outreville, J. (2015). The relationship between relative risk aversion and the level of education: a survey and implications for the demand for life insurance. *Journal of Economic Surveys*, 29: 97–111.

Runco, M. and S. Sakamoto. (1999). Experimental studies of creativity, in *Handbook of Creativity*, R. Sternberg (ed.), Cambridge University Press: UK.

Sala, G. and F. Gobet. (2016). Do the benefits of chess instruction transfer to academic and cognitive skills? A meta-analysis, *Educational Research Review*, 18: 46–57.

Schildberg-Hörisch, H. (2018). Are risk preferences stable? *Journal of Economic Perspectives*, 32: 135–54.

Scholz, M., H. Niesch, O. Steffen, B. Ernst, M. Loeffler and E. Witruk et al. (2008). Impact of chess training on mathematics performance and concentration ability of children with learning disabilities. *International Journal of Special Education*, 23: 138–148.

Schneider, W., H. Gruber, A. Gold and K. Opwis. (1993). Chess expertise and memory for chess positions in children and adults. *Journal of Experimental Child Psychology*, 56: 328–349.

Spadoni, L. and J. Potters. (2018). The effect of competition on risk taking in contests. *Games*, 9: 72.

Sutter, M., M. Kocher, D. Glätzle-Rützler and S. Trautmann. (2013). Impatience and uncertainty: Experimental decisions predict adolescents' field behavior. *American Economic Review*, 103: 510–531.

Tietjen, K., A. Rahman and S. Spaulding. (2004). Bangladesh Educational Assessment Time to Learn: Teachers' and Students' Use of Time in Government Primary Schools in Bangladesh, USAID.

Torrance, E. (1966). The Torrance Tests of Creative Thinking – Norms: Technical Manual Research Edition—Verbal Tests, Forms A and B— Figural Tests, Forms A and B. Princeton, NJ: Personnel Press.

Torrance, E., O. Ball and H. Safter. (2008). Torrance Tests of Creative Thinking: Streamlined scoring guide for figural forms A and B. USA: Scholastic Testing Service, Inc.

Travis, F. (1998). Cortical and cognitive development in 4th, 8th and 12th grade students: The contribution of speed of processing and executive functioning to cognitive development. *Biological Psychology*, 48: 37–56.

Trinchero, R. and G. Sala. (2016). Can chess training improve Pisa scores in Mathematics? An experiment in Italian primary schools. *Eurasia Journal of Mathematics, Science and Technology Education*, 12, 655–668.

Tversky, A. and D. Kahneman. (1981). The framing of decisions and the psychology of choice. *Science*, 211(4481): 453–58.

Unterrainer, J., C. Kaller, R. Leonhart and B. Rahm. (2011). Revising superior planning performance in chess players: the impact of time restriction and motivation aspects. *American Journal of Psychology*, 124: 213–225.

Waters, A., F. Gobet and G. Leyden. (2002). Visuospatial abilities in chess players. *British Journal of Psychology*, 30: 303–311.

Wechsler, D. (1991). Wechsler Intelligence Scale for Children (3rd ed.). San Antonio, TX: Psychological Corporation.

Figure 1: Location of Treatment and Control Schools in Bangladesh

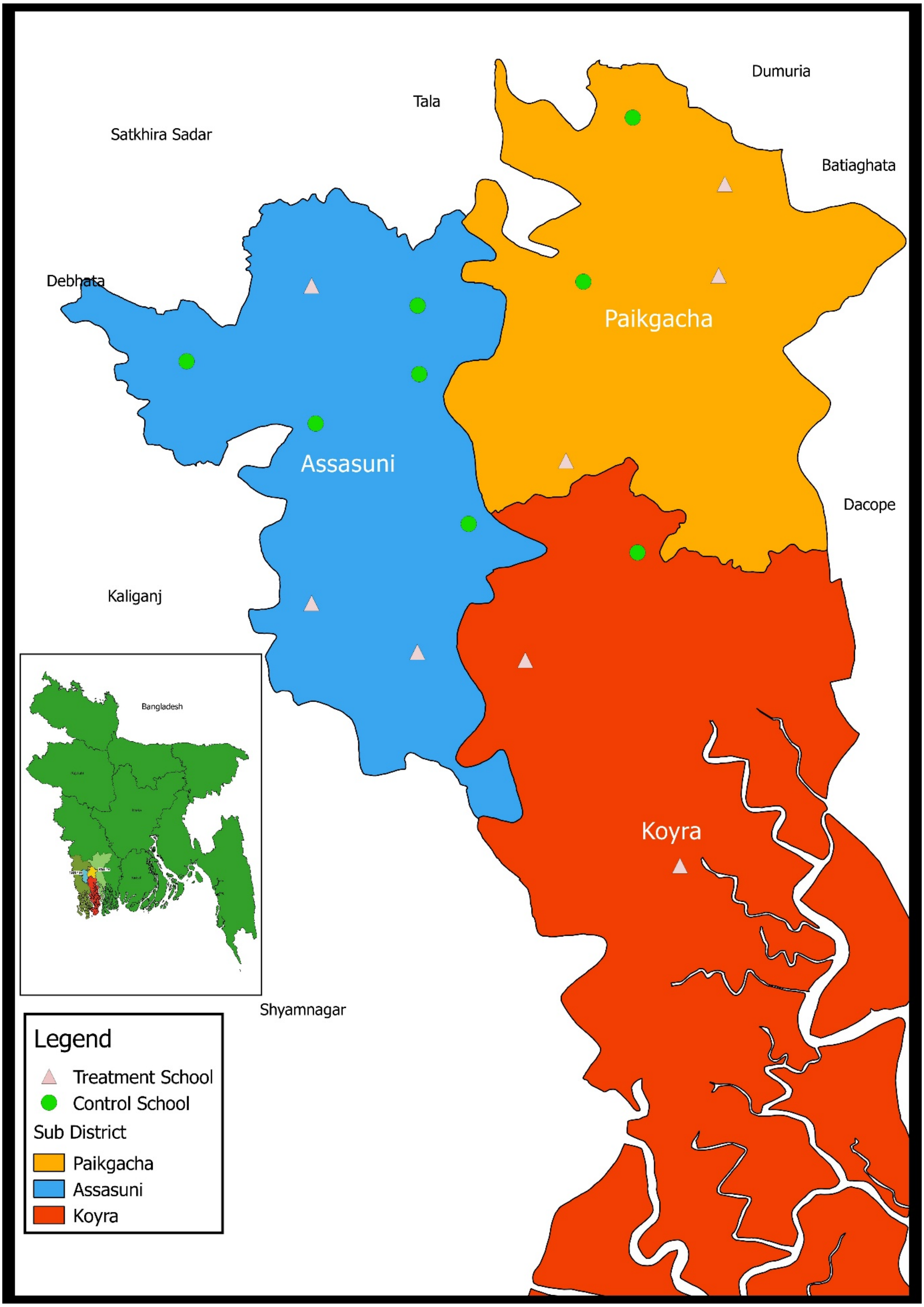
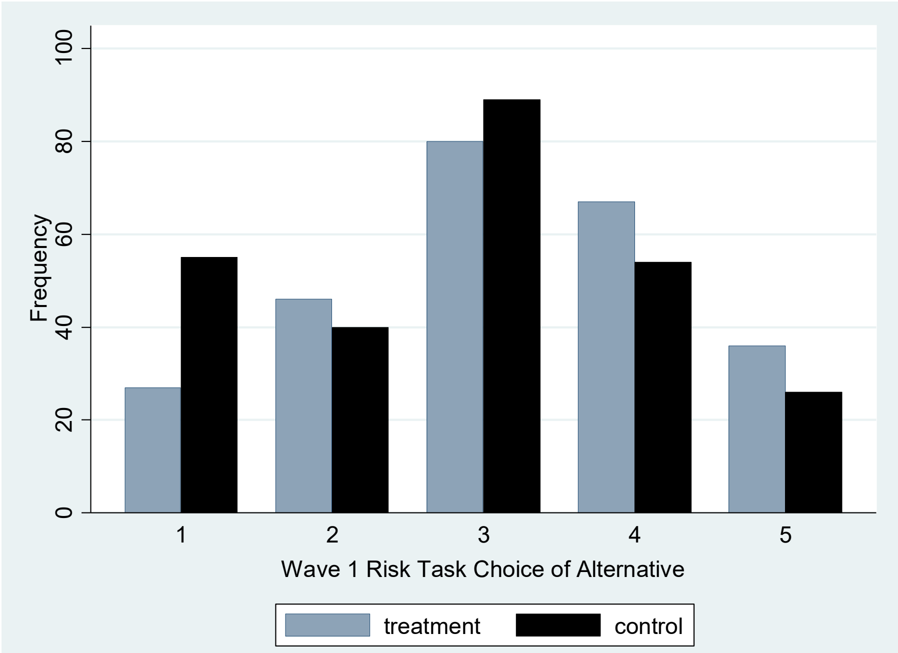


Figure 2: Intervention Timeline

Data collected prior to the start of the program (Nov 2015 – Jan 2016)	Wave 1 data collected at the end of the three week program (Jan/Feb 2016)	Wave 2 data collected 9-10 months later (Oct/Nov 2016)
• Parent survey (Nov/Dec 2015) • Pre-program chess knowledge test (Jan 2016) • Short personality test (Jan 2016) • Pre-program math test (Jan 2016)	• Time preferences test, Wave 1 • Risk preferences test, Wave 1 • Creativity test • Post-program math test • Post-program chess knowledge test for the treatment group • Network survey for the treatment group	• Time preferences test, Wave 2 (Oct 29/30, 2016) • Risk preferences test, Wave 2 (Oct 29/30, 2016) • Attention/focus test (Oct 29/30, 2016) • Network survey for the treatment group • Primary School Certificate (PSC) national examination (Nov 20-27, 2016)

Note: The chess program was conducted from Saturday to Tuesday over a period of three weeks. Note that Friday is considered the weekly holiday in Bangladesh (equivalent to Sunday in other developed countries) and that the school week runs from Saturday to Thursday. There were a total of 12 program days where chess lessons were provided.

Figure 3: Distribution of Choices across Groups, and Waves in the Risk-Elicitation Task



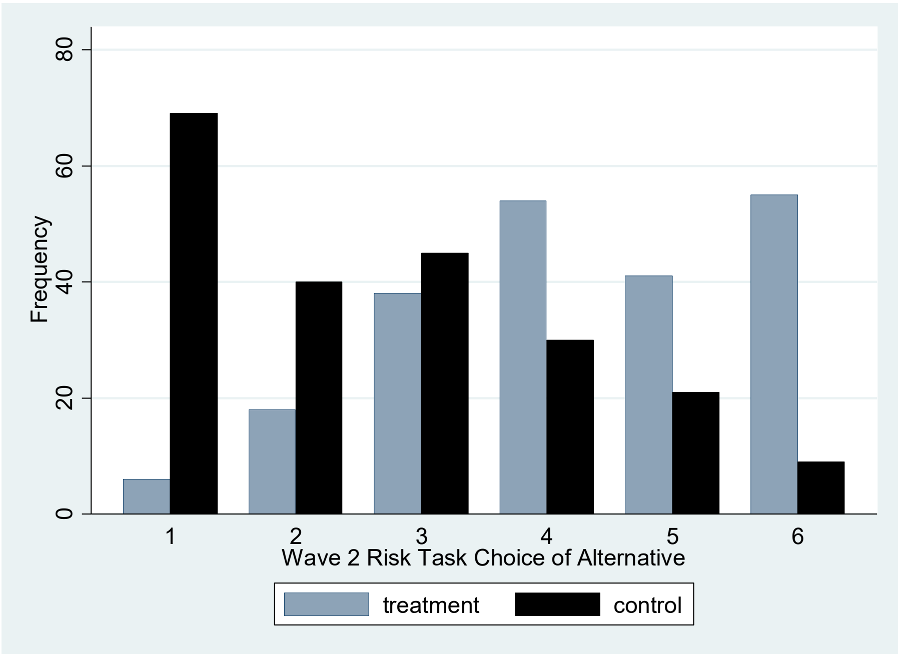


Table 1: Treatment/Control Raw Mean Differences in Household Characteristics

Variable	Treatment Mean	Control Mean	Difference
Household income (in takas)	8377.2	8771.0	-393.8
			(544.5)
Number of household members	4.406	4.351	0.055
			(0.135)
Sanitary ring latrine in the house	0.626	0.642	-0.016
			(0.060)
Drinking water in the house from tube well	0.633	0.816	-0.183
			(0.164)
Existence of electricity supply in the house	0.338	0.497	-0.158
			(0.171)
Distance of the school from the home (km)	1.115	0.674	0.441
			(0.358)
Value of total assets except land (in takas)	68089.0	63041.7	5047.3
			(10326.3)
Household religion (Muslim = 1)	0.932	0.938	-0.005
			(0.032)
Do any of the parents know how to play chess	0.103	0.066	0.037
			(0.030)
Someone with more than grade 10 education in household	0.139	0.132	0.007
			(0.029)
Father's years of schooling	4.12	4.37	-0.244
			(0.655)
Mother's years of schooling	4.13	4.08	0.048
			(0.732)
Father's age	39.96	39.97	-0.011
			(0.603)
Mother's age	33.64	33.61	0.029
			(0.643)
Father works as labourer/in agriculture	0.676	0.608	0.068
			(0.076)
Mother is a housewife	0.986	1.000	-0.014**
			(0.005)
Two-parent household	0.996	1.000	-0.003
			(0.003)
Gender of student (male =1)	0.430	0.494	0.064
			(0.049)
N	281	288	

Notes: Standard errors in parentheses and are clustered at the school level. **p*-value<0.1 ** *p*-value<0.05 *** *p*-value<0.01.

Table 2: Mathematics (Wave 1)

Variable	Control Mean	Unadjusted Impact	Regression Adjusted Impact
Math pre-marks	18.71	0.506 (3.168)	1.362 (2.719)
		[0.820]	[0.608]
		{0.999}	{0.705}
N	215	494	445
Math post-marks	14.38	1.304 (3.019)	2.072 (2.414)
		[0.680]	[0.442]
		{0.999}	{0.648}
N	209	478	428

Notes: Standard errors in parentheses and are clustered at the school level, with conventional p-values reported as **p*-value<0.1 ** *p*-value<0.05 *** *p*-value<0.01. The associated wild bootstrapped p-values are reported in square brackets, while false discovery rate (FDR) sharpened q-values (Benjamini et al., 2006) computed using the procedure in Anderson (2008) are reported in curly brackets. Covariates included in the regression adjustment are from Table 1. The wild bootstrap *p*-values are based on 1000 replications. Control means are based on the regression adjusted sample.

Table 3: Primary School Certificate (PSC) National Exam Scores (Wave 2)

Variable	Control Mean	Unadjusted Impact	Regression Adjusted Impact
Bangla	3.76	0.282 (0.224)	0.347* (0.197)
		[0.312]	[0.180]
		{0.622}	{0.370}
Math	2.93	0.718* (0.357)	0.705** (0.283)
		[0.086]	[0.030]
		{0.520}	{0.161}
Science	3.60	0.341 (0.287)	0.292 (0.294)
		[0.282]	[0.426]
		{0.622}	{0.648}
English	2.90	0.457 (0.334)	0.398 (0.330)
		[0.222]	[0.338]
		{0.622}	{0.583}
Social Science	3.63	0.240 (0.371)	0.306 (0.319)
		[0.612]	[0.434]
		{0.999}	{0.648}
Religious Studies	3.95	0.387 (0.230)	0.405* (0.209)
		[0.142]	[0.084]
		{0.520}	{0.283}
Overall GPA	3.45	0.413 (0.242)	0.414* (0.214)
		[0.124]	[0.086]
		{0.520}	{0.283}
N	190	434	395

Notes: Standard errors in parentheses and are clustered at the school level, with conventional p-values reported as **p*-value<0.1 ** *p*-value<0.05 *** *p*-value<0.01. The associated wild bootstrapped p-values are reported in square brackets, while false discovery rate (FDR) sharpened q-values (Benjamini et al., 2006) computed using the procedure in Anderson (2008) are reported in curly brackets. Covariates included in the regression adjustment are from Table 1. The wild bootstrap *p*-values are based on 1000 replications. The conversion from letter grades to scores is as follows: A+ = 5 points, A = 4 points, A- = 3.5 points, B = 3 points, C = 2 points, D = 1 point, F = 0 points. Control means are based on the regression adjusted sample.

Table 4: Risk preferences (Waves 1 and 2)

Variable	Control Mean	Unadjusted Impact	Regression Adjusted Impact
Wave 1 (Min 1, Max 5), higher value = less risk averse	2.84	0.319* (0.166)	0.301 (0.175)
		[0.084]	[0.144]
		{0.520}	{0.370}
N	225	520	450
Wave 2 (Min 1, Max 6), higher value = less risk averse	2.65	1.647*** (0.437)	1.752*** (0.442)
		[0.000]	[0.002]
		{0.001}	{0.028}
N	191	426	381

Notes: Standard errors in parentheses and are clustered at the school level, with conventional p-values reported as **p*-value<0.1 ** *p*-value<0.05 *** *p*-value<0.01. The associated wild bootstrapped p-values are reported in square brackets, while false discovery rate (FDR) sharpened q-values (Benjamini et al., 2006) computed using the procedure in Anderson (2008) are reported in curly brackets. Covariates included in the regression adjustment are from Table 1. The wild bootstrap *p*-values are based on 1000 replications. Control means are based on the regression adjusted sample. False discovery rate (FDR) sharpened q-values (Benjamini et al., 2006) are computed using the procedure in Anderson (2008).

Table 5: Time preferences (Waves 1 and 2)

Variable	Control Mean	Unadjusted Impact	Regression Adjusted Impact
Wave 1			
Impatience (0-5)	1.26	0.038 (0.062)	-0.026 (0.062)
		[0.630]	[0.716]
		{0.999}	{0.723}
Delayed impatience (0-5)	1.32	0.016 (0.065)	-0.040 (0.061)
		[0.810]	[0.540]
		{0.999}	{0.681}
Time inconsistency (binary)	0.28	-0.086** (0.037)	-0.162*** (0.040)
		[0.068]	[0.008]
		{0.464}	{0.053}
Time inconsistency (0-5)	0.38	-0.091 (0.064)	-0.234*** (0.060)
		[0.210]	[0.006]
		{0.622}	{0.053}
Non-monotonicity (binary)	0.14	-0.089*** (0.028)	-0.121*** (0.018)
		[0.010]	[0.002]
		{0.150}	{0.028}
N	224	521	450
Wave 2			
Impatience (2-10)	5.19	-0.338 (0.331)	-0.087 (0.365)
		[0.354]	[0.898]
		{0.622}	{0.951}
Delayed impatience (2-10)	5.27	-0.120 (0.270)	-0.151 (0.257)
		[0.738]	[0.634]
		{0.999}	{0.705}
Time inconsistency (binary)	0.74	-0.073 (0.046)	-0.060 (0.042)
		[0.136]	[0.202]
		{0.520}	{0.389}
Time inconsistency (0-2)	1.13	-0.145 (0.084)	-0.129* (0.065)
		[0.112]	[0.090]
		{0.520}	{0.283}
Non-monotonicity (binary)	0.67	-0.107* (0.052)	-0.126** (0.055)
		[0.054]	[0.062]
		{0.510}	{0.283}
N	191	426	381

Notes: Standard errors in parentheses and are clustered at the school level, with conventional p-values reported as *p-value<0.1 ** p-value<0.05 *** p-value<0.01. The associated wild bootstrapped p-values are reported in square brackets, while false discovery rate (FDR) sharpened q-values (Benjamini et al., 2006) computed using the procedure in Anderson (2008) are reported in curly brackets. Covariates included in the regression adjustment are from Table 1. The wild bootstrap p-values are based on 1000 replications. Control means are based on the regression adjusted sample. False discovery rate (FDR) sharpened q-values (Benjamini et al., 2006) are computed using the procedure in Anderson (2008).

Table 6: Subgroup Analysis

Subgroup	Effect on PSC Math	Effect on risk (Wave 2)	Effect on time inconsistency (Wave 1 binary)	Effect on time inconsistency (Wave 1 0-5)	Effect on non-monotonicity (Wave 1 binary)	Effect on non-monotonicity (Wave 2 binary)
Male	0.924** (0.332)	1.664*** (0.445)	-0.214*** (0.057)	-0.241** (0.104)	-0.142*** (0.037)	-0.212** (0.087)
Female	0.528 (0.311)	1.822***(0.492)	-0.121** (0.046)	-0.228*** (0.076)	-0.103*** (0.027)	-0.058 (0.069)
p-value for subgroup difference = 0	0.181	0.642	0.174	0.927	0.472	0.181
Low math ability	0.710** (0.309)	1.950*** (0.501)	-0.147** (0.060)	-0.157 (0.090)	-0.108*** (0.033)	-0.101 (0.059)
High math ability	0.785* (0.389)	1.679*** (0.472)	-0.199*** (0.060)	-0.339*** (0.097)	-0.153*** (0.036)	-0.173 (0.101)
p-value for subgroup difference = 0	0.851	0.456	0.547	0.210	0.395	0.523
Low household income	0.786** (0.315)	1.778*** (0.502)	-0.194*** (0.029)	-0.264*** (0.043)	-0.135*** (0.026)	-0.109* (0.062)
High household income	0.556 (0.312)	1.739*** (0.484)	-0.106 (0.074)	-0.179 (0.123)	-0.094** (0.041)	-0.161* (0.082)
p-value for subgroup difference = 0	0.427	0.936	0.161	0.468	0.479	0.566

Notes: Standard errors in parentheses and are clustered at the school level. **p*-value<0.1 ** *p*-value<0.05 *** *p*-value<0.01. Covariates included in the regression adjustment are discussed in Section 3 of the paper.

APPENDIX: ADDITIONAL NON-COGNITIVE OUTCOMES

Appendix A: Creativity

Many different conceptions of creativity have emerged that attempt to explain the psychological meaning of the construct. Attempting to come up with a unifying definition for creativity appears to be a daunting task, as it has been argued by various researchers that creativity is domain-specific (e.g. Csikszentmihalyi, 1990). Like the existence of challenges over the definition of creativity, there are challenges in the methodologies used to measure it.

Unlike simpler games such as tic-tac-toe or checkers, the numerous permutations of moves available in chess make it impossible to solve, even by the world's most powerful computers. This unsolvability results in each game being different from the previous one, thereby encouraging exploration and discovery. Although the strongest computers are now better at chess than the strongest human chess players, it often has been remarked that computers play a different style of chess that relies on brute-force calculations. In comparison, a more human style of playing chess is often referred to as being more elegant and creative. The majority of chess opening theory and combination patterns on the chessboard have been created by creative and imaginative human players. Even today, although grandmasters regularly use computers as an aid in their opening preparations for tournaments, it is typically human input and intuition that determine the particular key move and branch that are analyzed in greater detail using the power of the computer and modern chess software.

The Torrance Tests of Creative Thinking (TTCT) is the most well-known and widely used test for measuring creativity. The TTCT was developed by Torrance in 1966 (Torrance, 1966). Although primarily designed as an assessment for identifying gifted children, the TTCT is utilized extensively in both the educational field and the corporate world, and it is more widely used and referenced than other measures of creative or divergent thinking.

The original TTCT comprises two components: the TTCT-Verbal and the TTCT-Figural. This study uses the TTCT-Figural test to assess children's creativity as the hypothesis is that if chess has any effects on creativity, they are likely to be manifested in idea-based creativity, rather than verbal creativity. Artistic talent is not required to receive credit on TTCT-Figural tests. The first activity comprises two pages of lines (15 pairs) that the subject must use to create a picture or pictures. The second activity requires the subject to use 10 incomplete figures to make an object or picture. For both activities, the key is to make the lines or incomplete figures part of the drawing. Once the drawing is complete, they are required to add a title that is "clever" and "unusual" to help interpret the drawing. The TTCT-Figural has been found to be fair in terms of gender, race, community status, language background, socioeconomic status, and culture (Cramond, 1993). In the field implementation of the test, the instructions were translated into Bengali and pilot-tested to ensure that students understood the test instructions.

Our scoring of the TTCT-Figural test is based on a 0-3 scale and performed by two markers that were blind to the treatment/control allocation. For both activities, responses that are deemed not creative were provided as guidelines to the markers.[32] Drawings that were deemed not creative were scored a "1" to reflect an attempt made at responding to the question. Other possible scores for each drawing are "2" (somewhat creative) and "3" (creative), and were based on subjective assessments made by the markers. Missing or non-attempts score a '0' as quite a few questions were not answered by the children. For example, about 50% of the children did not attempt to create all 15 drawings in the first activity. The TTCT score we computed was an average of the two markers' scores.

A second creative test that we implemented was Guilford's (1967) alternative uses test, in which examinees are asked to list as many possible uses for a common household item as they can (e.g. brick, shoe, paper clip). The alternative uses test is a standard test of divergent thinking. In our application of the test, participants are asked to list up to 10 alternative uses for a shoelace in a fixed amount of time to gauge both the quantity and novelty of ideas. Our scoring of the alternative-uses test follows the approach we used for the TTCT-Figural test. Responses were scored a "1" (not creative), "2" (somewhat creative),"3" (creative), or "0" for missing answers or non-attempts. The total score was obtained by adding up the points across all answers, with 30 being the maximum possible score. Once again, the score for the alternative-uses test was computed as an average of the two markers' scores.

Creativity Results

Results from the TTCT-Figural test are presented in Table A.1. The average score for the control group is 16.57 for the TTCT pairs-of-lines test and 14.47 for the TTCT picture completion. For both activities, somewhat surprisingly, the control group actually registers slightly higher scores for creativity than the treatment group. However, the differences are not statistically significant. On the other hand, according to the alternate uses test, the treatment group appears to be able to generate more novel ideas. The estimated impact was 0.889 relative to a control group mean score of 14.38. Again, however, the difference in means between the two groups is not statistically significant. Therefore, it appears that chess instruction does not have significant short-term effects on student creativity.

[32] These are based on guidelines provided by Torrance et al. (2008).

Table A.1: Creativity (Wave 1)

Variable	Control Mean	Unadjusted Impact	Regression Adjusted Impact
TTCT pairs of lines	16.57	-0.285 (0.656)	-0.087 (0.572)
		[0.692]	[0.934]
		{0.999}	{0.951}
TTCT picture completion	14.47	-0.554 (1.462)	0.119 (1.311)
		[0.748]	[0.940]
		{0.999}	{0.951}
Guilford's alternate uses test	14.38	0.889 (1.429)	1.727 (1.069)
		[0.580]	[0.150]
		{0.999}	{0.370}
N	223	483	432

Notes: Standard errors in parentheses and are clustered at the school level, with conventional p-values reported as *p-value<0.1 ** p-value<0.05 *** p-value<0.01. The associated wild bootstrapped p-values are reported in square brackets, while false discovery rate (FDR) sharpened q-values (Benjamini et al., 2006) computed using the procedure in Anderson (2008) are reported in curly brackets. Covariates included in the regression adjustment are from Table 1. The wild bootstrap p-values are for the regression adjusted impacts and are based on 1000 replications. Control means are based on the regression adjusted sample. False discovery rate (FDR) sharpened q-values (Benjamini et al., 2006) are computed using the procedure in Anderson (2008).

Appendix B: Attention and Focus

We employ two frequently used tests for the assessment of attention: the digit-cancellation test (Diller et al., 1974) and the digit-symbol test (Wechsler, 1991).

Cancellation tests are used to measure focused and selective attention, speed of information processing, short-term memory, and cognitive flexibility. In the first part of the digit-cancellation test, participants were given a pencil and were asked to use it to cross out all the "8" digits presented on six rows on the form as quickly and as accurately as possible. They were instructed to start with the top row, proceeding from left to right. Upon completion, they were presented with another form, then were asked to cross out all the "5" digits presented over six rows on the form as fast and as accurately as possible. In total, there were 624 digits organized in 12 rows to process. The time to completion together with omission and commission errors were recorded. The cancellation score was computed as the number of targets hit minus the number of errors.

The digit-symbol test is a subtest of the Wechsler Intelligence Test and taps processing speed, visual tracking and scanning, visual-motor coordination, focused and sustained attention, short-term memory, cognitive flexibility and rapid shifting, and the ability to learn a new task. The form consists of four rows of 25 empty boxes in each row, with the first row used for a demonstration and practice trial. Participants were instructed to work as quickly as possible, using a pencil, and go from one box to the next, from left to right.[33] This was a timed test, and the number of correct symbols copied within 120 seconds was recorded in this test.

Results for Attention and Focus

The two tests implemented in Wave 2 of the field experiment are well-known tests that might be able to discern any medium-term effects of playing chess on one's ability to focus and concentrate. These tests were conducted 9-10 months after the conclusion of the chess training program.

Despite anecdotal evidence suggesting that playing chess might improve one's ability to focus, the performance of the treatment and control groups in the Digit Cancellation and Digit Symbol test were very similar, resulting in there being no significant differences in their respective group means in terms of test scores or time to complete the former task (see Table B.1). Our results therefore do not suggest that there are any medium-term effects of chess instruction on attention and focus.

[33] The original instructions were to use 10 of the 25 boxes in the first row as a practice trial. Unfortunately, the field implementation resulted in field workers using the first 25 boxes for the trial.

Table B.1: Digit Cancellation and Digit Symbol Test (Wave 2)

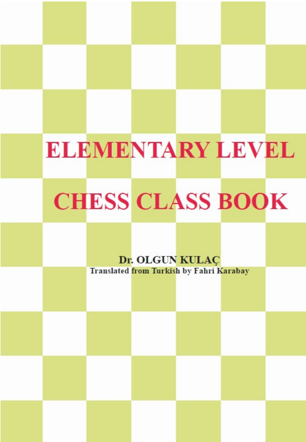
Variable	Control Mean	Unadjusted Impact	Regression Adjusted Impact
Digit cancellation test: Total score	205.89	-0.751 (0.691)	-0.601 (0.566)
		[0.318]	[0.334]
		{0.622}	{0.583}
Digit cancellation test: Time to completion	225.56	9.589 (7.856)	4.892 (6.438)
		[0.242]	[0.492]
		{0.622}	{0.650}
Digit symbol test	32.72	1.579 (1.331)	1.075 (1.326)
		[0.268]	[0.482]
		{0.622}	{0.650}
N	190	425	380

Notes: Standard errors in parentheses and are clustered at the school level, with conventional p-values reported as *p-value<0.1 ** p-value<0.05 *** p-value<0.01. The associated wild bootstrapped p-values are reported in square brackets, while false discovery rate (FDR) sharpened q-values (Benjamini et al., 2006) computed using the procedure in Anderson (2008) are reported in curly brackets. Covariates included in the regression adjustment are from Table 1. The wild bootstrap p-values are for the regression adjusted impacts and are based on 1000 replications. Control means are based on the regression adjusted sample. False discovery rate (FDR) sharpened q-values (Benjamini et al., 2006) are computed using the procedure in Anderson (2008).

ONLINE APPENDIX

(Not for Publication)

Online Appendix 1: Syllabus of the Chess Training Program



CONTENTS

CONTENTS

Available for free download at: http://cis.fide.com/en/teaching-materials

Online Appendix 2: Field Experiment Setting

Online Appendix 3: Attrition

Table C.1: Treatment/Control Balance in Characteristics for Various Samples (*p*-values)

Variable	PSC sample	Risk/Time Wave 1 Sample	Risk/Time Wave 2 Sample	Creativity Tests Sample	Digit Tests Sample
Household income (in takas)	0.832	0.379	0.878	0.240	0.899
Number of household members	0.979	0.896	0.694	0.887	0.711
Sanitary ring latrine in the house	0.194	0.197	0.121	0.177	0.122
Drinking water in the house from tube well	0.243	0.461	0.546	0.479	0.525
Existence of electricity supply in the house	0.388	0.382	0.491	0.386	0.479
Distance of the school from the home (km)	0.243	0.225	0.224	0.150	0.220
Value of total assets except land (in takas)	0.387	0.358	0.336	0.594	0.318
Household religion (Muslim = 1)	0.796	0.909	0.827	0.907	0.820
Do any of the parents know how to play chess	0.239	0.228	0.428	0.473	0.434
Someone with more than grade 10 education in household	0.439	0.778	0.229	0.564	0.166
Father's years of schooling	0.882	0.900	0.834	0.727	0.783
Mother's years of schooling	0.835	0.733	0.768	0.989	0.760
Father works as labourer/in agriculture	0.269	0.385	0.529	0.253	0.555
Mother is a housewife	0.538	0.393	0.405	0.325	0.411
Two-parent household	0.433	0.508	0.746	0.664	0.741
Father's age	0.682	0.504	0.429	0.364	0.434
Mother's age	0.853	0.727	0.814	0.553	0.817
Gender of student (male =1)	0.220	0.530	0.111	0.347	0.113
N	395	450	381	432	380

Notes: Sample sizes are based on the regression adjusted samples for each outcome, with standard errors clustered at the school level. **p*-value<0.1 ** *p*-value<0.05 *** *p*-value<0.01.

Table C.2: Treatment/Control Balance in Characteristics for Various Samples (Means)

	PSC Sample			Risk/Time Wave 1 Sample		
Variable	Treat Mean	Control Mean	Diff	Treat Mean	Control Mean	Diff
Household income (in takas)	8580.4	8439.4	141.0	8197.7	8702.4	-504.7
Number of household members	4.35	4.21	0.14	4.36	4.26	0.1
Sanitary ring latrine in the house	0.62	0.63	-0.01	0.64	0.65	-0.01
Drinking water in the house from tube well	0.59	0.84	-0.25	0.63	0.85	-0.22
Existence of electricity supply in the house	0.331	0.479	-0.148	0.35	0.49	-0.14
Distance of the school from the home (km)	1.05	0.661	0.389	1.01	0.64	0.37
Value of total assets except land (in takas)	77604.8	62721.0	14883.8	74395.5	60680.0	13715.5
Household religion (Muslim = 1)	0.94	0.93	0.01	0.92	0.93	-0.01
Do any of the parents know how to play chess	0.102	0.063	0.039	0.106	0.066	0.04
Someone with more than grade 10 education in household	0.151	0.121	0.03	0.13	0.14	-0.01
Father’s years of schooling	4.33	4.22	0.11	4.28	4.36	-0.08
Mother’s years of schooling	4.27	4.12	0.15	4.26	4.01	0.25
Father’s age	39.78	40.04	-0.26	39.62	40.02	-0.4
Mother’s age	33.51	33.63	-0.12	33.43	33.65	-0.22
Father works as labourer/in agriculture	0.66	0.57	0.09	0.662	0.595	0.067
Mother is a housewife	0.95	0.96	-0.01	0.955	0.968	-0.013
Two-parent household	0.98	0.97	0.01	0.982	0.973	0.009
Gender of student (male =1)	0.51	0.41	0.1	0.48	0.44	0.04
N	205	190	395	225	225	450

Table C.2 (continued)

	Risk/Time Wave 2 Sample			Creativity Tests Sample		
Variable	Treat Mean	Control Mean	Diff	Treat Mean	Control Mean	Diff
Household income (in takas)	8378.9	8481.9	-103	8136.7	8799.3	-662.6
Number of household members	4.36	4.27	0.09	4.35	4.26	0.09
Sanitary ring latrine in the house	0.62	0.65	-0.03	0.64	0.63	0.01
Drinking water in the house from tube well	0.61	0.86	-0.25	0.62	0.85	-0.23
Existence of electricity supply in the house	0.37	0.47	-0.1	0.33	0.47	-0.14
Distance of the school from the home (km)	1.10	0.61	0.49	0.94	0.61	0.33
Value of total assets except land (in takas)	73989.4	61303.6	12685.8	70287.0	64062.2	6224.8
Household religion (Muslim = 1)	0.93	0.92	0.01	0.928	0.933	-0.005
Do any of the parents know how to play chess	0.094	0.068	0.026	0.103	0.076	0.027
Someone with more than grade 10 education in household	0.153	0.120	0.033	0.139	0.158	-0.019
Father's years of schooling	4.34	4.18	0.16	4.21	4.46	-0.25
Mother's years of schooling	4.32	4.09	0.23	4.14	4.13	0.01
Father's age	39.51	39.99	-0.48	39.65	40.19	-0.54
Mother's age	33.39	33.54	-0.15	33.49	33.87	-0.38
Father works as labourer/in agriculture	0.65	0.59	0.06	0.67	0.59	0.08
Mother is a housewife	0.95	0.97	-0.02	0.95	0.97	-0.02
Two-parent household	0.98	0.97	0.01	0.98	0.98	0
Gender of student (male =1)	0.53	0.40	0.13	0.49	0.42	0.07
N	190	191	381	209	223	432

Table C.2 (continued)

	Digit Tests Sample		
Variable	Treat Mean	Control Mean	Diff
Household income (in takas)	8378.9	8463.4	-84.5
Number of household members	4.36	4.27	0.09
Sanitary ring latrine in the house	0.62	0.65	-0.03
Drinking water in the house from tube well	0.61	0.86	-0.25
Existence of electricity supply in the house	0.37	0.48	-0.11
Distance of the school from the home (km)	1.10	0.61	0.49
Value of total assets except land (in takas)	73989.4	60836.8	13152.6
Household religion (Muslim = 1)	0.93	0.92	0.01
Do any of the parents know how to play chess	0.094	0.068	0.026
Someone with more than grade 10 education in household	0.152	0.116	0.036
Father's years of schooling	4.34	4.14	0.2
Mother's years of schooling	4.33	4.08	0.25
Father's age	39.51	39.89	-0.38
Mother's age	33.39	33.53	-0.14
Father works as labourer/in agriculture	0.65	0.60	0.05
Mother is a housewife	0.95	0.97	-0.02
Two-parent household	0.98	0.97	0.01
Gender of student (male =1)	0.53	0.40	0.13
N	190	190	380

Online Appendix 4: Risk Preference Task

(All tasks described in this appendix are conducted using the Bengali version.)

(A) Wave 1 task

Each student is handed and read out the preference test sheet.

Instructions

In this activity, you have to choose 1 option from 5 different options. There is no right or wrong option. You should choose the option that you like the most. Please circle your chosen option number.

For each option, there are two possible outcomes: "Heads" or "Tails". After everyone has made their choice, a coin will be flipped to determine which outcome occurs. If the coin turns out Heads, then you will receive the number of stationary items under the Heads column corresponding to your choice. If the coin is Tails, you will receive the number of stationary items under the Tails column corresponding to your choice.

Here is an example:

Option Number	If the coin is Heads I get	If the coin is Tails I get
1	4 items	4 items
2	6 items	3 items
3	8 items	2 items
4	10 items	1 items
5	12 items	0 items

Benu has chosen option number 5 by circling the number '5'. The teacher then flips a coin and it shows Heads. This means Benu will receive 12 items as depicted on the stationary sheet. If the coin had shown Tails Benu would have got nothing.

Before making your decision, please answer the following question using the table above:
If I choose Option Number 1, and the coin is flipped and it turn out to be Tails I will receive $\boxed{}$ items from the stationary box.

(*Teacher or person in charge checks all answers first before proceeding*).

Please make your decision now by circling the Option Number that you choose:

Option Number	If the coin is Heads I get	If the coin is Tails I get
1	4 items	4 items
2	6 items	3 items
3	8 items	2 items
4	10 items	1 items
5	12 items	0 items

1	**1**
	
2	**2**
	
3	**3**
	

4	4
	
5	**5**
SüPER Marshal SET OF MATHEMATICAL INSTRUMENTS CLASSIC & PRECISION	

(B) Wave 2 task

In this activity, you will have the opportunity to earn some **tokens**. [show picture of tokens]

These tokens can be exchanged for items from these two bags. [show two bags with items]

Items in the small bag cost two tokens. [show the items that are available in the small bag – take the time to ensure each child is aware of the variety of items in the bag]

Items in the big bag cost five tokens. [show the items that are available in the large bag – take the time to ensure each child is aware of the variety of items in the bag]

You can also exchange 1 token for 1 chocolate gold coin. [show chocolate gold coin]

There are multiple copies of each item -- you will have ample opportunity to exchange for the item you want once the tasks are complete.

Items costing 5 tokens

Items costing 2 tokens


Items costing 1 token

In this activity, you have to choose 1 option from 6 different options. There is no right or wrong option. You should choose the option that you like the most. Please circle your chosen option number.

For each option, there are two possible outcomes: "Heads" or "Tails". After everyone has made their choice, a coin will be flipped to determine which outcome occurs. If the coin turns out Heads, then you will receive the number of tokens shown under the Heads column corresponding to your choice. If the coin is Tails, you will receive the number of tokens under the Tails column corresponding to your choice.

Here is an example:

Option Number	If the coin is Heads I get	If the coin is Tails I get
1	5 tokens	5 tokens
2	7 tokens	4 tokens
3	9 tokens	3 tokens
4	11 tokens	2 tokens
5	13 tokens	1 tokens
6	15 tokens	0 tokens

Abida has chosen option number 5 by circling the number '5'. The teacher then flips a coin and it shows Heads. This means Abida will get 13 tokens. If the coin had shown Tails Abida would have gotten 1 token.

Before making your decision, please answer the following question using the table above: If I choose Option Number 2, and the coin is flipped and it turn out to be Tails I will receive ☐ tokens.

[*Teacher or person in charge checks all answers first before proceeding*].

Please make your decision now by circling the Option Number that you choose:

Option Number	If the coin is Heads I get	If the coin is Tails I get
1	5 tokens	5 tokens
2	7 tokens	4 tokens
3	9 tokens	3 tokens
4	11 tokens	2 tokens
5	13 tokens	1 tokens
6	15 tokens	0 tokens

Description of Risk Preference Variables

Variables	Range	Description
Wave 1		
Risk	1-5	Number indicates alternative chosen in the risk task, with a higher number corresponding to a higher risk
Wave 2		
Risk	1-6	Number indicates alternative chosen in the risk task, with a higher number corresponding to a higher risk

Online Appendix 5: Time Preference Task

(All tasks described in this appendix are conducted using the Bengali version.)

(A) Wave 1 task

Instructions

There are 2 parts to this activity. In each part, you have to make 5 decisions. This means in total, you will make 10 decisions. Each of these decisions involve choosing whether you prefer to receive some candy earlier, or later. Once all students have completed all the activities for today, one of your 10 decisions from the two parts in this activity will be chosen. You will only be able to collect the candy for the chosen decision. Which decision is chosen will be determined by randomly drawing a piece of paper from a jar. The jar will contain 10 pieces of paper numbered from 1 to 10. So, for example, if the piece of paper drawn shows '6' then everyone will be able to collect pens based on their choice in decision number 6.

Part One

In this part of the activity you have to make 5 different decisions. There is no right or wrong answer for each decision. You should choose the option that you like the most.

For each of the 5 decisions, you choose to receive some candy either tomorrow (*state day/date*) (Earlier) or in 8 days (*state day/date*) (Later).

Here is an example:

☒	☐
Earlier: Receive 4 candies tomorrow	Later: Receive 8 candies in 8 days
	

Benu is deciding what to do for her first decision. She is choosing between receiving 4 pieces of candy tomorrow (Earlier) versus 8 pieces of candy in 8 days (Later). After some thinking, she decides that she prefers to have 4 pieces of candy tomorrow. She puts a cross X on the box corresponding to Earlier as shown above.

Please make your decisions by putting a cross X on the box corresponding to your decision.

Decision 1

☐ Earlier: Receive 4 candies tomorrow 	☐ Later: Receive 4 candies in 8 days 

Decision 2

☐ Earlier: Receive 4 candies tomorrow 	☐ Later: Receive 6 candies in 8 days 

Decision 3

☐ Earlier: Receive 4 candies tomorrow 	☐ Later: Receive 8 candies in 8 days 

Decision 4

Earlier: Receive 4 candies tomorrow 	Later: Receive 10 candies in 8 days 

Decision 5

Earlier: Receive 4 candies tomorrow 	Later: Receive 12 candies in 8 days 

Part Two

Part 2 of this activity (not shown) uses the same figures, where the only difference is that the ‘early’ option is now 8 days and the ‘later’ option is 15 days. As before, students have to make 5 decisions here.

(B) Wave 2 task

In this activity, you will have the opportunity to earn some **tokens**. [show picture of tokens]

There are 2 parts to this activity. In each part, you have to make 3 decisions. This means in total, you will make 6 decisions.

Once all students have completed all the activities for today, one of your 6 decisions will be chosen and you will receive tokens based on those decisions. You will only receive tokens for the one decision that was chosen. Which decision is chosen will be determined by randomly drawing a piece of paper from a jar. The jar will contain 6 pieces of paper numbered from 1 to 6. So, for example, if the piece of paper drawn shows '6' then everyone will receive tokens based on their choice in decision number 6.

Part One

In this part of the activity you have to make 3 different decisions. There is no right or wrong answer for each decision. You should choose the option that you like the most.

For each of the 3 decisions, you must choose 1 from 5 options. Each option gives you a different number of tokens today (earlier) and in 7 days (later). Tokens that you receive today can be exchanged immediately for items from the bags. Tokens that you receive in 7 days can be exchange for items from the bag when you receive them. We will refill the bags after today so that it will contain the same items that you see today. Therefore, the things that you can exchange your tokens for in 7 days will be the same as what you can exchange for today.

Here is an example:

Example

Option Number	TOKENS RECEIVED TODAY	TOKENS RECEIVED IN 7 DAYS
1	8 tokens	0 tokens
2	6 tokens	3 tokens
3	4 tokens	6 tokens
4	2 tokens	9 tokens
5	0 tokens	12 tokens

Abida is deciding what to do for her first decision. If she chooses Option 1, she will receive 8 token today and no tokens in 7 days. If she chooses Option 4, she will receive only 2 tokens today, but also 9 tokens in 7 days. After some thinking, she decides to choose Option 2 to receive 6 tokens today, and 3 tokens in 7 days. She chooses Option 2 by circling the number '2'.

Please make your decisions by circling the Option Number that you choose for each of the following 3 decisions.

Decision 1

Option Number	TOKENS RECEIVED **TODAY**	TOKENS RECEIVED IN **7 DAYS**
1	12 tokens	0 tokens
2	9 tokens	3 tokens
3	6 tokens	6 tokens
4	3 tokens	9 tokens
5	0 tokens	12 tokens

Decision 2

Option Number	TOKENS RECEIVED **TODAY**	TOKENS RECEIVED IN **7 DAYS**
1	12 tokens	0 tokens
2	9 tokens	4 tokens
3	6 tokens	8 tokens
4	3 tokens	12 tokens
5	0 tokens	16 tokens

Decision 3

Option Number	TOKENS RECEIVED **TODAY**	TOKENS RECEIVED IN **7 DAYS**
1	12 tokens	0 tokens
2	9 tokens	5 tokens
3	6 tokens	10 tokens
4	3 tokens	15 tokens
5	0 tokens	20 tokens

Part Two

Part 2 of this activity (not shown) uses the same figures, where the only difference is that the choice is now between receiving tokens in 7 days or 14 days. As before, students have to make 3 decisions here.

Description of Time Preference Variables

Variables	Range	Description
Wave 1		
Impatience; Impatience Delayed	0-5	Participants make five decisions over 'earlier' or 'later'. A score of 1 is assigned if the 'earlier' alternative is chosen. The final number is the sum of these scores; e.g. 0 indicates the participant always chose 'later'.
Time Inconsistency	0-5	Participants make the original five decisions over a standard, and delayed set. A score of 1 is assigned if the choice is the standard set does not correspond with the delayed set. The final number is the sum of these scores; e.g. 0 indicates that all choices in the standard set always corresponds with the delayed set.
Time Inconsistency (binary)	0-1	A 1 indicates that at least one of the choices in the standard set does not correspond to the delayed set.
Non-monotonicity (binary)	0-1	A 1 indicates either someone who switched from 'later' to 'earlier' as the interest rate increased, and/or someone who choose 'later' at an interest rate of 0%.
Wave 2		
Impatience; Impatience Delayed	2-10	Participants have two choice sets, with five alternatives in each. Each choice set receives a score 1-5 based on the alternative chosen, with a higher score indicating a more impatient choice. The final number is the sum of these two scores; e.g. 2 indicates the participant chose the most patient choice across both sets.
Time Inconsistency	0-2	A score of 1 is assigned for each standard choice set where the alternative chosen does not correspond to the delayed choice set. The final number is the sum of these two scores; e.g. 0 indicates both choices in the standard set corresponds with the delayed set.
Time Inconsistency (binary)	0-1	A 1 indicates that at least one of the choices in the standard set does not correspond to the delayed set.
Non-monotonicity (binary)	0-1	A 1 indicates someone who switched to a more impatient choice as the interest rate increased