Children, Chess, and Mathematics: Experimental Evidence from Bangladesh

Asadul Islam
Monash University

Wang-Sheng Lee*
Deakin University and IZA, Bonn

Aaron Nicholas
Deakin University

(This version: 30 April 2018)

Abstract

We investigate the benefits of an intensive 30 hour chess training program on students starting grade five in rural Bangladesh using a randomized experiment. A key outcome we focus on are academic results from a compulsory public exam all grade five students in Bangladesh must sit for – the Primary School Certificate (PSC) exam. The exam took place 9-10 months after the completion of the chess training. While the previous literature has emphasized potential links between chess and academic outcomes, a novel contribution of this paper is consideration for the following non-cognitive outcomes: risk preferences, patience, creativity and attention/focus. We find that chess training has a moderately significant positive impact on math scores in the national exam and reduces the treatment group's level of risk-aversion. While the program has insignificant effects on students' level of impatience, it reduces the incidence of time-inconsistency as well as the incidence of non-monotonic time-preferences. A potential mechanism by which the effect on math occurs is by reducing the treatment group's risk-aversion. Effects of chess training on the other academic outcomes and creativity are not significant.

JEL codes: C93, D80, I21, I25

Keywords: randomized trial, non-cognitive skills, chess, creativity, math, patience, risk

We are grateful to Deakin University and Monash University for generous research funding. We are grateful to Liang Choon Wang for helpful comments and Foez Mojumder for providing excellent research support. We would like to thank participants at the Labour Econometrics Workshop 2017 in Auckland, the 2017 Asian and Australasian Society of Labour Economics (AASLE) conference in Canberra, and workshops at Deakin University and Singapore Management University for comments. This research would also have not been possible without cooperation from the Department of Primary Education (DPE) in Bangladesh, local school teachers and administrative staff, NGO partner Global Development Research Initiative (GDRI), the hard work and support of field staff in Bangladesh, and our chess coaches. * Corresponding author. E-mail: wang.lee@deakin.edu.au.

1. Introduction

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 labour market success (Heckman et al., 2006; Carneiro et al., 2007). The evidence from a large number of recent studies on the importance of non-cognitive skills has drawn the attention of policymakers who are interested in identifying programs that can help develop such important 'soft' life skills in children in their formative years. A recent review by Gutman and Schoon (2013), funded by the Education Endowment Foundation (EEF) and Cabinet Office in the UK, summarizes the existing evidence on how non-cognitive skills can be defined and measured and assesses the role of select interventions that aim to improve non-cognitive skills in children and young people.

In this paper, we add to the literature on interventions that can potentially help nurture non-cognitive skills and improve cognitive skills in children. Specifically, we investigate whether undertaking intensive chess lessons can affect children's academic performance, risk preferences, time preferences, creativity, and attention/focus. In order to estimate causal effects of chess training, we conduct a randomized controlled trial (RCT) among rural primary school children in Bangladesh. Estimated impacts on academic outcomes are based on results from a compulsory national exam for primary school students in Bangladesh. We examine both short-run effects, based on assessments that were made shortly after the conclusion of the training program, as well as medium-term effects that were based on assessments that were conducted 9-10 months after the program ended. We also add to the debate regarding the nature of the synergistic relationship between cognitive and non-cognitive skills by exploring alternative causal pathways to help decompose the effects of the chess training program.

Chess is a popular game and played by millions of people worldwide. It is a game with very low barriers to entry and is played across the socio-economic spectrum. While the rules of the two-player strategic board game are easy to learn, it is difficult to master. Becoming a good chess player involves understanding the nuances of the different phases of the game (opening, middlegame, endgame), learning how to attack and when to take calculated risks, learning how to think strategically and creatively, and learning how to be patient and remain focused on the end goal (checkmating the opponent's king).

It is therefore not surprising that chess is often portrayed as a useful education tool. The European Parliament has expressed its favourable opinion on using chess courses in schools as an educational tool (Binev et al., 2011) and chess has even become part of the formal primary school curriculum in Armenia in 2011 and Poland in 2017.[1] Several organizations in the US have helped spur an increasing interest in formal chess instruction. America's Foundation for Chess (AF4C) introduced the First Move chess program into elementary school classrooms as a tool for enhancing critical and creative thinking skills. In 2014, the program reached over 80,000 students in 24 states and four countries. Since 1986, a non-profit organization "Chess in the Schools" has taught, inspired, and empowered more than 500,000 students in low-income New York City public schools.

The popularity of chess as an education tool is at least in part attributable to its perceived effect on cognitive skills in general and mathematics in particular. Studies such as Scholz et al. (2008) and Trinchero and Sala (2016) suggest that chess improves children's mathematical skills because the game has some elements in common with the mathematical domain and because it promotes suitable habits of mind. In 2014, School Library Journal's best education pick of the year was a chess related product called Yamie Chess that is backed by Harvard and MIT academics. Yamie Chess features an interactive coloring math comic book written by experienced math teachers for K-8 supplemental math learning. In a meta-analysis, Sala and Gobet (2016) quantitatively evaluate the available empirical evidence examining whether skills acquired during chess instruction in schools positively transfer to mathematics, reading and general cognitive skills. Their results suggest that chess instruction improves children's mathematical, reading, and cognitive skills moderately, with a tendency for a stronger effect on mathematical than reading skills. However, most previous studies are based on non-experimental data or have small sample sizes.[2] An exception is Jerrim et al. (2016, 2018) who report results from a large-scale clustered randomized field experiment conducted in the UK evaluating the impact of teaching Year 5 school children how to play chess. They found no evidence that the program raised children's attainments in mathematics. There were also no impacts on their secondary outcomes reading and science.

Similar to Jerrim et al. (2016, 2018), we also conducted a randomized experiment on children in grade five. One difference is that the children in our experiment come from rural primary schools who do not have previous experience playing chess. Jerrim et al. (2016: 46) report in their study that chess playing activity at their baseline was 48% in treatment schools

[1] For 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).

[2] As reported in Jerrim et al. (2018), the median sample size of previous studies conducted is just 54 children.

and 45% in control schools. Our setting seems particularly well suited to test the benefits of a chess training program as unlike children in an urban area, the majority of children will never have been exposed to the game of chess before or similar cognitively demanding games.

The primary outcome we examine are test scores from a standardized compulsory public exam all grade five students in Bangladesh must sit for – the Primary School Certificate (PSC) exam – which took place 9-10 months after the completion of the chess training. We are particularly interested in examining the effects on math test scores because of the perceived benefits between chess and mathematics. While an emphasis will be placed on examining their results for mathematics, their first language and science (as in Jerrim et al., 2016, 2018), we are also able to examine the test scores on other subjects as the PSC exam comprises six mandatory subjects in total – Bengali, English, science, social science, mathematics and religion. In addition, in an effort to provide a rigorous assessment of the links between chess, cognitive skills and non-cognitive skills, extensive post-program data were collected to test for the effects of chess training on non-cognitive outcomes. We investigate whether undertaking intensive chess lessons can affect children's risk preferences, time preferences, creativity, and attention/focus.

The literature offers no concrete evidence regarding the nature of the relationship between cognitive and non-cognitive skills. Cunha and Heckman (2008) show that non-cognitive skills promote the formation of cognitive skills. However, Burks et al. (2009), Dohmen et al. (2010) and Benjamin et al. (2013) find evidence for an alternative pathway – that higher cognitive skills systematically affect non-cognitive skills elicited through risk taking and time discounting tasks. The synergy between cognitive and non-cognitive skills and how they interact dynamically to shape the evolution of subsequent capabilities is of great interest to policy makers who are interested in designing a modern school curriculum relevant for tomorrow's workforce.

To further explore the synergistic relationship between cognitive and non-cognitive skills, using the framework described in Imai et al. (2010, 2013), we conduct a causal mediation analysis to assess the potential pathways that chess training has on our measured outcomes. We examine both possible pathways – how chess training might have an effect on academic outcomes that are mediated by non-cognitive outcomes, or alternatively how chess training might have an effect on non-cognitive outcomes that are mediated by academic outcomes.

We find that chess training has a moderately significant positive impact on math scores in the national exam. Despite having no significant effect on impatience, it reduces risk-aversion, the incidence of time inconsistency, as well as the incidence of non-monotonic time preferences. Similar patterns with regards to the results on time preferences are found in Lührmann et al. (2015), where the intervention was financial education rather than chess education. Effects of chess training on the other academic outcomes, creativity and attention/focus were not significant. In exploring causal pathways, we find that an important mechanism by which the effect of chess training on academic outcomes occurs is by decreasing the treatment group's risk-aversion. We find weaker evidence for the alternative pathway that higher cognitive skills systematically affect non-cognitive skills.

One novel contribution of this paper is a focus on the link between chess and non-cognitive outcomes, an issue which is generally ignored in the previous literature. We focus on measures of non-cognitive skills that are relevant to chess playing and labor market outcomes: risk, time preferences, patience, creativity, attention and focus. A second contribution is with regards to attempting to better understand the evolution of cognitive and non-cognitive skills in children in response to an intervention that has the potential to help develop both sets of skills. We are not aware of any other work that uses formal mediation analysis to help better appreciate the underlying mechanisms that might explain why chess might be of benefit to children. A third contribution is to examine the effectiveness of chess instruction in a developing country context. Our study focuses upon the impact of chess instruction on cognitive and non-cognitive outcomes 9-10 months after the intervention has finished, allowing us to examine if there is a lasting effect. In terms of academic outcomes, as in Jerrim et al. (2018), 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." To our knowledge, ours is the first RCT to be conducted examining the lasting impact of chess instruction in a developing country context.

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 cognitive and non-cognitive outcomes measured in this study. Section 5 outlines our empirical strategy to estimate causal effects and possible causal mechanisms. Section 6 presents the results of the intervention. Finally, Section 7 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 skill transfers to other domains of learning. Although near transfer might be possible (where transfer occurs between closely related domains such as maths and physics), several studies have shown that chess players' skill tends 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 where participants had to solve 16 four-, five-, and six-move problems each, chess planning skills did not improve the ability of chess players to solve the task (Unterrainer et al., 2011). Chess skill also does not predict performance in the economic game known as the ‘beauty contest’ (Bühren and Frank, 2010). In this game, participants were asked to state a number (not necessarily an integer) between 0 and 100, the winning number being the one closest to two-thirds of the average. The average guess was 32.15; chess Grandmasters’ average guess was 32.96. However, Levitt et al. (2011) find that stronger chess players (grandmasters) perform better in backward induction games such as the race to 100 relative to weaker chess players.

We are not aware of any work that has explored in-depth the link between chess skills and non-cognitive skills, although some previous work has focused on the effects of chess on focused attention and metacognition (Scholz et al., 2008; Kazemi et al., 2012). This is despite an observation made more than two centuries ago from a notable chess enthusiast, the renowned inventor Benjamin Franklin who wrote in “The Morals of Chess” (1786) that:

> “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 (considers the consequences that may attend an action), circumspection (surveys the whole chess-board, the relations of the several pieces and situations, and the dangers they are respectively exposed to) and caution (not to make moves too hastily and to abide all the consequences of one’s rashness). Circumspection implies a person thinks carefully before doing or saying anything and is expected to be correlated with patience. Combining foresight and caution implies a person will learn to make calculated risks, hence linking chess playing style and skill with risk preferences.

Some suggested benefits of chess to children are highlighted in *Brooklyn Castle*, a 2012 documentary film about an after-school chess program in Intermediate School 318, an inner-city public school in Brooklyn, New York. The film follows five of the school's chess team members for one year, and documents their challenges and triumphs both on and off the chessboard, and how chess helped them with their self-esteem and made them more motivated learners. Their chess team became the first middle school team to win the United States Chess Federation's national high school championship.

3. The Intervention

The intervention took place in primary schools in remote rural communities in two districts in the southwest of Bangladesh in January/February 2016. Our chess experiment is a clustered randomized controlled trial with randomization at the school level. The experiment involved nearly 600 students starting grade five in 2016 in 16 primary schools located in two districts of southern Bangladesh: Khulna and Satkhira. These schools were first 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. One of the co-authors had spent his childhood and attended primary and secondary schools 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 schools, and district level administration that facilitated logistics for implementing the intervention.

The 16 schools were randomly divided into two groups: eight in the treatment group and eight in the control group. Students in the treatment schools received 12 days of chess training (spread over three weeks).

In the design phase of the project, while randomization at the class level was considered and deemed to be preferable, it was ruled out for several reasons. First, there is the possibility of contamination between treatment and control classes. Second, there are organizational issues that may arise. When one class is receiving the intervention, students from other classes might also want to join in. Third, most schools in rural Bangladesh only had one class of students for each grade.

A pre-program baseline test of chess knowledge suggests that the majority of children in our analysis sample did not know how to play chess. In response to the survey question: "do you know how to play chess?" 8.4% in the treatment group and 4.7% in the control group responded that they knew how to play chess. When asked further specific questions about how chess pieces move and capture in two chess positions that were provided in diagrams, it emerged that 4.0% in the treatment group and 2.3% in the control group knew the chess rules but this difference was not statistically significant (p-value = 0.252). Training sessions were conducted separately in each school and carried out at the beginning of the academic year in January and February of 2016. The chess instruction involved teaching the rules of chess and basic chess strategy.

The lesson plan was based on free instructional chess material available from the Chess in Schools Commission of the World Chess Federation (FIDE) (see Appendix 1 for the syllabus used for the chess lessons). We hired two instructors in total to deliver the entire chess program to the eight treatment schools. Both instructors are qualified chess coaches and have extensive experience in teaching chess to children. One is a FIDE master and former national champion of Bangladesh. The other is 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.

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 a number of field staff who are amateur chess enthusiasts. During the training sessions each pair of students received a chess set to practice in class. The chess classes were held either before or after school.[3] The intervention involved providing a total of 24 hours of chess instruction (12 days times two hours per lesson) and about 6 hours of supervised chess practice playing against an opponent, which allowed the students to apply any new skills that was just covered. Therefore, in total, the program consisted of approximately 30 hours of chess training. This is above the 25 hours Sala and Gobert (2016) report as the threshold above which chess instruction produces substantial effects.

Student feedback on the chess lessons was very positive. Of the 248 (out of 294) respondents in the treatment group that provided feedback on the chess lessons, 100% of the students responded that they liked playing chess and 99.2% said they would like more chess lessons. 94.5% of the children said that in Week 1, they played or discussed chess with at least one classmate during a time outside of the chess program; this percentage remained high in Week 2 (87.5%). The chess sets used in the training program were donated to each respective

[3] Some pictures of the field setting can be found in Appendix 2, where normal classrooms have been used to conduct the chess lessons. Some schools have double shifts, and grade five students start classes after 11 am in those schools. Students in the control group schools spent their time before and after school engaged in their usual unstructured play activities. We conducted chess lessons in those schools in the morning. It is to be noted that the school curriculum during the start of the school year (January and February) is relatively light and teachers are not always teaching academic subjects during school time during this period of the year. Our timing of the intervention was planned to minimize pressure on students from their day-to-day academic class lessons.

school at the conclusion of the three-week training program in order to facilitate the children continue playing and practising chess after lessons had ended. The students' interest in chess do not appear to be transitory. When we checked to see if treatment group members were still playing chess 9-10 months later, we found that 94.3% of them had played chess with a classmate in the past week, and that 87.5% of them had played chess with other friends or relatives in the past week.

Before the chess training program launched, a household survey was carried out in November and December 2015 to collect some basic household information and the demographic profile of children and their parents; the respondents were the parents of the children participating in the chess experiment. We also tested their pre-program math skills and chess knowledge. At the conclusion 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 1 describes the key timelines of the project. Short-run outcomes (Wave 1) were measured at the conclusion of the three-week chess training program (the day after) and longer-term outcomes (Wave 2) were measured about 9-10 months after the training concluded – at the end of October 2016. We also assessed if the program had an impact on academic performance based on results from a national exam students in grade five had to sit for during the period November 20-27, 2016.

3.1 Baseline Balance

Using data from the baseline household survey data carried out in November and December 2015, we examined a list of household characteristics to determine if there were any systematic observed differences between the treatment and control groups. This includes the following: household income, the number of household members, whether there is a sanitary ring latrine in the house, whether drinking water in the house is from a tube well, whether there is electricity supply in the house, the distance of the school from the home, the value of total assets except land, household religion, whether either of the parents know how to play chess, whether there is someone with more than a grade 10 education in the household, whether father works as labourer/in agriculture, whether mother is a housewife, whether family is a two-parent household.

Table 1 presents the differences in means of household characteristics for the treatment and control groups. None of the differences are statistically significant, suggesting that the randomization process was well implemented.

In general, it can be seen that the children are mostly underprivileged and have 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 grade 10. As far as occupations are concerned, 64% of fathers are engaged in agriculture or day labor, another 29% in small business activities and 6% in services; almost all mothers are housewives. In our sample, the average household size is 4.3, and the average monthly income is less than 8500 takas (about US$110).

We collected baseline data for 294 treatment group students and 300 control group students in November 2015 but there was some data loss over time as we measured outcomes in Waves 1 and Wave 2. For example, for the PSC exam taken about one year later, we only had results for 220 treatment group students and 214 control group students. In general, attrition did not pose a problem for the integrity of the experimental design. As Table E.1 in Appendix 5 shows, there were no significant mean differences in characteristics between the treatment and control groups in any of the outcomes examined. Comparing the sample we have follow-up data on outcomes for with those from the initial sample that we have lost to attrition, we find that there are no significant differences in characteristics. This suggests that attrition in our sample is not systematically related to any particular set of characteristics and is likely to be random.

4. Outcomes

4.1 Academic Outcomes

We use exam marks from the Primary School Certificate (PSC) that is administered nationwide annually in Bangladesh to all grade five students as the primary outcomes for cognitive abilities. The PSC is a written examination, administered face-to-face and delivered through paper-pencil tests at the end of grade five. This exam took place in November 2016 and was approximately 9 to 10 months after the conclusion of the chess program. The PSC comprises six mandatory subjects – Bengali, English, science, social science, mathematics and religion. All subjects are administered in Bengali, except for the English language test which is conducted in English.

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%.[4]

As we had a particular interest in the potential links between chess and math, in addition, two separate mathematics tests were developed to measure students' math skills before and after the chess training. The tests intended to assess problem-solving capacities in mathematics, 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). In order to develop the tests, the local mathematics textbook for grade four students in Bangladesh were consulted. Local school teachers and educators were also consulted to help develop the test. The tests were conducted to assess both content and cognitive domains of students. Content domains include: addition, subtraction, multiplication and division, (including money and product transactions), fractions, geometric skills, and reading, comparing and interpreting graphical representation of data. As our analysis sample are students from rural areas, with students generally coming from poorer socio-economic backgrounds with lower academic knowledge than their urban counterparts, we factored into consideration the background of the students when designing the tests.

4.2 Risk Preferences

Risk-aversion is a trait that is typically associated with welfare-relevant later life outcomes. Hence, its detection (and potential manipulation) from an early age may be of particular policy interest. For example, Cohn et al. (1975) find that households that tend to acquire wealth are those that are relatively less risk averse to begin with. Risk aversion is also shown to predict key household decisions such as choice of occupation, portfolio selection, moving decisions and exposure to chronic diseases in ways consistent with theory (Guiso and Paiella, 2008). Uncertainty is inevitable from birth and learning how to take calculated risks is important in many life situations. Davis and Eppler-Wolff (2009) argue that parents need to understand the significance of risk-taking as a teaching experience for children.

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.

[4] Due to privacy reasons, we were not able to access the numeric scores awarded to each student for each of the exams taken. However, we were able to obtain the letter grades awarded to each 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, F = 0 points.

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.[5] 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 3 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. Hence, 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 risker, 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) increase with the level of risk. Hence, a risk-neutral or risk-loving person should always choose 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 3.

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

In the second wave conducted in late October 2016, the same task was used with two changes made. First, in order 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 which could be used to purchase several new attractive items (see the second part of Appendix 3). 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 to investment remains the same as in the first wave. Details of the task undertaken in Wave 2 can also be found in Appendix 3.

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

4.3 Time Preferences

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 – behaviour that has been shown to be predictive of economic success.

Time preferences were elicited in both waves and at the same time as risk preferences, with the order of the two tasks being reversed across waves. In the first wave (January/February 2016), we used a multiple price list format as popularised by Coller and Williams (1999). Unlike risk preferences, it is not common to find single-decision implementations of elicitation tasks for time preferences. Additionally, it is common for the multiple price list format to be implemented on children.[6]

In this task, students make five decisions; in each decision they choose between receiving four pieces of candy tomorrow ('earlier'), versus receiving x pieces of candy in eight days ('later'), where $x \in \{4,6,8,10,12\}$. This is close to the design adopted in Alan and Ertac (2017), where the choice was between two gifts today, versus y pieces one week later, where $y \in \{2,4,6,8,10\}$. We chose candy to differentiate 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 incentivise time-preference elicitation of children 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 were presented on the same day the task was performed. In doing so, the aim is to minimize any apparent impatience that would have arisen from a lack of trust of the experimenters, or any psychological discontinuities that may arise from imagining payment in the future versus 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 remain identical with the only difference being that the earlier alternative was paid out in eight days, and the later alternative, 15 days. This delay is similar to the seven days (earlier), 14 days (later) implementation used in Alan and

[6] For example, Bettinger and Slonim's (2007) study involved children aged 5-16 in the USA; Castillo et al.'s (2011) analysis involved children aged 13-14 in the USA; Sutter et al.'s (2013) study involved children aged 10-18 in Austria; Alan and Ertac's (2017) study involved children aged 9-13 in Turkey.

Ertac (2017). Time inconsistency is particularly relevant to our implementation since it has often been tied to self-control, commitment problems and procrastination (see Hoch and Lowenstein, 1991; Frederick et al., 2002). It is unclear a priori if the effect of chess training will be stronger on patience, or 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 4).

Our Wave 2 time preference task was refined and chosen after observing the results from Wave 1: students were extremely patient, 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 was also done in Alan and Ertac (2017) 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). 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; (iii) We rewarded students with tokens which 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 eight days (later). For each decision, the most impatient alternatives results in receiving 12 tokens earlier and no tokens later, while the most patient alternatives results 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 a further three decisions which 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 (2017) 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 view choosing to receive tokens later at $r = 0$ as 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 3). Instructions for the convex time budget task are given in Appendix 4.

4.4 Creativity

Much has been written about the importance of nurturing creativity in children (e.g. Sternberg, 2003; Maker et al., 2008). Indeed, one of the essential goals of a national education system is to teach children to think critically and creatively. Sir Ken Robinson asserts in his celebrated 2006 Ted Talk ‘How Schools Kill Creativity’ that the current modern educational system fosters conformity, compliance and standardization rather than creative approaches to learning. There is an increasing focus on test scores that can be compared across students, across schools, and even across nations. 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 with children and university students (see Runco and Sakamoto, 1999, for a review).

Many different conceptions of creativity have emerged attempting to understand 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; Tardif and Sternberg, 1988). Similar to the existence of challenges in the definition of creativity there are also challenges in the methodologies used to measure it.

Unlike simpler games like 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 insolvability results in each game being different to the last thereby encouraging exploration and discovery. The 'play' state of mind is crucial for learning and creativity in children (e.g. Bateson and Martin, 2013; Gray, 2013). Although the strongest computers are now better in chess than the strongest human chess players, it is often 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 determines the particular key move and branch that is 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 of measuring creativity (Kim, 2006). The TTCT was developed by Torrance in 1966 (Torrance, 1966). Although primarily designed as an assessment for the identification of 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 of two components: the TTCT-Verbal and the TTCT-Figural. This project focuses on using TTCT-Figural tests to assess the creativity of children as the hypothesis is that if chess has any effects on creativity, it is likely to manifest itself on ideas-based creativity rather than verbal creativity. Artistic talent is not required to receive credit in the TTCT-Figural tests. The first activity is composed of two pages of lines (15 pairs) which the subject is to use in creating 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 which is 'clever' and 'unusual,' helping to 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 piloted to ensure that students understood the test instructions.

Our scoring of the creativity TTCT-Figural test was 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 to be not creative were provided as guidelines to the markers.[7] Drawings that were evaluated to be not creative were scored a '1' to reflect an attempt made at responding to the question. Other possible scores for each drawing are a '2' (somewhat creative) and a '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 make all 15 drawings in the first activity. The TTCT score we computed was an average of the two marker's scores.

A second creative test that we implemented was Guilford's (1967) alternative uses task, where examinees are asked to list as many possible uses for a common household item (e.g. brick, shoe, paper clip). The alternative uses test is a standard test of divergent thinking. In our application of the test, participants were asked to list up to 10 alternative uses for a shoelace in a fixed amount of time in order to gauge both the quantity of ideas and novelty of ideas. Our

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

scoring of the alternate uses test follows the approach we used for the TTCT-Figural test. Responses were scored a ‘1’ (not creative), ‘2’ (somewhat creative) or a ‘3’ (creative) and a ‘0’ for missing answers or non-attempts. The total score was obtained by summing up the points across all answers, with 30 as the maximum possible score. Once again, the score for the alternative uses task was computed as an average of the two marker’s scores.

4.5 Attention and Focus

The study of the development of attention occupies a central place in cognitive developmental psychology. 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).

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 considered 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 fast 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 and were then asked to cross out all the “5” digits presented on 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, where the first row (of 25 boxes) was used for a demonstration and practice trial. Participants were instructed to work as quickly as possible, using a pencil, and going from one box to the next from left to right.[8] This was a timed test and the number of correct symbols copied within 120 seconds was recorded in this test.

[8] 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.

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 grade five in schools that were randomly assigned to receive chess training – is the following OLS regression:

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

$Y_{i,s}$ denotes outcomes for individual i in school s. As randomization was done at the school level, all students in grade five in 2016 in the treatment schools were invited to participate in the chess training program. Attendance to the chess lessons was voluntary.[9] We regression adjust our results using a set of baseline covariates $X_{i,s}$ which 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.

5.1 Estimating Causal Mechanisms

We exploit the fact that our Wave 2 measures of non-cognitive outcomes and the PSC exam were collected at approximately the same time (within 3-4 weeks of each other) to examine possible pathways that chess training might affect cognitive and non-cognitive outcomes. In this section, we briefly discuss the framework we will be using to conduct our exploratory causal mediation analysis. In section 5.2, we discuss the possibility that the effects of our chess training program on academic outcomes are mediated by non-cognitive skills that themselves were impacted by the chess training program. In section 5.3, we discuss the alternative possibility of how chess training might have an effect on non-cognitive outcomes that are mediated by cognitive outcomes.

While randomized experiments can be useful in helping to provide unbiased estimates of average treatment effects when evaluating the effectiveness of a program, one criticism often

[9] Unfortunately, daily attendance to the chess training was not recorded. However, the students were very enthusiastic to learn chess. Most students attended all the chess sessions based on casual observations from the field workers and chess instructors.

made about such experiments is that an accompanying discussion on causal mechanisms is often lacking and experiments are usually silent about why such impacts occur. In this section, we perform causal mediation analysis, a statistical framework for analysing causal mechanisms that has become increasingly popular in social and medical sciences in recent years. Such an analysis defines a mechanism as a process where a causal variable of interest (i.e. a treatment) influences an outcome via an intermediate variable. The aim is to decompose the total effects of the treatment on an outcome into direct and indirect effects. The indirect effect proposes an explanation for why the treatment works, and represents the amount of the total effect that is explained by the mediator. The direct effect represents all other possible causal mechanisms and explanations for why the treatment works.

Such effect decompositions into direct and indirect effects are often represented in diagrams such as Figure 3. In the top panel (Figure 3a), we illustrate how non-cognitive skills (e.g. risk preferences) might mediate the effects of chess training on a cognitive outcome (e.g. math PSC exam marks). In the bottom panel (Figure 3b), as an alternative, we illustrate how cognitive skills might mediate the effects of chess training on a non-cognitive outcome. Although skills beget skills, it is less clear whether non-cognitive skills beget cognitive skills or vice versa. We consider both possible pathways as we are agnostic about the nature of the relationship between cognitive and non-cognitive skills in the context of our chess training program. On the one hand, we find that learning and playing chess has medium term effects on math scores in the PSC exam. This effect could operate via mediators such as risk preferences, time preferences, or attention/focus and be consistent with Cunha and Heckman (2008) who show that non-cognitive skills promote the formation of cognitive skills. On the other hand, we also find that our chess training program has an effect on risk preferences, as well as time consistency and rationality. Finding that higher cognitive skills systematically affect non-cognitive skills as captured by elicited preferences will be consistent with the findings of Burks et al. (2009), Dohmen et al. (2010) and Benjamin et al. (2013).

In a typical field experiment where no mediators are analysed, the total average treatment effect is equivalent to the direct effect and allows for many possible explanations. As this direct effect remains a black box, mediation analysis can be helpful in uncovering what lies within the black box and mechanisms that drive any observed treatment effects.

Even in the context of a randomized experiment, some additional assumptions are necessary to identify indirect and direct effects. Imai et al. (2010) formalize this using an assumption they refer to as ‘sequential ignorability’, which comprises of two components. Under the first part, sequential ignorability implies that treatment assignment must be ignorable (which will be true if the treatment if randomized). The second part of the sequential ignorability assumption requires that conditional on pre-treatment covariates, the mediator must be as if randomized. This implies that if there are any pre-treatment covariates that affect both the mediator and outcome, these covariates must be conditioned on in order to identify the direct and indirect effects. The second part of this assumption is very strong and is in principle untestable. Nevertheless, such a mediation analysis can be helpful if there are proposed mechanisms by which treatment is hypothesized to have an effect, and if the necessary covariates and measures of possible mediators have been included in the research design.

We estimate mediation effects using the following set of linear equations:

$$\begin{aligned} M_{i,s} &= \alpha_1 + \delta_1 treat_s + \beta_1 X_{i,s} + \varepsilon_{i,s} \\ Y_{i,s} &= \alpha_2 + \delta_2 treat_s + \eta M_{i,s} + \beta_2 X_{i,s} + \gamma_{i,s} \end{aligned} \tag{2}$$

In equation (2), $Y_{i,s}$ is the outcome of interest and $M_{i,s}$ is the proposed mediating variable (which we consider one at a time). We apply the approach to mediation analysis as discussed in Imai et al. (2010, 2013). The mediation package in R (Tingley et al., 2017) is used to estimate equation (2).

In order to make the sequential ignorability assumption plausible such that $Y_{i,s} \perp M_{i,s} \mid treat, X_{i,s}$, we include as additional control variables in $X_{i,s}$ the Big Five personality characteristics of the students. These personality characteristics can be viewed as the unobservables that affect both the mediator and outcomes simultaneously. In other words, we assume that preferences for risk and time are as if randomly assigned once we control for these characteristics. The students' personality variables were measured prior to the start of the chess training, using an instrument adapted from Barbaranelli et al. (2003).

5.2 Risk and Patience as Mediators for Cognitive Outcomes

Non-cognitive skills have been found to be strong predictors of educational attainment and other economic outcomes (Heckman and Kautz, 2012; Heckman and Rubinstein, 2001). Cunha and Heckman (2008) show that whilst non-cognitive skills promote the formation of cognitive skills, the reverse does not hold.

There is some evidence that risk preferences are related to education, but most research have focused on the level of education attained. Higher education can be viewed as a risky investment because it requires sacrificing present consumption and absorbing substantial psychic costs in return for future rewards, but success is not a certain outcome. Recent evidence supports the view that more risk-averse individuals have a lower tendency to pursue a university education (Outreville, 2015, provides a recent survey). Less research has focused on the relationship between risk preferences and school performance. In this line of research, attention has generally been on the format of the test. For instance, Heath (1989) finds evidence that supports a sustained male advantage when assessment is through the multiple-choice format. Halek and Eisenhauer (2001) suggest that differentials in performance in multiple-choice formats can be explained by individual risk preferences as well as demographic factors. However, this is unlikely to be very relevant in our context since in the PSC math exam, multiple choice questions represent only one-quarter of the total marks allocated.

Time preferences can also affect education outcomes but the empirical evidence on this is mixed. While Golsteyn et al. (2014) find using longitudinal data from Sweden that high discount rates are related to worse school performance, Bettinger and Slonim (2007) fail to find any correlation between patience and school performance.

Teaching children how to play chess in a prescribed systematic fashion might help benefit their development of specific non-cognitive skills, and these non-cognitive skills might in turn be related to academic outcomes. We focus on risk preferences, the various dimensions of time preferences, and attention/focus as possible mechanisms.

Chess can be useful to illustrate the concept and benefits of taking calculated risk. For example, sacrifices in material (i.e. giving up a pawn, bishop, knight, rook or queen) are often made by chess players if it helps checkmate the opponent’s king and win the game. Such sacrifices are inherently risky because if one’s calculations are inaccurate, then the loss of material could prove to be fatal and eventually lead to a quick loss of the game. Gambits and sacrifices can be made in any of the three different phases of a chess game, be it in the opening, middlegame or endgame. Gerdes and Gränsmark (2010) and Dreber et al. (2013) measure risk in chess by exploiting the fact that chess players start each game by choosing a strategic development scheme for their pieces (called a chess opening). By exploiting a standardized classification of these openings and expert assessments, they label the chosen strategies in each game as being either risky (aggressive) or safe (solid). All chess players have a regular set of openings they use to start the game as white or black and this is referred to as their opening repertoire. Considerable effort is dedicated to creating an opening repertoire that matches one's personality. Players who are more temperamental will typically choose more aggressive openings, while a calm, agreeable person will more often choose a solid opening.

As chess players who are beginners are unlikely to have a well thought out opening repertoire, it is not possible to use the above mentioned approach previously seen in the literature to measure risk preferences. However, the point is that chess is a game that lends itself easily to allowing risk to be assessed and expressed. By comparing children who have been subject to three weeks of intensive chess lessons with children who have essentially no knowledge of chess, we hypothesize that learning how to play chess and an appreciation of basic chess strategy can help in the development and articulation of risk preferences in children.

In addition, chess might help teach children to be more patient, more focused, and have more self-control. It can potentially motivate children to become willing problem solvers and 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; one needs to look a few steps ahead in a chess game and consider and evaluate alternative scenarios. Chess can teach children how to focus and how to visualize by imagining a sequence of events before it happens. The schematic thinking approach in chess is similar to trees and branches in 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).

5.3 Cognitive Skills as Mediators for Non-Cognitive Outcomes

Some research has focused on the alternative pathway that cognition has an effect on non-cognitive outcomes. Burks et al. (2009) examine how an individual's cognitive skills are related to the individual's economic preferences. They find that individuals with better cognitive skills are more patient in both the short- and long-run. Better cognitive skills are also associated with a greater willingness to take calculated risks. Frederick (2005) finds that performance on a range of cognitive tests correlates negatively with impatience and risk aversion. Dohmen et al. (2010) find these relationships in a large, representative sample from Germany. Benjamin et al. (2013) uses three lab experiments with high school students in Chile as subjects and find that short-term discounting and small-stakes risk aversion are more prevalent among students with lower measured cognitive ability.

Based on theories of choice bracketing (Tversky and Kahneman, 1981), the tendency for lower cognitive ability to be less able to recognize how individual risky decisions integrate with other assets like lifetime wealth, or to conceptualize and integrate future considerations with current goals could be one mechanism explaining the link between cognition and economic preferences. This theory is consistent with the above-mentioned empirical results that have been found in the literature.

‘Two-system’ theories, which postulate a causal relationship between cognitive skills and the expression of behavioral biases also support the above-mentioned empirical results. According to this theory, decision-making results from the interaction of a deliberative system, which is patient and risk neutral, and an emotional system, which is impulsive and risk averse. Results from lab experiments suggest that higher cognition tends to mitigate risk aversion and impatience, especially if we consider cognitive ability to be an empirical proxy for the player’s long-run cognitive resources. For example, Benjamin et al. (2013) find that inducing a cognitive load (e.g. requiring participants to remember a string of seven numbers while they are engaged in the task of interest) leads to more impatient and more risk-averse decisions. Thus, a two-system decision process is another potential mechanism explaining the negative relationship of risk aversion and impatience with cognitive ability.

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 7. In theory, the results do not need to be regression adjusted due to randomization. We do so as a robustness check and to increase the precision of the experimental estimates. The sample sizes for unadjusted and regression adjusted results vary and are dependent 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 partly reflects the fact that on any given day, student absenteeism is high in primary schools in rural Bangladesh.[10] We also estimated unadjusted impacts for the regression adjusted samples presented; the results are in general very similar to the unadjusted impacts seen in Tables 2 to 7 (results available upon request).

Three alternative sets of *p*-values are presented. First, using asterisks in the columns for unadjusted and regression adjusted impacts, we present standard *p*-values from a regression model based on clustered standard errors. Second, *p*-values using the wild bootstrap (1000 replications) proposed by Cameron et al. (2008) are presented in the following column. Third, as there are many outcomes examined, it might be important to address the issue of multiple hypothesis testing. To deal with this, we compute false discovery rate sharpened q-values (Benjamini et al., 2006) using the procedure in Anderson (2008). These *q*-values are provided in the last column of Tables 3 to 7.

[10] See, for example, Chaudhury et al. (2006) for the high absence of children in schools in Bangladesh and other developing countries.

6.1 Academic Results

We consider two types of test scores to measure cognitive ability. The first involves the use of a project-administered mathematics test. The treatment group scored slightly better in the pre-program mathematics 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 group 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. The results of the PSC exam for the six mandatory subjects – Bengali, English, science, social science, mathematics and religion – are provided in Table 3.

We find a significant positive effect of our intensive chess instruction program on grades for mathematics in the PSC exam (p-value = 0.03 using the wild cluster bootstrap). The treatment-control difference of 0.71 points is approximately equivalent to between half and a full letter grade in mathematics. There is a smaller impact on Religious Studies (0.41) which is significant at the 10% level using the wild cluster bootstrap, but impacts on the remaining subjects in the PSC exam are not statistically significant. This suggests that the main driver of the differences in GPA in the PSC exam come from differences in math scores. The overall GPA was 3.86 for the treatment group and 3.45 for the control group. This difference of 0.41 is statistically significant using conventional clustered standard errors and the wild bootstrap (p-value = 0.086).

6.2 Risk Preferences Results

The average value of the alternative chosen in the risk-elicitation task was used for assessing a treatment effect on risk preferences, where a higher value indicates a riskier choice. The values range from 1-5 in Wave 1, and 1-6 in Wave 2.[11] Results are depicted in Table 4. In Wave 1, treated students invested, on average, 0.3 more items into the risky 'asset' (regression adjusted impact, 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. Figure 2 breaks down the treatment effects according to each available alternative and highlights the changes between Waves 1 and 2. For

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

both waves we can see that the largest difference emerges for alternative 1 – the safest alternative. There is, in addition, a strong effect 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.

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 for assessing 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 choice set containing five alternatives.[12] For each student, we again summed a count of impatient alternatives chosen. The results (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 do two additional tests. The first involves a test for time inconsistency. In both waves, we have students make decisions over an original and one-week-delayed set that differ only in having payoffs in the latter realised seven-days later than in 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. In Wave 1, both the binary and continuous measures of time inconsistency (third and fourth row of Table 5) are found to be significantly different from zero (p-values = 0.008 and 0.006 respectively). In Wave 2, however, only the continuous measure of time inconsistency (fourth row of Table 5) is found to be significantly different (p-value = 0.090). Nonetheless, in all cases, the sign of

[12] 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.

the effect remains the same. Students in the treatment group are less likely to make time inconsistent decisions.

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 Table 5. In Wave 1, students in the treatment group are 12.1% less likely to violate the monotonicity requirement (p-value = 0.002). Similarly, the corresponding coefficient is 12.6% (p-value =0.062) in Wave 2.

Overall, while we find no statistically significant effects on impatience, we find that students in the treatment group are both less likely to be time inconsistent. They are also less likely to exhibit non-monotonic time preferences.

6.4 Creativity Results

It is reasonable to hypothesize that the experience of learning and playing chess intensively over a three-week period can lead to improving one's creativity. Chess offers a way of thinking generally unavailable in school subjects because of its focus on imagination, patterns and structures. Results from the TTCT-Figural test are presented in Table 6. The average score for the control group was 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 obtains 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.

6.5 Results for Attention and Focus

Anecdotal evidence from watching children playing chess in tournaments suggests that playing chess as a regular activity might help in making young children less fidgety and able to develop a better ability to concentrate on a task at hand. Chess players are known, for example, to be able to tune out background noise when focusing on a chess game that they are playing. 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 (see Table 7). Our results therefore do not suggest that there are any medium term effects of chess instruction on focus and attention.

6.6 Results of Mediation Analysis

Table 8 presents the results of our exploratory causal mediation analysis. Columns (1)-(3) focus on non-cognitive mediators and cognitive outcomes. Columns (4)-(7) focus on cognitive mediators and non-cognitive outcomes. The 'total effect' presented is the sum of the average causal mediation effect (ACME) and the direct effect. Note that the analysis conducted in this section is based on a slightly smaller sample size due to missing data for our measures of the Big Five personality characteristics for some students. As a result, the 'total effect' estimates presented in this table will not necessarily coincide with the average treatment effects presented in earlier tables.

Focusing first on cognitive outcomes with non-cognitive mediators (columns 1-3), we can see that the most statistically significant total effects are for PSC math grades. The novel evidence provided in Table 8 is that about 27% of the effect operates through risk preferences (when all the other potential mediating variables are excluded). This average causal mediation effect (ACME) is significant at the 10% level (p-value = 0.082). This results provides support for the findings in Cunha and Heckman (2008) who show that non-cognitive skills promote the formation of cognitive skills.

When we examine the alternative possibility that cognitive skills help mediate our non-cognitive outcomes (columns 4-7), we find less evidence for such a mechanism. In column (4) of Table 8, although the proportion mediated via match is also statistically significant, only about 3% of the effect operates through math.

Nevertheless, we stress that these results are exploratory in nature. As our measures of cognitive outcomes were collected slightly after (3-4 weeks) our Wave 2 measures of non-cognitive skills and economic preferences, it is possible that such a time ordering of the data might have an effect on the results. In addition, the sequential ignorability assumption is untestable and there likely remain unobserved variables that affect our causal mediation analysis.

6.7 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:

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

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 9 reports the subgroup impact results. Overall, we fail to detect any subgroup differences by gender, baseline math ability or baseline household income.

6.8 Multiple Hypotheses Testing

As there are many outcomes examined, this raises the issue of multiple hypothesis testing. To control for the false discovery rate, we provide sharpened q-values (Benjamini et al., 2006) using the procedure implemented in Stata by Anderson (2008). The last column of Tables 3 to 7 show the FDR-adjusted sharpened q-values for the outcomes of interest. Their interpretation 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 group the tests into families based on the domain tested: academic, risk, time preferences, creativity, and attention/focus, whose results are presented in Tables 3 to 7.

For PSC math scores, when the three main subjects in primary school in Bangladesh – Bangla, math and science – are considered (as examined in the evaluation by Jerrim et al. (2016, 2018) in the context of the UK), the FDR q-value for math scores increases from 0.030 (wild-bootstrap) to 0.099 (FDR) and remains marginally statistically significant. For risk preferences, the highly significant *p*-value for Wave 2 remains after making the FDR adjustment. For time preferences, in Wave 1, the results remained unchanged following the FDR adjustment. Both the binary and continuous measures of time inconsistency, as well as non-monotonicity, remain highly significant. However, the impacts in Wave 2 for time inconsistency and non-monotonicity are no longer statistically significant following the FDR adjustment.

7. Summary and Conclusions

This paper evaluates the potential benefits of learning chess as a child. We conducted a randomized experiment on grade five students in rural Bangladesh whereby the intervention comprised of a 30 hour training program based on a curriculum approved by the World Chess Federation. The benefits of playing chess regularly has 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 curriculum.

Chess is generally viewed by parents and teachers as a highly regarded extra curricula activity in primary school. In recent years, chess coaching for children has become increasingly popular in the US and other developed countries.[13] By employing a field experiment to examine the potential benefits of having chess lessons within the educational framework, we have provided credible estimates of the benefits chess instruction for children can have for cognitive and non-cognitive outcomes.

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

A second contribution is with regards to attempting to better understand the evolution of cognitive and non-cognitive skills in children in response to an intervention that has the potential to help develop both sets of skills. We are not aware of any other work that attempts to empirically examine the underlying mechanisms that helps explain why chess might be of benefit to children. Benjamin Franklin (1786) opines that by learning and playing chess, "several very valuable qualities of the mind, useful in the course of human life, are to be acquired or strengthened by it." The outcomes we measure in our chess intervention aim to quantify some of these "valuable qualities." We choose to examine how chess training might

[13] For example, in the US, the Chess Club and Scholastic Center of Saint 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).

have an effect on academic outcomes that are mediated by non-cognitive outcomes, or alternatively how chess training might have an effect on non-cognitive outcomes that are mediated by academic outcomes

In terms of outcomes, we examine both short-run effects, based on assessments that were made shortly after the conclusion of the program, as well as medium-term effects based on assessments that were conducted 9-10 months after the program ended. We also measure impacts on academic outcomes based on results from a compulsory public exam all grade five students in Bangladesh must sit for.

We find that chess training has a significant positive impact on math scores in the national exam and reduces the treatment group's level of risk-aversion. While the program has insignificant effects on students' level of impatience, it decreases the incidence of time inconsistent decisions as well as decreases the incidence of non-monotonic time-preferences. Our results for risk preferences, time inconsistency and non-monotonic time-preferences remain highly significant even after accounting for multiple hypothesis testing. The significant (at the 5% level) longer term impacts on PSC math scores weaken somewhat after the FDR adjustment. In exploring causal mechanisms, we find some evidence that decreased risk aversion is a better mediator for the effect on math scores than vice versa.

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 a rural developing country and results obtained do not necessarily have external validity. Nonetheless, by focusing the intervention on a group of children who essentially had no prior experience with playing chess and who do not have access to many contemporary toys and games that children in developed countries do (e.g. standard play materials, mobile devices, computer games) that provide mental stimulation, we potentially allow for a fuller impact of chess lessons (if any) to emerge and be realized.

References

Alan, S. and S. Ertac. (2017). Fostering patience in the classroom: results from randomized educational intervention. *Journal of Political Economy*, forthcoming.

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

Barbaranelli, C., G. Caprara, A. Rabasca and C. Pastorelli. (2003). A questionnaire for measuring the Big Five in late childhood. *Personality and Individual Differences*, 34: 645–664.

Bateson, P. and P. Martin. (2013). *Play, playfulness, creativity and innovation*. Cambridge University Press: Cambridge.

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 Slonim, R., 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.

Bühren, C. and B. Frank. (2010). Chess players' performance beyond 64 squares: A case study on the limitations of cognitive abilities transfer. MAGKS Joint Discussion Paper Series in Economics, 19-2010.

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.

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

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

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.

Cawley, J. and C. Ruhm (2012). The economics of risky health behaviors, in *Handbook of Health Economics*, Volume 2. T. McGuire, M. Pauly, and P. Barros (Eds). Elsevier: New York.

Chabris C., D. Laibson, C. Morris, J. Schuldt and D. Taubinsky. (2008). Individual laboratory-measured discount rates predict field behaviour. *Journal of Risk and Uncertainty*, 37: 237–69.

Charness, G., Gneezy, U. and Imas, A., (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.

Cohn, R., W. Lewellen, R. Lease and G. Schlarbaum. (1975). Individual investor risk aversion and investment portfolio composition. *Journal of Finance:* 30: 605–620.

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

Cramer, J., J. Hartog, N. Jonker and C. Van Praag. (2002). Low risk aversion encourages the choice for entrepreneurship: an empirical test of a truism. *Journal of Economic Behavior and Organization*, 48: 29–36.

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 behavioral 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.

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

Frederick, Shane (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.

Gerdes, C. and P. Gränsmark. (2010). Strategic behavior across gender: a comparison of female and male expert chess players. *Labour Economics*, 17: 766–775.

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

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

Gray P. (2013). *Free to learn*. New York: Basic Books.

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.

Gutman, L. and I. Schoon. (2013). The impact of non-cognitive skills on outcomes for young people, Institute of Education, University of London.

Halek, M and J. Eisenhauer. (2001). Demography and risk aversion. *The Journal of Risk and Insurance*, 68: 1–24.

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

Heath, J. (1989). An Econometric Model of the Role of Gender in Economic Education. *The American Economic Review*, 79: 226–230.

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.

Hoch, S.J. and Loewenstein, G.F., (1991). Time-inconsistent preferences and consumer self-control. *Journal of consumer research*, 17: 492–507.

Imai, K., L. Keele, and T. Yamamoto. (2010). Identification, inference, and sensitivity analysis for causal mediation effects. *Statistical Science*, 25: 51–71.

Imai, K., D. Tingley, and T. Yamamoto. (2013). Experimental designs for identifying causal mechanisms. *Journal of the Royal Statistical Society, Series A*, 176: 5–51.

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*, forthcoming.

Kazemi, F., M. Yektayar and A. Abad. (2012). Investigation the impact of chess play on developing meta-cognitive ability and math problem-solving power of students at different levels of education. *Procedia-Social and Behavioral Sciences*, 32: 372–379.

Kim, K. H. (2006). Can we trust creativity tests? A review of the Torrance Tests of Creative Thinking (TTCT). *Creativity Research Journal*, 18: 3–14.

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

Lührmann, M., M. Serra-Garcia and J. Winter. (2015). The impact of financial education on adolescents' intertemporal choices. *Institute for Fiscal Studies Working Paper 14/18.*

Maker, C., J. Sonmi and O. Muammar. (2008). Development of creativity: The influence of varying levels of implementation of the DISCOVER curriculum model, a non-traditional pedagogical approach. *Learning and Individual Differences*, 18: 402–417.

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.

Scholz, M., H. Niesch, O. Steffen, B. Ernst, M. Loeffler and E. Witruk, E., 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 Experimenal Child Psychology*, 56: 328–349.

Sutter, M., Kocher, M.G., Glätzle-Rützler, D. and Trautmann, S.T., (2013). Impatience and uncertainty: Experimental decisions predict adolescents' field behavior. *The American Economic Review*, 103: 510–531.

Sternberg, R. (2003). Creative thinking in the classroom. *Scandinavian Journal of Educational Research*, 47: 325–338.

Tardif, T. and R. Sternberg. (1988). What do we know about creativity?, in R. Sternberg, ed., *The Nature of Creativity*, Cambridge University Press, New York, pp. 609–628.

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.

Tingley, D., T. Yamamoto, K. Hirose, L. Keele, and K. Imai. (2017). Mediation: R Package for Causal Mediation Analysis. Version 4.4.6.

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.

Vuchinich, R. and C. Simpson. (1998). Hyperbolic temporal discounting in social drinkers and problem drinkers. *Experimental and clinical psychopharmacology,* 6: 292.

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: Intervention Timeline

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

Note: The chess program was conducted from Saturday to Tuesday over a period of three weeks. Day 1 is a Saturday. There were a total of 12 program days where chess lessons were provided.

Figure 2: Distribution of Choices across Groups, and Waves in the Risk-Elicitation Task

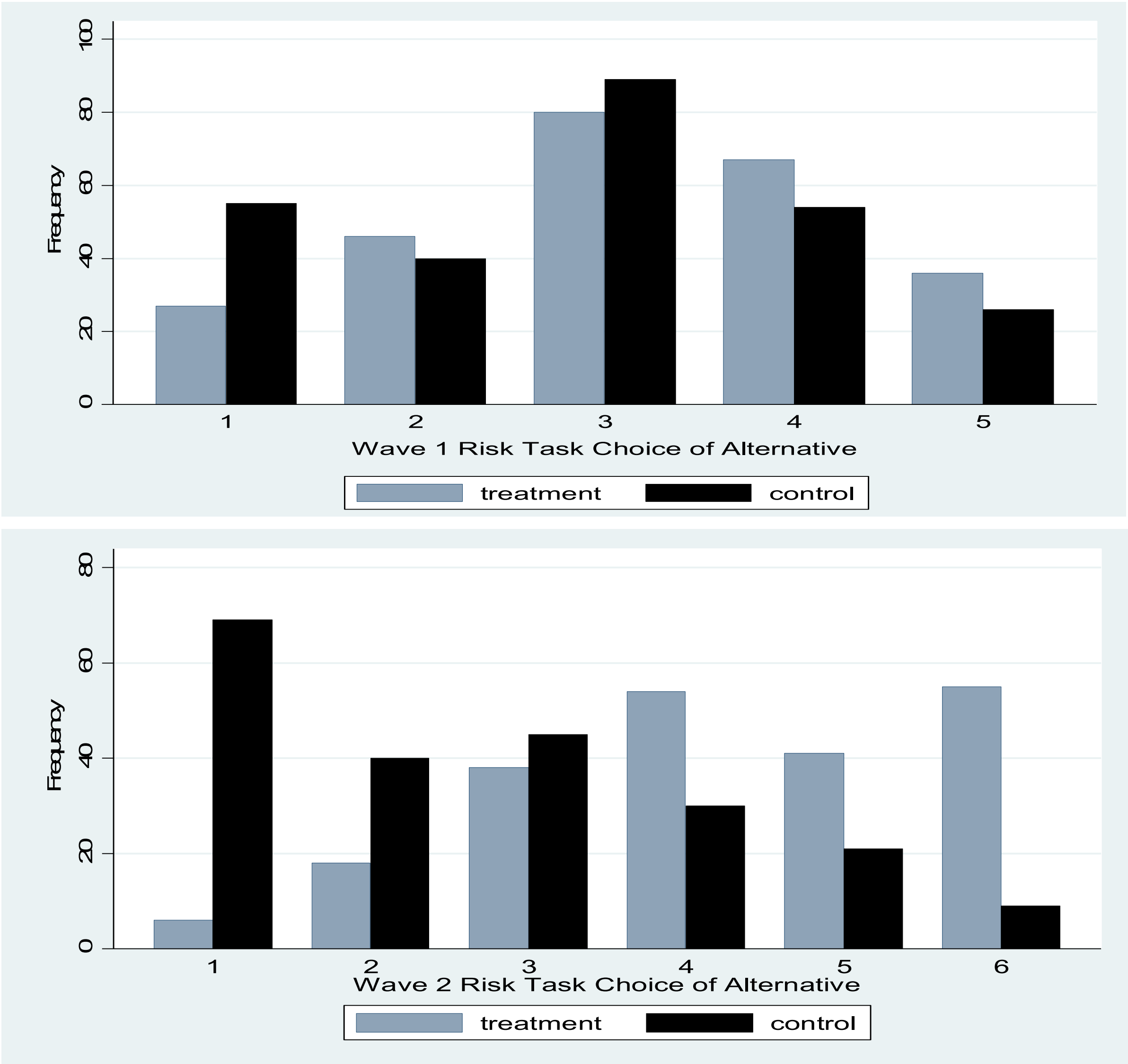
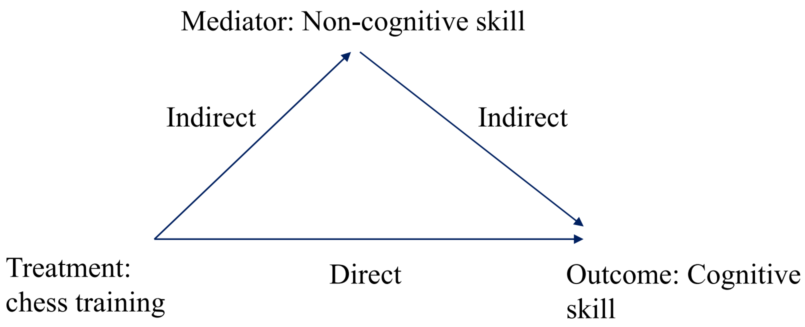


Figure 3: Mediation Analysis Illustration

(a) How a Non-Cognitive Skill Mediates the Treatment Effect on Cognitive Skills



(b) How Cognitive Skills Mediates the Treatment Effect on a Non-Cognitive Skill

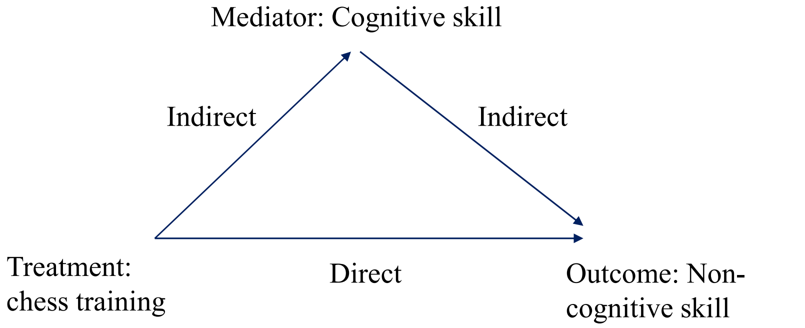


Table 1: Treatment/Control Raw Mean Differences in Household Characteristics

Variable	Treatment Mean	Control Mean	Difference
Household income (in takas)	8316.15	8657.38	-341.22
			(559.25)
Number of household members	4.357	4.293	0.064
			(0.122)
Sanitary ring latrine in the house	0.629	0.643	-0.014
			(0.063)
Drinking water in the house from tube well	0.636	0.816	-0.181
			(0.162)
Existence of electricity supply in the house	0.340	0.493	-0.153
			(0.173)
Distance of the school from the home (km)	1.120	0.672	0.448
			(0.362)
Value of total assets except land (in takas)	70656.46	61675.59	8980.87
			(12714.08)
Household religion (Muslim = 1)	0.932	0.940	-0.008
			(0.033)
Do any of the parents know how to play chess	0.102	0.063	0.039
			(0.029)
Someone with more than grade 10 education in household	0.136	0.133	0.003
			(0.028)
Father's years of schooling	4.236	4.345	-0.109
			(0.650)
Mother's years of schooling	4.12	4.07	0.05
			(0.739)
Father's age	39.94	40.00	-0.06
			(0.624)
Mother's age	33.64	33.59	0.054
			(0.528)
Father works as labourer/in agriculture	0.653	0.593	0.059
			(0.073)
Mother is a housewife	0.948	0.963	-0.014
			(0.013)
Two-parent household	0.972	0.973	-0.001
			(0.011)
Gender of student (male =1)	0.489	0.437	0.053
			(0.048)
N	294	300	

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	Wild bootstrap p-value
Math pre-marks	18.71	0.506 (3.168)	1.362 (2.719)	0.608
N	215	494	445	
Math post-marks	14.38	1.304 (3.019)	2.072 (2.414)	0.442
N	209	478	428	

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 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.

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

Variable	Control Mean	Unadjusted Impact	Regression Adjusted Impact	Wild bootstrap p-value	FDR adjusted q-values
Grade for Bangla	3.76	0.282 (0.224)	0.347* (0.197)	0.180	0.220
Grade for English	2.90	0.457 (0.334)	0.398 (0.330)	0.338	
Grade for Math	2.93	0.718* (0.357)	0.705** (0.283)	0.030	0.099
Grade for Social Science	3.63	0.240 (0.371)	0.306 (0.319)	0.434	
Grade for Science	3.60	0.341 (0.287)	0.292 (0.294)	0.426	0.370
Grade for Religious Studies	3.95	0.387 (0.209)	0.405* (0.209)	0.084	
Overall GPA	3.45	0.413 (0.242)	0.414* (0.214)	0.086	
N	190	434	395		

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 from Table 1. The wild bootstrap p-values are for the regression adjusted impacts and 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. False discovery rate (FDR) sharpened q-values (Benjamini et al., 2006) are computed using the procedure in Anderson (2008).

Table 4: Risk preferences (Waves 1 and 2)

Variable	Control Mean	Unadjusted Impact	Regression Adjusted Impact	Wild bootstrap p-value	FDR adjusted q-values
Wave 1 (Min 1, Max 5), higher value = less risk averse	2.84	0.319 (0.166)	0.301 (0.175)	0.144	0.078
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)	0.002	0.005
N	191	426	381		

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 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).

Table 5: Time preferences (Waves 1 and 2)

Variable	Control Mean	Unadjusted Impact	Regression Adjusted Impact	Wild bootstrap p-value	FDR adjusted q-values
Wave 1					
Impatience (0-5)	1.26	0.038 (0.062)	-0.026 (0.062)	0.716	0.508
Delayed impatience (1-5)	1.32	0.016 (0.065)	-0.040 (0.061)	0.540	0.508
Time inconsistency (binary)	0.28	-0.086** (0.037)	-0.162*** (0.040)	0.008	0.025
Time inconsistency (0-5)	0.38	-0.091 (0.064)	-0.234*** (0.060)	0.006	0.025
Non-monotonicity (binary)	0.14	-0.089*** (0.028)	-0.121*** (0.018)	0.002	0.021
N	224	521	450		
Wave 2					
Impatience (2-10)	5.19	-0.338 (0.331)	-0.087 (0.365)	0.898	0.561
Delayed impatience (2-10)	5.27	-0.120 (0.270)	-0.151 (0.257)	0.634	0.508
Time inconsistency (binary)	0.74	-0.073 (0.046)	-0.060 (0.042)	0.202	0.220
Time inconsistency (0-2)	1.13	-0.145 (0.084)	-0.129* (0.065)	0.090	0.145
Non-monotonicity (binary)	0.67	-0.107* (0.052)	-0.126** (0.055)	0.062	0.122
N	191	426	381		

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 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).

Table 6: Creativity (Wave 1)

Variable	Control Mean	Unadjusted Impact	Regression Adjusted Impact	Wild bootstrap p-value	FDR adjusted q-values
TTCT pairs of lines	16.57	-0.285 (0.656)	-0.087 (0.572)	0.934	1.000
TTCT picture completion	14.47	-0.554 (1.462)	0.119 (1.311)	0.940	1.000
Guilford's alternate uses test	14.38	0.889 (1.429)	1.727 (1.069)	0.150	0.819
N	223	483	432		

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 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).

Table 7: Digit Cancellation and Digit Symbol Test (Wave 2)

Variable	Control Mean	Unadjusted Impact	Regression Adjusted Impact	Wild bootstrap p-value	FDR adjusted q-values
Digit cancellation test: Total score	205.89	-0.751 (0.691)	-0.601 (0.566)	0.334	0.969
Digit cancellation test: Time to completion	225.56	9.589 (7.856)	4.892 (6.438)	0.492	0.969
Digit symbol test	32.72	1.579 (1.331)	1.075 (1.326)	0.482	0.969
N	190	425	380		

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 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).

Table 8: Mediation Analysis

Outcome Variable	PSC math grade (1)	PSC science grade (2)	PSC overall GPA (3)	Risk (Wave 2) (4)	Patience (Wave 2) (5)	Monotonicity (Wave 2) (6)	Time consistency (Wave 2) (7)
ACME of risk (Wave 2)	0.152 (0.082)	0.119 (0.110)	0.082 (0.178)				
Direct effect of chess training	0.397 (0.038)	0.016 (0.920)	0.188 (0.146)	–	–	–	–
Total effect	0.549 (0.000)	0.135 (0.320)	0.270 (0.008)				
Proportion mediated	**0.272 (0.082)**	0.578 (0.400)	0.304 (0.178)				
ACME of patience (Wave 2)	-0.005 (0.74)	-0.004 (0.750)	-0.01 (0.762)				
Direct effect of chess training	0.545 (0.000)	0.141 (0.280)	0.272 (0.018)	–	–	–	–
Total effect	0.540 (0.000)	0.136 (0.290)	0.267 (0.018)				
Proportion mediated	-0.001 (0.74)	-0.008 (0.880)	-0.01 (0.776)				
ACME of monotonicity (Wave 2)	0.027 (0.19)	0.028 (0.140)	0.026 (0.150)				
Direct effect of chess training	0.516 (0.000)	0.112 (0.370)	0.248 (0.020)	–	–	–	–
Total effect	0.544 (0.000)	0.139 (0.290)	0.275 (0.010)				
Proportion mediated	0.042 (0.190)	0.132 (0.390)	0.09 (0.150)				
ACME of time consistency (Wave 2)	0.029 (0.220)	0.020 (0.240)	0.022 (0.216)				
Direct effect of chess training	0.511 (0.000)	0.121 (0.350)	0.256 (0.024)	–	–	–	–
Total effect	0.539 (0.000)	0.141 (0.290)	0.278 (0.012)				
Proportion mediated	0.048 (0.220)	0.083 (0.430)	0.071 (0.220)				
ACME of PSC math				0.063 (0.092)	0.082 (0.150)	0.022 (0.108)	0.062 (0.034)
Direct effect of chess training	–	–	–	1.721 (0.000)	-0.191 (0.490)	0.079 (0.230)	0.127 (0.388)
Total effect				1.784 (0.000)	-0.109 (0.680)	0.101 (0.094)	0.189 (0.192)
Proportion mediated				**0.033 (0.092)**	-0.112 (0.740)	0.194 (0.190)	0.261 (0.222)
ACME of PSC science				0.020 (0.370)	0.028 (0.360)	0.009 (0.320)	0.017 (0.360)
Direct effect of chess training	–	–	–	1.768 (0.000)	-0.012 (0.660)	0.090 (0.170)	0.176 (0.180)
Total effect				1.789 (0.000)	-0.094 (0.730)	0.099 (0.150)	0.194 (0.160)
Proportion mediated				0.008 (0.370)	-0.025 (0.860)	0.067 (0.430)	0.064 (0.450)
ACME of PSC GPA				0.037 (0.190)	0.089 (0.042)	0.025 (0.020)	0.050 (0.028)
Direct effect of chess training	–	–	–	1.747 (0.000)	-0.185 (0.494)	0.072 (0.260)	0.133 (0.320)
Total effect				1.784 (0.000)	-0.095 (0.722)	0.097 (0.130)	0.184 (0.190)
Proportion mediated				0.019 (0.190)	-0.158 (0.736)	0.217 (0.140)	0.219 (0.214)
N	306	306	306	306	306	306	306

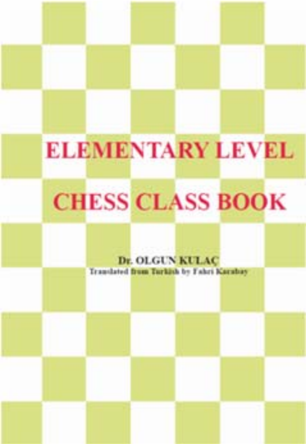
Notes: ACME = average causal mediation effect. *p*-values in parentheses. The calculation of confidence intervals in parentheses are based on quasi-Bayesian confidence intervals using 1000 simulations.

Table 9: 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 (N=182)	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 (N=213)	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 (N=184)	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 (N=177)	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 (N = 249)	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 (N=146)	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 1: Syllabus of the Chess Training Program



CONTENTS

UNIT-1
GETTING TO KNOW CHESS
- Getting to Know Chess 1
- What is Chess? 2
- Benefits of Chess 4
- Chess Board 5
- Chessmen 8
- Placing the Chessmen 9
- How they Move 10
- Movement of Rook 11
- Movement of Bishop 12
- Movement of Queen 15
- Movement of Knight 16
- Movement of Pawn 19
- Movement of King 20
- Capture by Rook 23
- Capture by Bishop 24
- Capture by Queen 27
- Capture by Knight 28
- Capture by Pawn 31
- Capture by King 32
- Attacking a chessman 35
- Piece Values 38

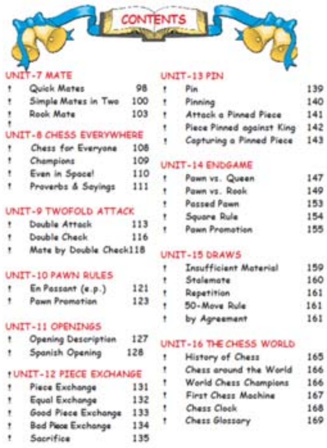
UNIT-2 CHECKMATE
- Check 40
- King under threat 43
- Checkmate 47
- Mate Positions 48
- Mate In One Move 49

UNIT-3
GETTING TO KNOW THE CHESSBOARD
- Rank 53
- File 53
- Diagonal 54
- Names of squares 55
- Central Squares 56

UNIT-4
SIMPLE MATES
- Queen Mate 59
- Cylinder Mate 60
- Fool`s Mate 61
- Scholar`s Mate 62

UNIT-5 ATTACK and DEFENCE
- Discovered Attack 66
- Discovered Check 69
- Protecting 72
- Moving Away 73
- Fork 76
- Skewer 82

UNIT-6 RULES
- Castling 86
- Notation 91
- Stalemate 95
- Scoring 95
- Sportsmanship 96



CONTENTS

UNIT-7 MATE
- Quick Mates 98
- Simple Mates in Two 100
- Rook Mate 103

UNIT-8 CHESS EVERYWHERE
- Chess for Everyone 108
- Champions 109
- Even in Space! 110
- Proverbs & Sayings 111

UNIT-9 TWOFOLD ATTACK
- Double Attack 113
- Double Check 116
- Mate by Double Check 118

UNIT-10 PAWN RULES
- En Passant (e.p.) 121
- Pawn Promotion 123

UNIT-11 OPENINGS
- Opening Description 127
- Spanish Opening 128

UNIT-12 PIECE EXCHANGE
- Piece Exchange 131
- Equal Exchange 132
- Good Piece Exchange 133
- Bad Piece Exchange 134
- Sacrifice 135

UNIT-13 PIN
- Pin 139
- Pinning 140
- Attack a Pinned Piece 141
- Piece Pinned against King 142
- Capturing a Pinned Piece 143

UNIT-14 ENDGAME
- Pawn vs. Queen 147
- Pawn vs. Rook 149
- Passed Pawn 153
- Square Rule 154
- Pawn Promotion 155

UNIT-15 DRAWS
- Insufficient Material 159
- Stalemate 160
- Repetition 161
- 50-Move Rule 161
- by Agreement 161

UNIT-16 THE CHESS WORLD
- History of Chess 165
- Chess around the World 166
- World Chess Champions 166
- First Chess Machine 167
- Chess Clock 168
- Chess Glossary 169

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

Day no	Unit number	Particulars name	Page number
Day-1	**Unit-1**	**Getting to know chess**	
	Unit-1	Getting to know chess	1
	Unit-1	What is chess	2
	Unit-1	Benefits of chess	4
	Unit-1	Chess board	5
	Unit-1	Chess men	8
	Unit-1	Placing the chessmen	9
	Unit-1	How they move	10
	Unit-1	Movement of rook	11
	Unit-1	Movement of bishop	12
	Unit-1	Movement of queen	15
	Unit-1	Movement of knight	16
	Unit-1	Movement of pawn	19
	Unit-1	Movement of king	20

Day no	Unit number	Particulars name	Page number
Day-2	Unit-1	**Getting to know chess**	
	Unit-1	Capture by rook	23
	Unit-1	Capture by bishop	24
	Unit-1	Capture by queen	27
	Unit-1	Capture by knight	28
	Unit-1	Capture by pawn	31
	Unit-1	Capture by king	32
	Unit-1	Attacking a chessman	35
	Unit-1	piece values	38

Day no	Unit number	Particulars name	Page number
Day-3	**Unit-2**	**Checkmate**	
	Unit-2	Check	40
	Unit-2	King under threat	43
	Unit-2	Checkmate	47

Day no	Unit number	Particulars name	Page number
Day-4	Unit-2	**Checkmate**	
	Unit-2	Mate position	48
	Unit-2	Mate in one move	49
	Unit-3	**Getting to know the chess board**	
	Unit-3	Rank	53
	Unit-3	File	53
	Unit-3	Diagonal	54
	Unit-3	Names of squares	55
	Unit-3	Central squares	56

Day no	Unit number	Particulars name	Page number
	Unit-4	**Simple mate**	

Day-5	Unit-4	Queen mate	59
	Unit-4	Cylinder Mate	60
	Unit-4	Fools mate	61
	Unit-4	Scholars mate	62
	Unit-6	**Rules**	
	Unit-6	Notation	91

Day no	Unit number	Particulars name	Page number
Day-6	**Unit-5**	**Attack and defense**	
	Unit-5	Discovered attack	66
	Unit-5	Discovered check	69
	Unit-5	Protecting	72
	Unit-5	Moving away	73

Day no	Unit number	Particulars name	Page number
Day-7	Unit-5	**Attack and defense**	
	Unit-5	Fork	76
	Unit-5	Skewer	82

Day no	Unit number	Particulars name	Page number
Day-8	**Unit-6**	**Rules**	
	Unit-6	Castling	86
	Unit-6	Stalemate	95
	Unit-6	Scoring	95
	Unit-6	Sportsmanship	96
	Unit-10	**Pawn rules**	
	Unit-10	En Passant (e.p.)	121
	Unit-10	Pawn promotion	123

Day no	Unit number	Particulars name	Page number
Day-9	**Unit-7**	**Mate**	
	Unit-7	Quick mates	98
	Unit-7	Simple mate in two	100
	Unit-7	Rook mate	103
	Unit-9	**Two fold attack**	
	Unit-9	Double attack	113
	Unit-9	Double check	116
	Unit-9	Mate by double check	118

Day no	Unit number	Particulars name	Page number
Day-10	Unit 10	Recap pawn rules	
	Unit-11	**Opening**	
	Unit-11	Broader survey on chess openings	
	Unit-11	Opening description	127
	Unit-11	Spanish opening	128

Day no	Unit number	Particulars name	Page number
Day-11	**Unit-12**	**Piece exchange**	
	Unit-12	Piece exchange	131
	Unit-12	Equal exchange	132
	Unit-12	Good piece exchange	133
	Unit-12	Bad piece exchange	134
	Unit-12	Sacrifice	135
	Unit-13	**Pin**	
	Unit-13	Pin	139
	Unit-13	Pinning	140
	Unit-13	Attack a pinned piece	141
	Unit-13	Piece pinned against king	142
	Unit-13	Capturing a pinned piece	143

Day no	Unit number	Particulars name	Page number
Day-12	**Unit-14**	**End game**	
	Unit-14	Pawn vs. Queen	147
	Unit-14	Pawn vs. Rook	149
	Unit-14	Passed pawn	153
	Unit-14	Square rule	154
	Unit-14	Pawn promotion	155
	Unit-15	**Draws**	
	Unit-15	Insufficient materials	159
	Unit-15	Stalemate	160
	Unit-15	Repetition	161
	Unit-15	50-Move rule	161
	Unit-15	By agreement	161
	Unit-16	**The chess world**	
	Unit-16	History of chess	165
	Unit-16	Chess around the world	166
	Unit-16	World chess champion	166
	Unit-16	First chess machine	167
	Unit-16	Chess clock	168
	Unit-16	Chess glossary	169

Appendix 2: Field Experiment Setting

Appendix 3: 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**
	

(B) Wave 2 task

Today you will participate in two different activities. In each 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 bag with items]

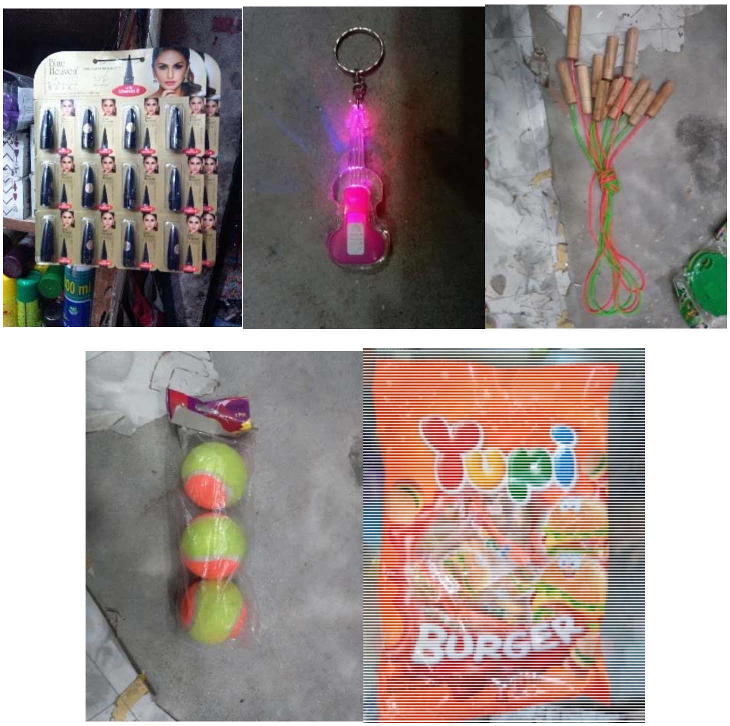
Items in the small bag cost two token. [show the items that are available in the small bag – take the time to ensure each child is aware of the variety of items in the bag]

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

Appendix 4: Time Preference Task

(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

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 amount 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.

Appendix 5: Attrition

Table E.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.