DISCUSSION PAPER SERIES

IZA DP No. 9370

Your Move:
The Effect of Chess on Mathematics Test Scores

Kamilla Gumede
Michael Rosholm

September 2015

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Your Move: The Effect of Chess on Mathematics Test Scores

Kamilla Gumede
TrygFonden's Centre for Child Research, Aarhus University

Michael Rosholm
TrygFonden's Centre for Child Research, Aarhus University
and IZA

Discussion Paper No. 9370
September 2015

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 9370
September 2015

ABSTRACT

Your Move: The Effect of Chess on Mathematics Test Scores*

We analyze the effect of substituting a weekly mathematics lessons in primary school grades 1-3 with a lesson in mathematics based on chess instruction. We use data from the City of Aarhus in Denmark, combining test score data with a comprehensive data base from administrative register. We use a difference-in-differences approach to investigate treatment effects on the treated and tend to find positive effects. Looking at sub groups, we find significant positive effects for native Danish children, while we find no effects for children of immigrants.

JEL Classification: I21, I28

Keywords: mathematics and chess, primary school, learning

Corresponding author:

Michael Rosholm
TrygFonden's Centre for Child Research
Aarhus University
Fuglesangs Alle 4
8210 Aarhus V
Denmark
E-mail: rom@econ.au.dk

* We are grateful to financial support from TrygFonden to TrygFonden's Centre for Child Research enabling this research, and we acknowledge useful comments from seminar participants at Aarhus University. Finally, we are grateful to the Danish School Chess Association and the City of Aarhus for cooperating with us on this evaluation, for providing crucial information on the intervention we study, and for making the math test score data available to us.

1. Introduction

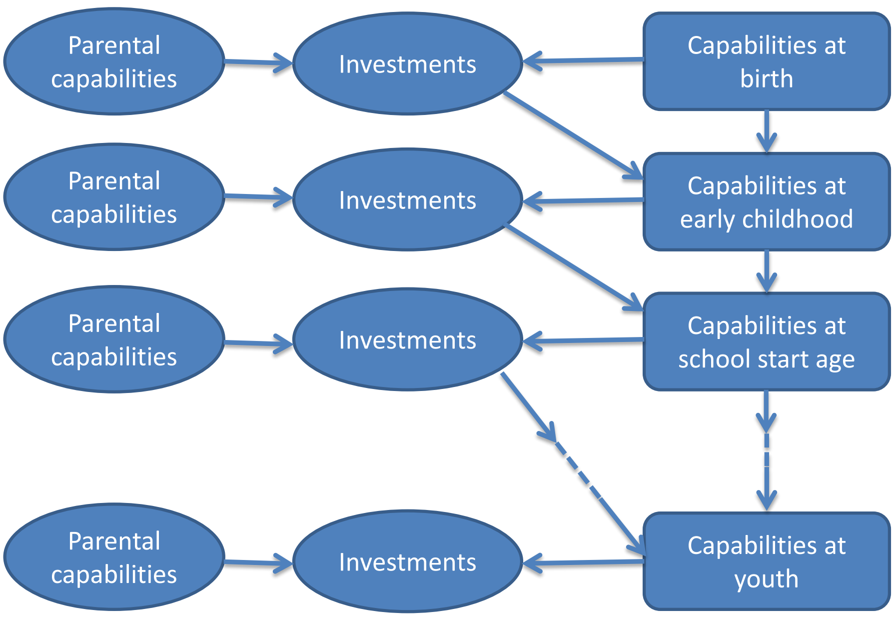
This paper studies the short run impacts of substituting a weekly 'normal' 45 minute math instruction with math instruction based on chess learning material in grade 1-3 in lower primary schools in the City of Aarhus, Denmark. We use a difference-in-differences approach controlling extensively for confounding factors.

The background for the study is that costs of primary and lower secondary schooling in Denmark are among the highest in OECD, one of the reasons being large amounts spent on children in special needs education; about one third of all the costs of primary and lower secondary education goes to special needs education (references), aimed at both students with learning disabilities but also students with behavioral problems. Hence, integration of special needs students into the ordinary classroom is high on the political agenda. Yet, looking at cross-country data from the PISA studies, Denmark is 'average' in OECD. In particular, we seem to have problems in lifting the weakest students and the very best students.

In the international literature, focus has recently switched towards increasingly focusing on non-cognitive skills when seeking explanations for educational/school failure. Factors such as personality/behavioral problems and lack of self-control/grit/conscientiousness are increasingly brought forth as explanations, see e.g. Almlund *et al.* (2011) and Moffitt *et al.* (2011).

The teaching of chess might be able to affect cognitive skills (in our case math skills) directly, as well as indirectly through non-cognitive skill formation. According to the life cycle skill formation model of Heckman (e.g. Heckman, 2012), a person has at any point in time a set of capabilities, cognitive as well as non-cognitive, that combine to make any investment in the person's capability formation more or less effective, see Figure 1.

Figure 1. The framework of life-cycle skill formation



Note: Adapted from Heckman (2012).

Chess is a game that requires calculation and planning ahead, which in turns requires the ability to concentrate and to memorize sequences of moves and resulting positions. It rewards the ability to exert patience and self-control, as 'quick' moves are often punished because they tend to lead to losses. Moreover, there are a set of rules of conduct during a chess game (you shake hands at the start and end of a game, you sit quietly during the game, you often discuss the game with your opponent and/or friends/teammates afterwards, hence teaching you how to learn from your mistakes and inspiring and illustrating the potential gains from learning).

Hence, learning chess may affect the ability to concentrate, the working memory, and other types of executive functions, as well as it may directly increase intelligence and problem solving abilities.

In this study, we report first results from a non-randomized trial and investigate how chess instruction affects the improvements in math test scores relative to a comparison group not receiving chess lessons. We control extensively for parental and child background using a rich data set based on administrative registers maintained by TrygFonden's Centre for Child Research at Aarhus University and located at Statistics Denmark.

We find that, on average, replacing one (out of four) weekly math instruction with an instruction based on chess learning material, during almost three quarters of a school year in grades 1-3 in primary school, leads to an improvement in subsequent math test scores of around 0.16-0.18 standard deviations. This is quite remarkable, given that a) the 'treatment as usual' given to the control group is an extra lecture in math instruction, and b) the 'normal' improvement during a school year in 3rd grade is in the order of 0.4 standard deviations.

Investigating differential impacts for subgroups, we find that the intervention appears effective at all levels of initial skills, that is, an interaction of treatment and pre-intervention test score is insignificant. We find no significant gender differences. We find no impact on 1st and 2nd generation immigrant children but fairly large impacts for native Danish children. Finally, it appears that the impact is larger in families where the mother has a qualifying education.

The remainder of the paper is organized as follows. In the next section, we briefly review the literature on the impacts of chess instruction, and discuss potential channels through which an impact on math test scores might appear; the theory of change. In Section 3, we present the intervention and derive some testable hypotheses based on the intervention and the theory of change, and in Section 4, the data used for the study is presented. Section 5 introduces the methodological approach, and in Section 6 the results are presented and discussed. Finally, Section 7 concludes and discusses policy implications.

2. Existing Evidence

In this section, we will first discuss the theoretical end empirical literature on the link between chess and learning, especially in mathematics but also other outcomes of potential relevance. By the end of the subsequent section, where the precise intervention is introduced, we use the evidence presented here in combination with the details on the intervention to derive a theory of change and some testable hypotheses.

What chess is and does

Chess is a sequential game, where the players make moves in turn with white and black pieces on the chess board with the aim of capturing the opponent's king. There are three possible outcomes of a chess game: you win, you draw (a tie), or you lose.

Since each player has 16 pieces with different types of mobility, it is a very complex game that requires calculation and planning ahead, which in turn requires the ability to concentrate and to memorize sequences of moves and resulting positions. It rewards the ability to exert patience and self-control, as 'quick' moves are often punished because they tend to lead to losses. Moreover, there are a set of rules of conduct during a chess game (you shake hands at the start and end of a game, you sit quietly during the game, you often discuss the game with your opponent and/or friends/teammates afterwards), hence teaching you how to learn from your mistakes and simultaneously demonstrating the potential gains from learning.

Still, it is an open question whether improvements in chess transfer into other domains. Boruch (2011) mentions that chess players do have uncommonly good memory and concentration abilities (see also Kirschner *et al.,* 2006), but notes that this may of course be entirely due to self-selection. Kirschner *et al.* (2006) also discuss the potential link to problem solving abilities, noting that expert chess players are better than novices at reproducing glimpsed board configurations from real chess games, but not when it comes to random board configurations. Based on this, they argue that persons good at problem solving are so because of experience stored in long term memory that can be used to determine the best course of action. They emphasize the importance of long-term memory to cognition, thus arguing against the transferability of e.g. problem solving skills in chess to other domains. On the other hand, Hong & Bart (2007) argue that chess playing requires higher order thinking skills, and that chess instruction is conducive to the acquisition of such skills. They consider these skills to be transferable, a point of view also supported by Ericsson & Staszewski (1989); Ericsson & Kintsch (1995); and Gobet & Simon (1996).

Berkman (2004) explicitly discusses the link between chess and mathematics – i.e. the issue of transferability of skills learnt at the chess board to the study of mathematics - and argues that chess promotes higher-order thinking skills, and that the analysis of chess positions has much in common with problem solving in mathematics. It works with concepts as correlation (deciding which piece is better to sacrifice at a given point in a game), it introduces the coordinate system, it introduces geometric concepts as rows and columns (called ranks and files in chess), diagonals and orthogonals, and it requires continuous calculation. It also develops visual memory, attention span (concentration), spatial reasoning skills, capacity to predict and anticipate consequences, critical thinking, self-confidence, self-respect, and problem solving skills (see also Boruch, 2011; Berkman, 2004; Hong & Bart, 2007; Campitelli & Gobet, 2008).

Bart (2014) argues that to play chess well, you need to understand and evaluate chess positions, taking into account the different mobility patterns of the pieces, requiring fluid intelligence and concentration capacity. You then have to formulate and evaluate possible moves, requiring executive functioning, recognition of patterns, and critical thinking. He goes on to argue that for these reasons, chess may lead to cognitive improvements.

Based on arguments similar to the findings above, Storey (2000) suggests using "chess as an instructional strategy for reinforcing skills such as concentration, problem identification, problem solving, planning strategies, creativity, and lucid thinking."[1]

Gobet & Campitelli (2006) review the existing empirical evidence on the relation between chess and several educational outcomes, and conclude - based on 7 studies, of which only two are published in peer reviewed journals - that a) it is still an open question whether chess instruction improves learning in other areas than chess, b) compulsory instruction is not recommended due to motivational problems, and c) the positive impacts found appear to be at the beginner's level, where skills beneficial to learning in general, such as the ability to concentrate, memory, and other types of executive functions, are improved. Even those conclusions are only tentative, based on our own reading of the literature, as most studies are seriously underpowered (numbers of observations range from 20 to 92 in these studies), and the technical details are not always very well described.

Bart (2014) summarizes a few studies performed more recently and/or not included by Gobet & Campitelli (2006), and is slightly more positive in his conclusions, stating that chess instruction has indeed positive effects on scholastic achievements. These studies tend to document positive causal links from chess instruction to mathematics achievement and non-verbal cognitive ability (Smith & Cage, 2000), from chess instruction to intelligence and problem solving ability (Aciego *et al.*, 2012), from chess to cognitive ability and math test scores (Kazemi *et al.*, 2012), from chess to math test scores for low ability students with IQs in the ranges of 70-85 (Scholz *et al.*, 2008), from chess to math test scores (and end-of-year grades) for pupils with special education needs (Barrett & Fish, 2011), and from chess to non-verbal intelligence for students at risk of academic failure (Hong & Bart, 2007). However, these studies also tend to have very small sample sizes.

Sigirtmac (2012) compares two non-randomly selected samples of children, one that received chess instruction at age 6 and one that did not. The two groups were matched on socioeconomic background but not on other variables. A significant effect on understanding of spatial concepts is found. This study has severe methodological problems, though.

An interesting study is Trinchero (2013), who investigates the impact of chess instruction on PISA math test scores. Using non-randomized data on children aged 8-10 in Italy and a raw difference-in-differences approach, he finds that chess instruction in school improves problem solving abilities. The estimated effect is statistically significant but rather small.

Boruch (2011) is to our knowledge the only study exploiting a sufficiently powered randomized trial to study the link from chess instruction to learning. The experiment took place in 33 Italian schools and consisted of 30 hours of chess instruction by certified staff. Each school participates with at least two 3rd grade classes, one of which, chosen at random, does not receive the intervention before 4th grade and therefore is a control group for the 3rd grade intervention. A pre- and post-test was administered to all classes. The main finding is that chess instruction increases math achievement by a third of a standard deviation. Foreign born pupils have significantly better impacts than native borns.

[1] Storey (2000), p. 47.

Potential indirect channels; non-cognitive skills

Chess instruction thus may potentially have both direct impacts on the capacity for learning mathematics, but there may also be indirect impacts operating through a set of non-cognitive skills. In the following, we briefly review a few studies on non-cognitive skills that we find particularly important. For a more general overview, we refer the reader to e.g. Almlund *et al.* (2011).

According to the discussion above, chess instruction may positively affect visual memory, attention span-persistence, concentration capacity, spatial reasoning skills, critical thinking, self-confidence, self-respect, executive functioning etc.

McLelland *et al.* (2012) study the relation between preschool attention span-persistence and various outcomes at ages 21 and 25. They find that, after controlling for academic achievementat age 7, kindergarten vocabulary skills, gender, maternal education and a few other variables, attention span-persistence at kindergarten age still significantly predicts math (and reading) achievements by age 21 and the odds of having completed college by age 25. The authors cite a long list of research documenting similar results for traits related to executive functions and self-control.

Dumontheil & Klingberg (2011) show that visuospatial working memory has significant positive effects on the ability to learn mathematics. They show that activity in the intraparietal sulcus (which plays a role in processing symbolic numerical information and visuospatial working memory) during visuospatial working memory tasks (which correlates with interindividual differences in visual memory capacity) predicts math abilities two years later independently of included behavioral measures.

Moffitt *et al.*, (2011) study how a very robust composite measure of self-control in childhood (age 3-11) predicts later life outcomes (at age 32 years), and find that low self-control predicts worse health (in terms of physical health and substance dependence/use problems), worse financial situation (lower SES, income, single-parent child rearing, and broad measures of financial problems), and crime (conviction), even after correcting for childhood family SES and IQ. They also provide suggestive evidence that at least part of the correlation is causal, by exploiting changes over the life course in self-control.

Almlund *et al.* (2011) summarizes the literature on ‘non-cognitive skills’ and find, among other things, that ‘Conscientiousness’ (in the ‘Big Five’ categorization of personality skills of psychologists, related to non-cognitive skills such as attention span and self-control) is important in determining how many total years of education individuals complete in their lifetimes. Locus of control and self-esteem are also important for adolescent schooling decisions. Moreover, the empirical evidence suggests that Conscientiousness may be as predictive as cognitive ability in predicting and possibly causing higher course grades. Conscientiousness plays an important role in predicting achievement tests above and beyond cognitive ability. ‘Conscientiousness’ also appears to be the most important Big Five trait in predicting late life health outcomes. Personality likely affects health through behaviors such as smoking, eating, and exercising. Studies that model the dynamic evolution of health over the life cycle find that personality affects health outcomes at least as much as cognitive measures. Few studies have examined the relationship between the Big Five and criminal behavior. The available evidence suggests that Big Five ‘Conscientiousness’ and ‘Agreeableness’ are important protective factors against criminal activity.

Hence, summarizing, there is evidence that chess instruction may affect math abilities directly through skill transferability, as well as indirectly through a number of 'non-cognitive skills' linked to personality traits.

3. The Intervention an Hypothesized Impacts

In this section, we first give some background on the Danish compulsory school system, then we present the intervention studies in this paper, and finally we present a set of hypothesized impacts of the intervention, based on the literature presented above as well as the intervention itself.

Background on the Danish School System

The Danish school system is inherently public. Around 16% of pupils in compulsory school attend private schools, but even these are heavily subsidized and have to live up to a centrally determined curriculum. There are 10 mandatory grades, from grade 0 to 9, and an optional grade 10. Grade 0 is a sort of preschool taking place in the schools, where pupils learn to go to school, but there is no explicit instruction in mathematics, Danish, or any other subject.

The typical school starting age is 6. Until a recent reform implemented August 2014, which dramatically lengthened the school day, students in grades 1-3 were typically in school attending lectures for about 21-25 hours per week.[2]

In grades 1-4, there would on average be around four 45 minute lectures of math instruction per week, and students have to go through a certain curriculum during these four years, focusing on calculation, pattern recognition, geometry, and basic problem solving.

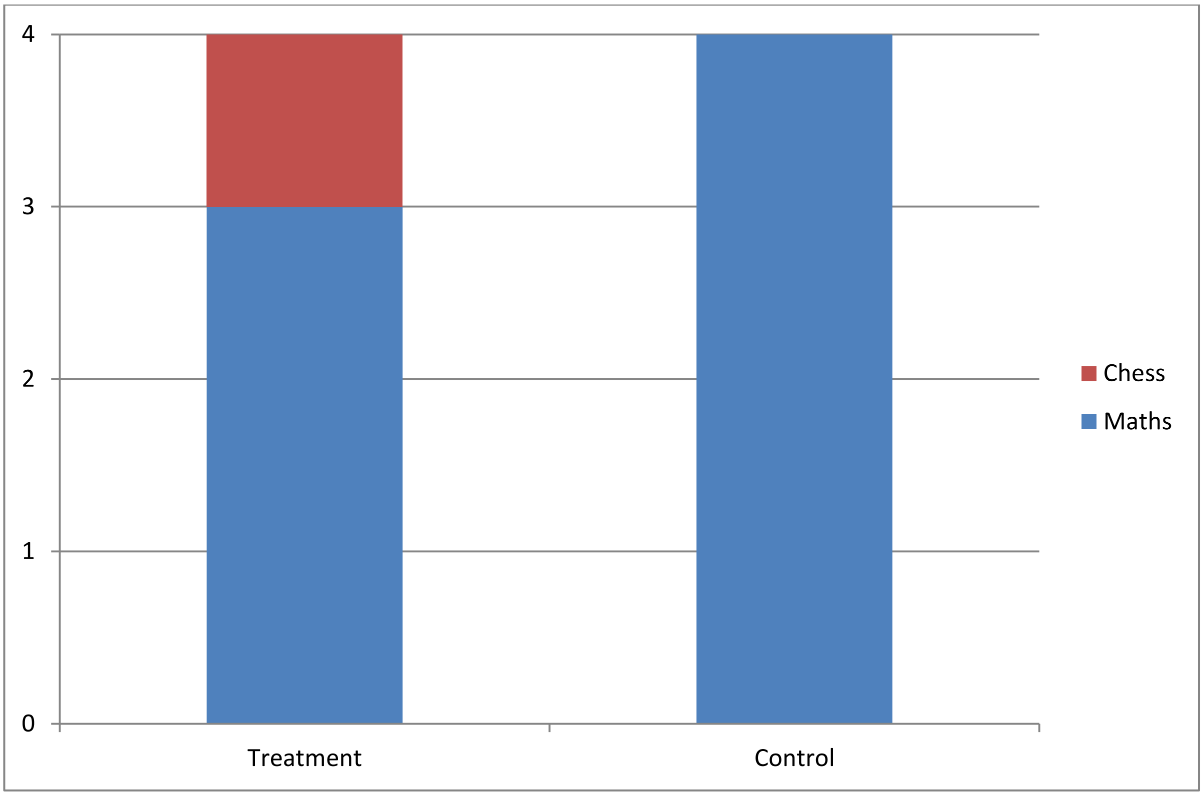
The Intervention

The intervention we study took place in five schools in the City of Aarhus, the 2nd largest city in Denmark. From these five selected schools, selected classes from grades 1 to 3 were chosen to participate in the study. Two of the schools are located in socially deprived sub-urban areas, two are in middle class suburban areas, and the fifth school is located close to downtown Aarhus, in a fairly middle-income area, but with some social housing.

For around three quarters of a school year, starting in January 2013 and ending mid October 2013, students in the treated classes would have one (in approximately four) weekly standard math lecture replaced by a weekly instruction based on chess. This is illustrated in Figure 2.

[2] After the reform, the minimum requirement is 30 hours of school attendance per week.

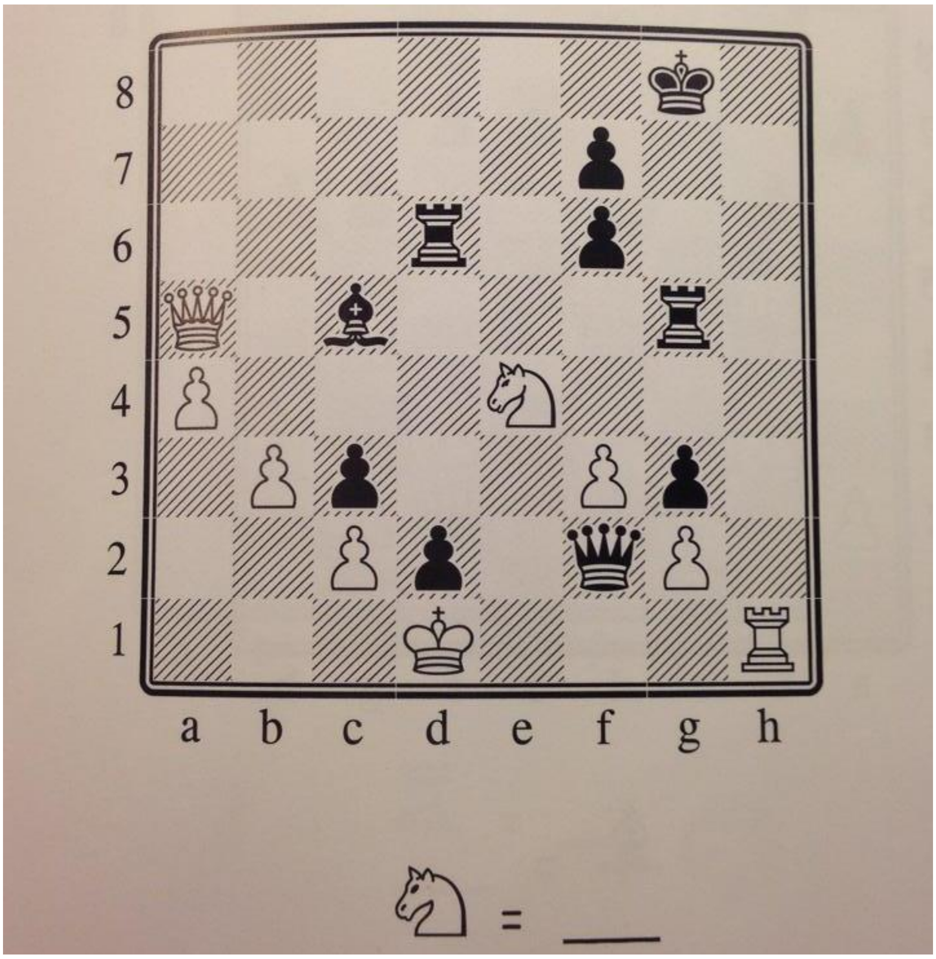
Figure 2. Approximate number of math related lectures in treatment and comparison groups



A dedicated mathematics teacher, who is also a club chess player, performed the teaching during treatment lectures in all five schools. The normal mathematics teacher was normally present in the classroom during chess lectures and would participate in playing if there were an odd number of students present in the classroom. Lectures consisted partly of instruction on the movements of chess pieces, and partly of practical chess playing exercises. The chess lectures were based on a book developed by the Danish School Chess Association (Dansk Skoleskak) called Chess+Math (Skak+Mat).[3] Figure 3 below depicts a typical exercise from the book. The instruction for the exercise is (translated from Danish): *"How many pieces can the knight take? Write your answer on the line below"*.

[3] In Danish, the word 'Mat' means 'mate' as it is used in chess, but is also an abbreviation for 'mathematics'.

Figure 3. A typical chess exercise from the book used for chess instruction



The intervention was originally intended to end before the summer holidays of 2013 (end of June, 2013), but due to a lock-out of the teachers as the result of a conflict in relation to central negotiations prior to the school reform of August 2014, all public schools were closed during most of April 2013. Hence, it was decided to extend the intervention into the next school year and end it by week 42 (mid October) of 2013, which is the autumn holiday week in Danish schools. In total, the students have thus had at most 30 math lectures replaced with a lecture of chess instruction.

Pupils in treatment as well as comparison classes in the same schools were administered a mathematics test immediately before the start of the intervention. Students were then administered a post-intervention test in November-December 2013, taking a test which was a grade level higher than the pre-intervention test.

The mathematics tests were designed to test students in calculation and geometry, pattern recognition (numbers and shapes), and basic problem solving. The calculation and geometry components are addition, subtraction, multiplication, division, basic equations (e.g. 4+x=9), and counting, say, the number of triangles in a cloud of shapes. Patten recognition is both number sequences (e.g. 1-2-4-8-16-_-_, 1-4-9-16-25 _ _, or 1-1-2-3-5-_-_) and finding shapes that fit into a sequence of shapes and colours. Finally, problem solving is of the type *"Mette arrives at a bridge. A troll is guarding it. He says 'if you want to pass the bridge and come back, I will double the amount of money you have in your pocket. Afterwards you have to give me 8 kroner'. Mette passes and gives the troll 8 kroner. Now Mette has no more money left in her pocket. How much did she have in her pocket when she arrived at the bridge?"* The tests were rather comprehensive in calculation and pattern recognition, while there was only one problem solving exercise.

The tests vary slightly from grade to grade. The test for 2nd grade has 4 questions on calculation giving a total of 14 points, it has one set of questions for pattern recognition giving a total of 10 points, and one problem solving exercise giving up to 4 points. For grades 1, 3 and 4, it is quite similar in terms of weighting.

In addition, there are three questions on well-being.

Theory of Change and Hypothesized Impacts

Since the intervention does not provide any extra teaching, rather it reduces the number of 'standard' math lectures and replaces them with chess instruction, it is not obvious *a priori* that the intervention will improve mathematics test scores, at least not in the short run.

From the literature study, we found evidence that chess instruction may affect mathematics abilities directly through skill transferability, as well as indirectly through a number of 'non-cognitive skills' linked to personality traits and executive functions. These insights, in combination with the description of the intervention above, basically form our theory of change.

Based on the literature and our theory of change, we hypothesize the following testable implications:

1. Chess instruction leads to improved mathematical abilities, especially in problem solving/pattern recognition tasks
2. Chess instruction has larger impacts for low initial ability students
3. Chess instruction has larger impacts for boys than for girls
4. Chess instruction has larger impacts for children of non-western ethnic origin
5. Chess instruction, through its potential impacts on non-cognitive skills, especially 'conscientiousness' may also affect school absence.

There are several other outcomes, such as test scores in other subjects, where no data are available (yet), where we might expect to find impacts, but since we are not able to test these with the data at hand, we shall not devote more time discussing potential impacts on other outcomes.

4. Data

The population under study consists of students in five public primary schools in the City of Aarhus, the 2nd largest city in Denmark. From these 5 schools, selected classes from grades 1 to 3 were chosen to participate in the study.

The mathematics tests were administered by the chess teacher in collaboration with an employee from the City's school administration. The tests were filled in by hand, and subsequently manually transferred to an excel spreadsheet by a student assistant at the school administration (who was not aware of the treatments status of the children). The pre- and post-intervention test scores were standardized to have mean zero and variance one.

For 482 children we had access to both pre- and post-intervention test scores and were able to match them on the basis of their names to their social security number and hence able to merge the test scores with background characteristics from a large set of administrative registers maintained by Statistics Denmark. Of the 482, 323 had one (in four) weekly mathematics lecture (45 minutes) replaced with chess instruction for the spring semester of 2013 and the first half of the fall semester (until mid october). 159 children in the comparison group did not participate in the program but had instead four weekly math lectures (treatment as usual). Unfortunately, the classes selected for participation were not randomly selected, but rather selected by the school principal. The criteria for selection of classes to treatment and control groups were unclear, but subsequent interviews suggest that they were close to random, or at least based on factors not immediately related to the children's capacity for learning.

The information on the test-scores and treatment status were merged with a large data base collected by TrygFonden's Centre for Child Research at Aarhus University based on administrative registers. The data base is physically located at Statistics Denmark, and contains information on parental background, immigrant status, gender, a few schooling outcomes (including days of school absence) and is used in order to control as comprehensively as possible for children's characteristics and parental background.

In terms of the included observable characteristics, Table 1 compares averages for the treated children and for children in the comparison group.[4]

Table 1. Descriptive statistics, background variables.

Variable	Treatment group	Comparison group
Standardized pre-intervention test-score	0.00	0.05
Boy	0.54	0.50
Girl	0.46	0.50
Age	9.57	9.45
1st or 2nd generation immigrant	0.28	0.25
Days of school absence 2012	9.21	9.94
Grade 1	0.19	0.31
Grade 2	0.45	0.33
Grade 3	0.36	0.36
# siblings	1.46	1.53
Age of mother	40.53	40.42
Mother lower secondary school	0.42	0.41
Mother high school	0.07	0.08
Mother vocational education	0.27	0.22
Mother short academic education	0.05	0.06
Mother medium academic education	0.07	0.08
Mother masters education or more	0.11	0.14
Mother's average ann. earnings past 3 years, DKK	195,276	188,578
Mother not working 2011	0,29	0,30

[4] We also have information on the schools attended, but for confidentiality concerns these are not reported here.

Father present	0.79	0.76
Age of father	43.28	42.28
Father lower secondary school	0.19	0.15
Father high school	0.05	0.04
Father vocational education	0.30	0.30
Father short academic education	0.08	0.10
Father medium academic education	0.14	0.13
Father masters education or more	0.17	0.18
Father's average ann. earnings past 3 years, DKK	278,564	284,405
Father not working 2011	0.19	0.20
N	323	159

Despite the fact that the distribution across grades differ quite a bit, with more treated children in grade 2 and fewer in grade 1, the children (and their parents) do not differ that much. None of the differences are statistically significant. The comparison group appears to have slightly higher test scores *ex ante*, but the standardized difference is small and not statistically significant. Moreover, mothers of children in the comparison group appear to be slightly better educated, although none of these differences are significant either. Still, given the potential non-random nature of the allocation to treatment and comparison groups, it is important to control for these background factors. In this respect, access to *ex ante* test scores is a large advantage, as it allows us to exploit a difference-in-differences strategy for identifying the causal impact of the intervention.

Figures 4 and 5 plot kernel density estimates of the distribution of the standardized pre- and post-intervention test-scores, respectively.

Figure 4. Pre-intervention standardized math test-scores.

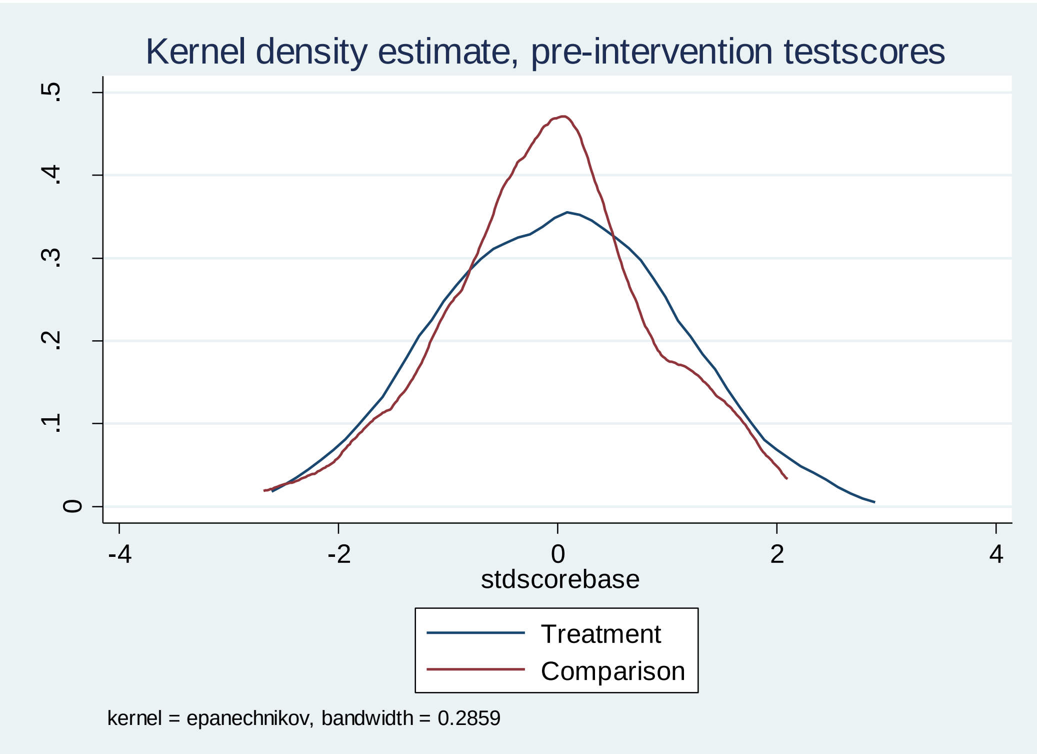
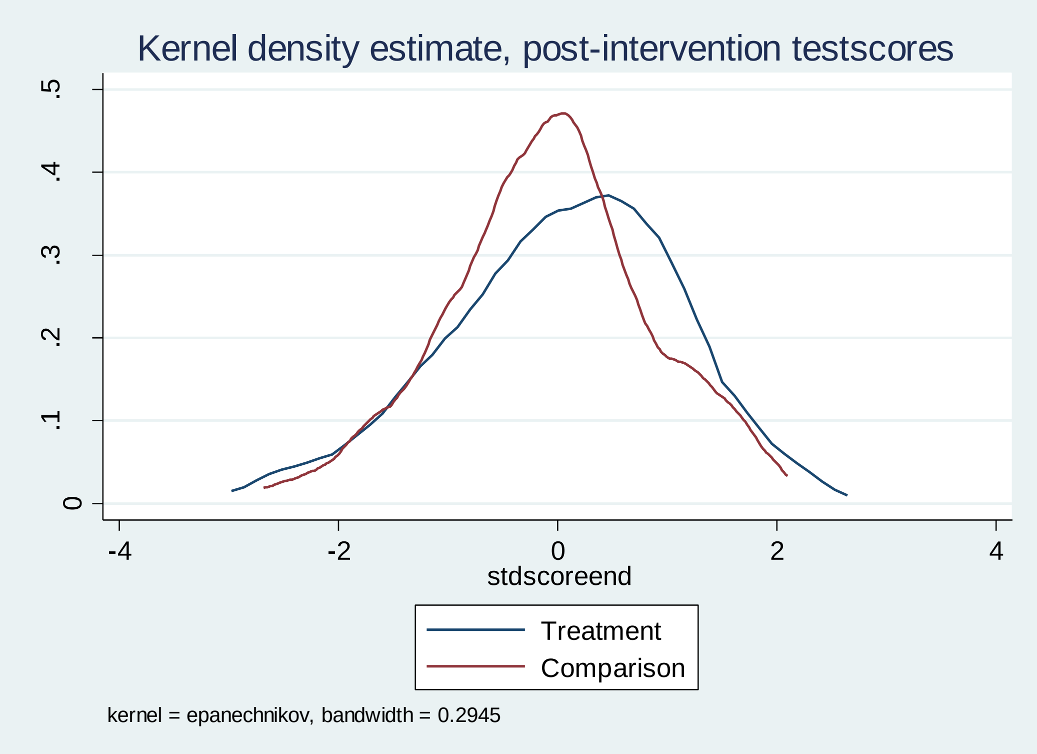


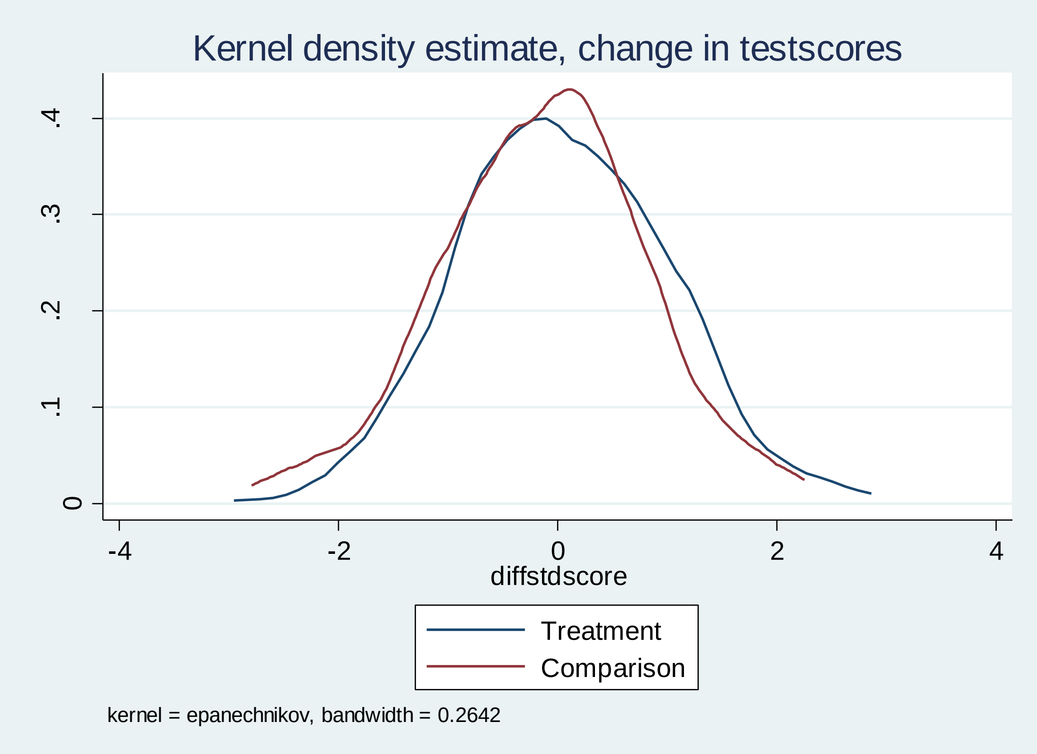
Figure 5. Post-intervention standardized math test-scores.



The pre-intervention standardized math test scores distribution of the treated group has more variance than in the control group, but as reported in Table 1, the means are almost identical. In the post-intervention distribution, there is almost a situation of first-order stochastic dominance of the treatment group distribution compared to the comparison group. The averages in the two distributions differ by 0.14 in favour of the treatment group, but are not significantly different.

Figure 6 plots the kernel density estimate of the change in the test score from pre- to post-intervention.

Figure 6. Change in standardized math test-scores.



Again, the distribution of the treatment group suggests almost first-order stochastic dominance, and the raw average difference-in-differences is 0.18 and statistically significant (at the chosen 95% level).

5. Methodology

Since the treatment and comparison groups are not randomly selected, the raw test-score differences between treatment and comparison groups cannot be given a causal interpretation. The raw difference-in-differences estimate (0.18 and statistically significant), can be interpreted causally under an assumption of common trends in learning speeds. In order to strengthen the causal interpretation, we would nevertheless prefer to be able to control for school factors, parental and child background variables.

Denote by Y_{it} the outcome of interest (e.g. the math test score) for individual *i* at time *t*. Let *t=1* denote the post-intervention period, and let *t=0* denote the pre-intervention period. Hence, Y_{i0} denotes the pre-intervention test score of student *i*, and Y_{i1} the post-intervention test score. Let D_i denote the treatment status of student *i* (D=1 denotes treatment, 0 comparison), and let X_{i0} denote a set of background characteristics (specifically, the variables reported in Table 1, except for the pre-intervention test score). Finally, let S_i denote a set of school indicators (fixed effects).

Our first model of interest then seeks to explain the post-treatment outcome of interest with the pre-treatment outcome of interest, a treatment indicator, a set of explanatory background characteristics, a set of school fixed effects, and a residual error term:

$$Y_{i1} = \mu + \alpha Y_{i0} + \gamma D_i + \beta X_{i0} + \delta S_i + \varepsilon_i \qquad (1)$$

Here, the level of the test score is explained by the pre-intervention test score, the treatment status, the school, and some explanatory variables.

An alternative formulation of the model would be 'learning' model, where the change in the test score is explained by the same variables (excluding the pre-intervention test score):

$$Y_{i1} - Y_{i0} = \tau + \theta D_i + \vartheta X_{i0} + \pi S_i + \rho_i \qquad (2)$$

This model is slightly different from (1) in the sense that it is the change in the test scores, that is, the speed of acquisition of mathematics capability, that is modeled. The learning speed, not the competence level, is in this specification a function of treatment and of some background characteristics. Hence, it is the capacity for learning quickly that might be affected by treatment, the home environment, the school, school attendance, etc., in equation (2).

In the results section, results for both specifications are reported, as they both have some merit, and moreover, it may serve as a test of robustness of the results to the empirical specification. As we have no prior reason for strongly believing that replacing mathematics with chess should improve math test scores in the short run, we choose to use two sided test statistics. Moreover, we have chosen the conventional 5% significance level.

Identification and validity

Ideally, we would have liked to perform a randomized trial, but since this was not an option, we have to employ this difference-in-differences evaluation strategy. Heckman *et al.* (1998) argue that this is often the best strategy when assignment to treatment is not exogenous.

The estimated effects are average treatment effects on the treated, which are identified under the assumption of parallel trends, that is, in the absence of the intervention, the level of mathematics capability or the speed of learning would be the same in the treatment and control groups, given the observed characteristics. Unfortunately, this assumption is not testable, as we have no information on previous test score results, since these were not digitalized by the City of Aarhus.

Moreover, the study has design flaws; first, only one teacher was involved in teaching the chess curriculum, and therefore, we cannot be certain that any effects can be found if the intervention were to be scaled up. Second, two teachers were present in the classroom during chess lectures, implying that an effect could also be caused by the additional teacher (a two-teacher effect). [5]

Finally, the study was severely underpowered, implying that it is unlikely to find significant impacts. There are 34 different classes involved in the intervention, implying that with an R-squared of 0.4 (as we find in the best of cases), we would have minimum detectable effect effects sizes around 0.23 with a power of 0.8 and a chosen significance level of 0.95.

Hence, we prefer to think of the project as a pilot study or a demonstration project of the potential beneficial effects of chess instruction to be supplemented with a properly designed randomized trial.

[5] Andersen *et al.* (2015) find positive but smaller effects than ours of a considerably more intensive two-teacher intervention, suggesting that a two-teacher effect is perhaps not the most likely causal mechanism behind our results.

6. Results

In this section, we present the results of the two models described in the previous section. The outcome variables – the mathematics test-scores - are standardized to have mean zero and standard deviation 1, so the estimated impacts are standardized effect sizes.

First, Table 2 shows estimation results, where the post-intervention test score is the dependent variable (corresponding to equation 1 in section 5). We show four different models; one where we do not include any explanatory variables beyond the indicator for having received a weekly chess-lesson instead of a weekly math lesson (model 1), one where we also include the pre-intervention test score (model 2), a model where in addition we add a set of child characteristics (model 3), and finally model 4 also adds characteristics of the mother and the father.

Table 2. Estimation results, post intervention test-scores.

Model	Impact estimate	Standard error	R-squared
1: Only chess dummy	0.14	0.10	0.00
2: 1+pre-intervention test score	**0.16**	0.08	0.33
3: 2+child and school characteristics	0.12	0.08	0.39
4: 3+mother's and father's characteristics	*0.14*	0.08	0.43

Note: Child characteristics include grade dummies, gender dummy, precise age, number of siblings and ethnicity. School characters include grade and school dummies. Mother's characteristics include age, being present in the household, 6 educational level dummies, gross earnings the previous year, and an indicator for not having worked. Father's characteristics are the same. **Bold numbers** imply statistical significance at the 95% level, *italicized numbers* indicate significance at the 90% level.

First of all, note that the pre-intervention test-scores explain about one third of the variation in the post-intervention test scores. This is rather low compared to some other studies in the literature on test scores, but these children are also fairly young, so the added learning during three quarters of a school year should be quite significant.

The estimated effect size becomes significant once we control for pre-intervention test scores, but the statistical significance disappears again when additional control variables are added. The effect sizes are 0.12-0.16.

Table 3 shows the effects from the learning model (equation 2 in section 5). When only the indicator for receiving chess instruction in class is included, the effect is significant, but the impacts decline slightly once additional characteristics are included, rendering it insignificant at the 95% level. The effect sizes are 0.16-0.18 in this model.

Table 3. Estimation results, change in test-scores.

Model	Impact estimate	Standard error	R-squared
2: Only chess dummy	**0.18**	0.09	0.01
3: 2+child and school characteristics	*0.16*	0.09	0.04
4: 3+mother's and father's characteristics	*0.17*	0.09	0.08

Note: Child characteristics include grade dummies, gender dummy, precise age, number of siblings and ethnicity. School characters include grade and school dummies. Mother's characteristics include age, being present in the household, 6 educational level dummies, gross earnings the previous year, and an indicator for not having worked. Father's characteristics are the same. **Bold numbers** imply statistical significance at the 95% level, *italicized numbers* indicate significance at the 90% level.

According to Lipsey *et al.* (2012), when evaluating test score results, a student learns (progresses) during a school year what corresponds to 0.5 units of a standard deviation in grade 3 and around 1 standard deviation in grade 1. Hence, the estimated effects of replacing an hourly math lecture with a chess lecture for less than three quarters of a school year corresponds to an additional learning similar to around one third of a school year of additional teaching of mathematics in grade 3 and one sixth of a school year in grade 1, if effects are homogenous across grades. Given that the weekly number of mathematics lectures in grades 1-3 are on average 3.75 (450 hours in total during the 3 years with 40 weeks of teaching during the year), this is quite impressive and suggests that the children really learn some capabilities during these lectures, which they can use more generally in school; it is hard to imagine such a large impact without positive external spillover effects on the ability to learn during the remaining three weekly mathematics lectures. We imagine, thus, that the impact operates through improved ability to concentrate and exert self-control, to memorize, to solve problems, and perhaps also through improved behavior (less noise in the classroom etc.). Most of these hypotheses are untestable, unfortunately. When data from national tests in Danish become available, we will be able to investigate if there are any impacts on learning Danish.

Table 4 estimates the effects of chess instruction separately by grade level. There are no significant differences – nor are there significant impacts - across grades, but this is probably due to the lack of sufficient statistical power, see section 5.

Table 4. Impact estimates by grade

	Model 1 Post-intervention test score effects	**Model 2 Change in test-score effects**
Chess dummy	0.10 (0.13)	0.23 (0.16)
Chess dummy * Grade 1	0.09 (0.20)	0.04 (0.24)
Chess dummy * Grade 2	0.04 (0.19)	-0.22 (0.22)

Note: Results are from the full model with all available background characteristics (i.e. model 4 in Tables 2 and 3). **Bold numbers** imply statistical significance at the 95% level, *italicized numbers* indicate significance at the 90% level.

In Table 5, we estimate a model with an interaction between the chess instruction indicator and an indicator for having below average pre-intervention test score. The interaction coefficient is negative but very small and not statistically significant. This might of course again be due to lack of statistical power, but at least we do not find that academically weaker children are significantly less affected by the intervention.

Table 5. Impact interaction with pre-intervention score

	Model 1 Post-intervention test score effects	**Model 2 Change in test-score effects**
Chess dummy	*0.20*	0.18
	(0.11)	(0.12)
Chess dummy * 1{pre score < avg}	-0.11	-0.03
	(0.16)	(0.17)

Note: Results are from the full model with all available background characteristics (i.e. model 4 in Tables 2 and 3). **Bold numbers** imply statistical significance at the 95% level, *italicized numbers* indicate significance at the 90% level.

In Table 6, we report effects with gender interactions. We find positive impacts for the reference group of boys, while there is a negative interaction term for girls, but it is not statistically significant, although it is fairly large in one of the models.

Table 6. Impacts by gender

	Post-intervention test score effects	**Change in test-score effects**
Chess dummy	**0.22**	0.20
	(0.11)	(0.13)
Chess dummy * girl	-0.17	-0.06
	(0.16)	(0.19)

Note: Results are from the full model with all available background characteristics (i.e. model 4 in Tables 2 and 3). **Bold numbers** imply statistical significance at the 95% level, *italicized numbers* indicate significance at the 90% level.

Table 7 reports impacts by immigrant status. There is an overall positive effect for the reference group of native Danish children, while the interaction term for 1st and 2nd generation immigrants is significantly negative, implying that they are not positively affected by the chess instructions, contrary to our *a priori* expectations. In fact, the net effect for 1st and 2nd generation immigrant children is negative, albeit statistically insignificant.

Table 7. Impacts by immigrant status

	Post-intervention test score effects	**Change in test-score effects**
Chess dummy	**0.23**	**0.24**
	(0.09)	(0.11)
Chess dummy * 1st or 2nd gen. immigrant	**-0.39**	-0.31
	(0.18)	(0.22)

Note: Results are from the full model with all available background characteristics (i.e. model 4 in Tables 2 and 3). **Bold numbers** imply statistical significance at the 95% level, *italicized numbers* indicate significance at the 90% level.

In Table 8, we report effects on the raw – non-normalized – test scores overall and separately by domains; pattern recognition, problem solving, and calculation. There is a significant positive impact only on pattern recognition, which was also an *a priori* hypothesis. On the contrary, there are no effects on problem solving, which we had also expected. The effects on calculation are almost as large as those on pattern recognition, but they are far from being statistically significant, due to a larger variance of test results in calculation.

Table 8. Impacts by different math domains – non-normalized test-scores

	Post-intervention test score effects	Change in test-score effects
Overall	*0.91*	*1.01*
	(0.54)	(0.57)
Pattern recognition	**0.51**	*0.43*
	(0.17)	(0.24)
Problem solving	-0.11	-0.01
	(0.16)	(0.21)
Calculation	0.47	0.56
	(0.45)	(0.46)

Note: Results are from the full model with all available background characteristics (i.e. model 4 in Tables 2 and 3). **Bold numbers** imply statistical significance at the 95% level, *italicized numbers* indicate significance at the 90% level.

We did not find any evidence of impacts on days of school absence, neither during nor after the end of the intervention period.

7. Conclusion

We find that replacing a weekly lecture of traditional mathematics with one based on chess instruction tends to increase subsequent results in math test scores. This holds especially for native Danish children, while for immigrant children, there is no impact.

The effect size corresponds roughly to what children in grade 3 learn during one third of a school year, that is, it is quite a remarkable effect, given that these children did not attend extra lectures; in fact they had fewer ordinary mathematics lectures than the comparison classes.

Regarding the hypotheses formed based on the literature review and intervention design, we did find positive effects on mathematics achievements in general, and in particular on pattern recognition. We did not, however, find any impact on problem solving. In retrospect, this may be due to there being only one problem solving exercise, which scored either 0 or 4 points. Hence, there was not a lot of variation in this particular outcome, compared to the two other outcomes. On the other hand, it may also just be the case that the particular way in which chess was used in the teaching material was more conducive to improving pattern recognition tasks. Regarding the hypothesis that chess instruction would be more beneficial to children with low initial mathematics skills, we did not find evidence of this. If anything, the results, albeit insignificant, pointed in the opposite direction. We did find a tendency to larger effects for boy than girls, but no significant differences were found. It would have been interesting to study also the importance of the gender composition in the classroom, but the sample was not nearly large enough to warrant such analyses. Contrary to what we had hypothesized, we found no (tending towards negative) impacts for children of immigrant origin, while there were rather large impacts for native Danish children, corresponding to around half a school year of mathematics learning. The mathematics teacher was a native dane, which may have played a role. Finally, we found no impact on school absence.

We cannot say much about the mechanisms behind the results, since children were only tested in mathematics, but the literature study suggested some mechanisms regarding what chess does; there may be a direct effect on mathematics ability, which could manifest itself in improved pattern recognition and problem solving abilities. This is to some extent confirmed. An indirect effect would go via the ability to concentrate, exert self-control, grit, conscientiousness or whatever you wish to call it. The size of the effect – in combination with the fact that the treated group did not receive extra mathematics, on the contrary – suggests that students gained more from all their mathematics lessons, or put differently, that what they learned during chess lessons generally enabled them to better understand mathematics.

In terms of costs and benefits, the available data does not enable us to conduct a cost benefit analysis. On the costs side, there are the additional learning material used in form of the book used, chess boards and pieces, chess demonstration boards and pieces but to the extent that these can be shared between classes in a school, and used for several years, this initial investment does not seem very large. In the experiment, an additional teacher was required, but if the intervention was to be scaled up, rather than costs of extra teachers, there would be costs of training mathematics teachers to use the books. On the benefit side, there would be learning gains and their consequences in the form of qualifying education obtained, life time earnings improvements, changes in health and crime outcomes due to changed education and employment situation. These benefits are all quite intangible in the sense that we do not know how primary school mathematics improvements affect these outcomes. Still, Joensen & Nielsen (2009) find that taking high level mathematics in high school causally leads to higher earnings in later life, so to the extent that primary school mathematics improvements are permanent and lead to more taking high level mathematics in high school, the gains may be considerable. The argument of dynamic complementarity (cf. Cunha & Heckman, 2007) suggests that early leaning interventions are more effective, because they also make later learning investments more effective; learning begets learning.

There are a few caveats to these conclusions, which are therefore only preliminary, awaiting a properly designed randomized trial. First, the study was severely underpowered, and selection of treated classrooms was not random. Second, only one teacher was involved in teaching the chess lessons. He might have been an excellent teacher, so that the result we find is really no more than a teacher fixed effect. In any case, the implication is that the external validity of these results is not large. Moreover, since the normal mathematics teacher was mostly also present in the classroom during chess lessons, there may also be a two- teacher effect operating. Finally, we only have access to mathematics test score results, so we cannot infer anything as to the mechanisms through which the effects manifest.

References

Almlund, M., A. L. Duckworth, J. J. Heckman & T. Kautz (2011), Personality psychology and Economics. In E.A. Hanushek, S. Machin & S. Woessman (eds). Handbook of the Economics of Education, Amsterdam: Elsevier.

Andersen, S. C., L. V. Beuchert-Petersen, H. S. Nielsen & M.K. Thomsen (2015), 2L Rapport: Undersøgelse af effekten af tolærerordninger. Politica (forthcoming)

Bart, W. M. (2014), On the effects of chess training on scholastic achievement. Frontiers in Psychology, Vol. 5, Article 762.

Berkman, R. M. (2004). The Chess and Mathematics Connection: More than Just a Game. Mathematics Teaching in the Middle School, Vol. 9, 246-250.

Boruch, R. (2011), Does Playing Chess Improve Math Learning? Promising (and Inexpensive) Results from Italy. Unpublished Manuscript.

Cunha, F. and J. J. Heckman (2007), The technology of skill formation. *American Economic Review*, Vol. 97, 31–47.

Dumontheil, i. & T. Klingberg (2011), Brain Activity during a Visuospatial Working Memory Task Predicts Arithmetical Performance 2 Years Later. Cerebral Cortex, doi:10.1093/cercor/bhr175.

Ericsson, K. & W. Kintsch (1995), Long-term working memory, *Psychological Review, 102,* 211-245.

Ericsson, K. & J. Staszewski (1989), Skilled memory and expertise: Mechanisms of exceptional performance. In D. Klahr and K. Kotovsky (Eds.), *Complex Information Processing: The Impact of Herbert A. Simon.* (pp. 235-267). Hillsdale, N.J.: Erlbaum.

Gobet, F. & G. Campitelli (2006), Educational benefits of chess instruction: A critical review. In T. Redman (ed), Chess and Education: Selected Essays from the Koltanowski Conference, Dallas TX: Chess program at the University of Texas at Dallas, 124-143.

Gobet, F., & Simon, H. (1996). The roles of recognition processes and look-ahead search in time-constrained expert problem solving: Evidence from grandmaster-level chess. *Psychological Science, 7,* 52-55.

Heckman, J. J., H. Ichimura, J. Smith & P. Todd (1998), Characterizing Selection Bias Using Experimental Data. *Econometrica*, Vol. 66, 1017-1098.

Heckman, J.J. (2012), The Economics and Econometrics of Human Development Part I: A Framework for Analyzing Human Development Over The Life Cycle. Presentation held at ‘TrygFonden Multi-Disciplinary Symposium Improving the Well-Being of Children and Youth’, January 29, 2012

Joensen, J. S. & H. S. Nielsen (2009), Is there a Causal Effect of High School Math on Labor Market Outcomes? *Journal of Human Resources*, Vol. 44, 171-198.

Kirschner, P. A., J. Sweller & R. E. Clark (2006), Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching. Educational Psychologist, Vol 41, No 2, 75-86.

Lipsey, M. W., K. Puzio, C. Yun, M. A. Hebert, K. Steinka-Fry, M. W. Cole, M. Roberts, K. S. Anthony & M. D. Busick (2012), Translating the Statistical Representation of the Effects of Education Interventions Into More Readily Interpretable Forms. US Department of Education, NCSER 2013-3000.

Moffitt, T. E., L. Arseneault, D. Belsky, N. Dickson, R. J. Hancox, H. Harrington, R. Houts, R. Poulton, B. W. Roberts, S. Ross, M. R. Sears, M. Thomson & A. Caspi (2011), A gradient of self-control predicts health, wealth, and public safety. Proceedings of the National Academy of Sciences of the United States of America, Vol. 108, 2693-2698.

Sigirtmac, A. D. (2012), Does chess training affect conceptual development of six-year-old children in Turkey? Early Childhood Development and Care, Vol. 182, 797-806.

Trinchero, R. (2013), Can chess training improve Pisa scores in mathematics? An experiment in Italian primary schools. Kasparov Chess Foundation Europe.