Is there a link between chess and arithmetic?

The present study had two aims: to evaluate the effects of primary school chess teaching on academic and cognitive outcomes and to make an analysis of the relationship between chess and arithmetic skills.

A series of evaluation studies have been made on the effects of chess teaching on performance in the academic, cognitive and behavioural domains. Gobet and Campitelli (2006) reported an analysis of studies on whether the teaching of chess generalized to domains such as mathematics, reading and reasoning. They found only a few investigations which met their criterions for high-quality evaluations studies. Because of the methodological shortcomings they observed, the authors concluded that there was no reliable evidence favoring the hypothesis that skills developed during chess study transfer to other skills.

In a literature review eight years later (Nicotera & Stuit, 2014), 24 empirical studies of chess teaching were categorized into three research quality levels. The main criteria were: teaching of chess in-school or after-school; use of academic, cognitive or behavioural outcome measures; inclusion of comparison groups in addition to the training groups; and pre- and post-testing in both the experimental and the control groups. Then effect sizes were computed for the studies in each of the three categories. The results suggested that chess intervention, both after-school and in-school, resulted in improved mathematical performance. But in two of the six most rigorously designed studies the authors (Christiaen, 1976; Scholz, et. al., 2008, in Nicotera & Stuit, 2014) did not find a statistically significant effect on mathematics. Both Gobet and Campitelli (2006) as well as Nicotera and Stuit (2014) voiced concerns over methodological shortcomings and also presented suggestions for improvement. One purpose with the present study was to follow these methodological advices as closely as possible in a renewed test of the effects of chess teaching on arithmetic.

As an alternative research strategy data can be collected by means of tests and subjected to psychometric analyses of interrelations between chess skill and test performances. A recurring finding in such studies has been correlations between arithmetic and chess performance (Grabner, Stern & Neubauer, 2007; Doll & Mayr, 1987, in Grabner et. al., 2007). Of particular interest in the latter study was the finding of a strong correlation between chess and number series performance. This result agrees with the hypothesis tested in the present study, namely the mental number sequence hypothesis (e.g., Johansson, 2013). This hypothesis states that when the mental number sequence skill has reached a threshold level (Johansson, 2013), it can be used for the solution of arithmetic problems. Results reported by Johansson (2005a, 2005b) suggest that early arithmetic solution strategies based on counting units (e.g., fingers or counters) are later supplemented with strategies based on representing numbers as numerals in the mental number sequence. It has been found (Johansson, 2005a; 2005b; 2013) that the ability to count backwards (e.g., “count backwards from 13 to 8”), solving doubles problems (2+2=?; 3+3=?, etc) and solving backwards number series problems (e.g., “Write the numbers that follow: 28, 25, 22, __, __, __, __, __) strongly predicts both number of correctly solved arithmetic problems and strategy used to solve the problems.

Children solving many number series problems were found to make frequent use of the immediate recall and the jump strategies, whereas children solving only a few such problems often used counting strategies (e.g., Johansson, 2005a). The jump strategy is interpreted to mean that the children are able to represent numbers as numerals in the mental number sequence. For example, when given the problem “Let’s say you have twenty-three apples and gives away six. How many do you have left?” the children mastering the mental number sequence translate the numbers into numerals “23 – 6” and “6” is segmented into “3 + 3”. This allows the solution “First I took away 3 from 23 which give 20, then 3 from 20 – 17”. Children mastering direct recall report something like “I know that 23 – 6 makes 17”. In contrast, children with a low level of mental number sequence skill often use the fingers to count one-by-one to 23 and then count backwards, also on the fingers, in six steps. Empirical data demonstrate that the former solution procedures are much more efficient, defined in terms of number of correct solutions, than the second one (e.g., Johansson, 2005a; 2005b).

These two types of strategies are assumed to reflect two qualitatively different ways of representing numbers: as concrete or imagined units (e.g., real or imagined fingers), or as meaningful entities - as numerals in the mental number sequence. In the latter case meaningfulness means that a given number is understood in terms of its relation to other numbers, e.g., the number 54 is understood as 1 more the 53, 10 less than 64, etc. The change from the first to the second way of understanding numbers is not a question of a growth of a concrete into an abstract understanding of number; instead, it is a question of a supplement of a concrete but limited way of understanding numbers – as units – with a still more concrete and flexible way of understanding numbers – as numerals. Numbers has become numerals.

In the discussion of the psychology of chess similar mechanisms have been suggested, that is, the development from understanding chess positions as some unrelated pieces occupying some squares on the chessboard to the immediate perception of meaningful chess relationships, where the latter allows a flexible search for and evaluation of efficient solution procedures (e.g., Holding, 1992; Charness, Reingold, Pomplum, & Stampe, 2001; Linhares, 2005).

The above line of reasoning means that the roots to the solution procedures used by children mastering the mental number sequence may be closely related to those behind solution procedures used by chess players whose level of skill has surpassed a given threshold. In both cases it is assumed to be a question of immediate perception of meaningful relationships. Understanding numbers as numerals in the mental number sequence means that a given number gets its meaning from other numbers in the sequence. In a similar way the meaning of a piece configuration is given by the relation between these pieces and the other pieces on the chess-board.

To test these speculations, data were collected that would allow psychometric analyses of the existence of such a link. Arithmetic skills studies made a few years ago (e.g., Johansson, 2013) demonstrated that this threshold level can be defined in terms of the prerequisite skills of numeral writing and number word sequence (Johansson, 2005b). In the case of chess no definition in terms of prerequisite skills has been found, instead the capacity to solve simple problems was taken as a measure of perceiving meaningful chess relationships.

Besides arithmetic, many researchers have investigated the visuospatial and reading components of chess play. De Groot (1965) and Chase and Simon (1973a, 1973b) have emphasized the role of pattern recognition for strong chess players. Studies of blindfolded chess point to the importance of a visual imagery component (Chabris & Hearst, 2003; Saariluoma & Kalakoski, 1998). It has also been found (Horgan & Morgan, 1990; Frydman & Lynn, 1992) that chess skill in children is related to the performance on visuospatial IQ-tests. On the other hand, there are studies that have failed to support the hypothesis of a visuospatial component (Nicotera & Stuit, 2014; Grabner, et. al., 2007; and Unterrainer, Kaller, Halsband, & Rahm, 2006). Also studies of the role of the reading component have produced inconclusive results (Nicotera & Stuit, 2014; Grabner et. al., 2007). Considering the frequent use of reading and visuospatial measures as predictors of the chess skill, (e.g., McDonald, n. d.) it was decided to include such measures in the present study as well.

Method

A quasi-experimental design was used with the chess training and control groups tested with the same tests before, during and after the chess intervention. The two teachers who were in charge of the chess intervention did not take part in the collection of data (except during the initial phase of the study), with a research assistant collecting and coding the data and a researcher (the author) doing the statistical analyses and the writing-up of the report.

Subjects and attrition. The study took place during a period of three academic years in a school in a small town in central Sweden. Each year the children entering the school formed two Grade 1 classrooms (in Sweden the children enter school at the age of seven). The school authorities followed the local regulations to make the two classrooms as equal as possible with respect to socio-economic background. When the children entered school no information was available about their levels of academic and chess skills. Therefore, it was assumed that there were no systematic differences between the two classrooms at the time of school entrance. The classroom whose students visited an after-school recreation center with chess activities (see below) was selected as the intervention classroom.

The first year of the study four classrooms were recruited (one Grade 1 and one Grade 2 pair); each of the second and third year one new pair of Grade 1 classrooms were recruited. This means that two pairs of classrooms participated during three years (one Grade 1 and one Grade 2 - at the years of entrance); three pairs during at least two years (two Grade 1 and one Grade 2 - at the year of entrance); and four pairs during at least one year (three Grade 1 and one Grade 2 - at the years of entrance). The students completed the pretests when entering the project and the post-tests at the end of each of the academic years they participated in the project.

Altogether 89 students in the experimental as well as the control classrooms took the pretests. Of these, one intervention and one control student entering the study the first academic year did not take all post-tests. The corresponding attrition rates for the second and third years were three experimental and two control students, and three experimental and three control students, respectively. These 13 students, however, took some of the post-tests; therefore they were only included in the analysis when there were complete data. These low attrition rates argue for an insignificant drop-out of students. In addition, a total of 14 students entered the classrooms after the pretests and also took some of the post-tests. These students were excluded from the data analysis. Due to absenteeism and sick-leaves all remaining students did not complete all tests; data on number of participating students are presented in the tables to follow.

For the intervention groups 10 lessons scheduled for mathematics were used for the teaching of chess, whereas the control groups received their ordinary math teaching during these lessons. This means that the classrooms participating all three years had 30 chess lessons, those participating two years 20 and those participating one year 10 chess lessons. At the end of the school day, a voluntary recreation center welcomed the children; one for the intervention and another for the control classrooms. One of the activities actively encouraged in the former recreation center was chess play. These chess play activities were initiated a few years before the project started by a staff person with a strong interest in chess. However, the chess activities dropped significantly after one project year because that staff person moved to a new job in another school.

Tests. Four different sets of tests were used:

1. *Tests of reading and visuospatial skills.*
   To test reading skill a test constructed by Järpsten (2004, 2009) was selected with one version for Grade 1 and 2 (Cronbach alpha = .80) and another for Grade 3 and 4 (Cronbach alpha = .76). For visuospatial performance the WISC symbol recognition test (test-retest reliability = .58) was used (Wechsler, 1999) throughout the project.
2. *Tests of arithmetic skills*. Three different sets of subtests were constructed:
   a. Number series. In one test the children were asked to solve forward jump series (e.g., "Continue the series 2, 4, 6 __, __, __, __, __") and in another backward jump series (e.g., "Continue the series 31, 29, 27 __, __, __, __, __"). These tests were not given to the students entering the project the first and third year.
   b. Numeral writing. Digit (Write the digits 0 to 9") and numeral writing ("Write seventeen/thirty-two, etc with digits") tests were constructed. These tests were not given to the students entering the project the first year. The digit writing test was only given to the Grade 1 students.
   c. Subtraction. Two subtraction tests were constructed: Arithmetic1 and Arithmetic2. In the first, mental calculation was to be used to find the answer. In the second, the students were allowed to calculate the answer with the aid of e.g., paper and pen or counters. In the pretest version for the Grade 1 students these tests were considered too difficult. Instead these children were tested with two multiple choice tests, one on addition and another on subtraction. In the statistical analyses these tests were treated as equivalent to the Arithmetic1 and Arithmetic2 tests, respectively.
3. *Chess*
   a. Chess test. The chess tests were constructed by one of the chess teachers in the project. The chess problems in the final post-test were grouped into three categories: Facts, Simple problems, and Advanced problems. In the Facts category the items probed the names (letter/digit pairs) of the chess board squares and how the different pieces could move (six questions in all). The category Simple problems consisted of six "Mate in one move" problems, and the category Advanced problems of three "Find a move for white that wins material" and six "Mate in two moves" problems.
   b. Chess questionnaire. Questions were constructed about spontaneous chess activities in-school (during breaks and in the recreation centers) and after-school (playing with family, with peers, on computer, membership in the local chess club and participation in chess tournaments). The questionnaire was given the last autumn of the project.

The same test of visuo-spatial-IQ was used throughout the project. The remaining tests were revised each year to be adapted to the children's current skill level. The aim was that mean level of performance should be in the 30 – 70 percent (of maximum performance) interval. Thus, changes in number of problems solved year by year give no valid information of increases or decreases in skill level. But, as identical tests were given to the experimental and control groups at each point of time, differences between these two groups inform about differences in level of performance.

To obtain reliability data for the arithmetic and the chess tests Cronbach's alphas were computed for each version and classroom pair. For the three arithmetic tests: multiple choice, mental calculation and calculation with aids, the alphas were found to vary between .69 – .91 (mdn=.86), .82 – .93 (mdn=.83), and .55 - .93 (mdn=.79), respectively; for numeral writing the alphas varied between .50 and . 92 (mdn=.63); and for counting forwards and backwards the corresponding coefficients were .69 - .74 (mdn = .71) and .70 - .89 (mdn = .83). Finally, for the chess test the alphas were found to vary from .76 to .92 (mdn = .89). Most tests seem to have an acceptable level of reliability, with somewhat low figures for the tests of calculation with aids and numeral writing. The somewhat low reliabilities for the last two tests seem to be related to the efforts to adjust the tests to the children's skill levels. The numeral writing test included only two or three items and some of the problems in the calculation with aid test were too difficult.

4. Arithmetic solution procedure interview. In individual interviews some students were given arithmetic problems to solve followed by solution procedure questions, but due to lack of resources (only about 50 % of the students were interviewed) these results will only be touched upon in the Discussion section.

Procedure
The tests and the questionnaire were administered as group tests and took about 3 lessons to complete. The students were instructed to try to solve as many tasks as possible, guessing was encouraged. Before testing started the parents were informed; all parents gave written consent for their children to participate.

Results

Validity check
First was checked whether earlier findings of the significance of the numeral writing and number word sequence skills for arithmetic performance were replicated (e.g., Johansson, 2005b). Separate analyses were run on the data collected at each of the three post-test versions. The numeral writing – subtraction test correlations were found to vary between .19 and .44 (mdn = .36), all significant at the $p<.01$ level or beyond, except the .19 coefficient which was significant at the $p<.05$ level. For the forward and backward jump tests the corresponding coefficients were .33 - .53 (mdn = .45) and .42 - .63 (mdn = .50), respectively. These coefficients were significant at the $p<.001$ level, with one exception – the .33 value that was significant at the $p<.01$ level. These results constitute a solid replication of the number sequence hypothesis -that advanced arithmetic performance grows out of the above two skills.

Intervention effects, initial analysis
As already detailed, the classrooms were partitioned into three groups: 1) Those with one year of chess training (one Grade 1 and one Grade 2 pair entering the project the first year; one Grade 1 pair entering the project the second year and one Grade 1 pair entering the project the third year), 2) Those with two years of chess training (the Grade 1 and Grade 2 pair entering the project the first year and the Grade 1 pair entering the project the second year), and 3) Those with three years of chess training (the Grade 1 and 2 pair entering the project the first year).

In Table 1 is given mean performance (in percent of max) on all tests at pretest, after one, after two, and after three years of chess teaching. The number of students taking each test is given in parenthesis after the percentage figure. In the three right hand columns are given the results of the analyses of covariance of the change from pre- to post-test with the relevant pretest data as covariates.

Starting the analysis with the chess results, it can be seen that the intervention and control classrooms differed somewhat in levels of chess skill when they entered the project. Inspecting the pretest differences within each of the four pairs showed that in two, performance in the experimental classrooms exceeded that of the control classrooms, whereas the opposite was the case in the two remaining classrooms. **t**-testing these differences revealed only one significant difference, namely that for the Grade 2 classrooms entering the project the first year, **t**(42)=2.61, p<.0.01. Probably the chess activities in the recreation center for the intervention classrooms did contribute to this finding. The tests of the effects of chess teaching were made by analyses of covariance with the relevant pretest data as covariates. Significances at the p<.05 (*), p<.01 (**) and p<.001 (***) levels are reported. Only results significant at the p<.01 level or beyond were considered unless the p<.05 results displayed a consistent pattern.

The validity of the use of the multiple choice pretests as covariates when testing the arithmetic results may be questioned. However, the results in Table 3 below, tested by analysis of variance, give the same results as those reported in Table 1.

Table 1. Mean pre- and post-test results (in percent of max) by number of intervention years for all tests in the study. In parentheses are given number of students tested.

Number of years in project and tests	Pretest				Post-test				F-statistics		
	Experim.		Control		Experim.		Control		F	df	η^2
One year of chess teaching											
Numeral writing	58	(43)	52	(45)	80	(84)	78	(85)	0.19	1/80	.00
Reading	47	(83)	48	(84)	55	(87)	65	(85)	5.93*	1/155	.04
Visuospatial	39	(86)	36	(83)	43	(84)	50	(83)	14.49**	1/151	.09
Arithmetic1	62	(89)	64	(89)	64	(86)	60	(87)	2.58	1/166	.01
Arithmetic2	33	(89)	24	(89)	69	(84)	60	(88)	3.87*	1/165	.02
Forward series	70	(20)	66	(22)	66	(84)	62	(83)	0.01	1/37	.00
Backwards series	35	(20)	32	(22)	63	(83)	57	(82)	0.19	1/35	.01
Chess	43	(88)	36	(90)	42	(84)	17	(81)	77.22***	1/158	.33
Two years of chess teaching											
Numeral writing	65	(18)	71	(21)	45	(60)	45	(62)			
Reading	56	(63)	55	(64)	74	(54)	76	(59)	0.69	1/109	.01
Visuospatial	40	(62)	36	(60)	59	(55)	60	(60)	0.16	1/111	.00
Arithmetic1	65	(65)	66	(66)	74	(60)	71	(62)	0.23	1/118	.00
Arithmetic2	27	(65)	15	(66)	56	(58)	52	(61)	0.29	1/115	.00
Forward series	70	(20)	66	(22)	58	(57)	55	(58)	0.03	1/111	.00
Backwards series	35	(20)	32	(22)	47	(57)	47	(58)	0.17	1/111	00
Chess	49	(66)	39	(66)	56	(59)	25	(61)	79.81***	1/116	.41
Three years of chess teaching											
Numeral writing	-	-	-	-	34	(40)	21	(33)	0.00	1/32	.00
Reading	64	(45)	67	(40)	79	(39)	80	(34)	0.08	1/70	.00
Visuospatial	39	(42)	32	(38)	62	(39)	67	(33)	5.03*	1/69	.07
Arithmetic1	68	(45)	65	(43)	75	(40)	77	(35)	1.21	1/72	.02
Arithmetic2	32	(45)	14	(43)	58	(40)	56	(35)	0.03	1/72	.00
Forward series	-	-	-	-	64	(40)	65	(33)			
Backwards series	-	-	-	-	47	(40)	37	(33)			
Chess	60	(45)	36	(43)	56	(38)	20	(33)	109.02***	1/68	.62

The chess results in Table 1 reveal that the intervention classrooms outperformed the control classrooms at all three post-test points of time; moreover, the intervention vs. control difference got stronger for each year of chess teaching. Having found a substantial increase in chess skill in the intervention classrooms, the next question is whether this effect transferred to arithmetic, reading and visuospatial performance.

This was studied in a new series of analyses of covariance of the remaining test results with the classrooms partitioned into the same three groups as described above and with the pretest data as covariates. Also the results of these analyses are given in the three right hand columns in Table 1. As can be seen only four significant effects were found, of which three are on the $p<.05$ level. The visuospatial variable was significant both after one and after three years of chess teaching. However, it should be observed that in both cases the result for the control group exceeded that of the experimental group. The lack of consistent results for reading and visuospatial IQ agree with those reported by Nicotera and Stuit (2014), whereas the negative results for arithmetic are surprising and contradict results usually reported by other researchers (see Nicotera & Stuit, 2014).

Psychometric analyses
It must be underlined that this first series of analyses tests whether mean performance in the chess classrooms exceeded that of the control classrooms. Next, a series of psychometric analysis were made of possible relationships between chess performance and performance on the arithmetic, reading and visuospatial tests. If such relationships exist, significant correlations should be found; moreover, if the strength of the relationship is related to amount of chess teaching, these correlations should increase with number of years of chess teaching in the intervention but not in the control classrooms. To test these speculations, the relevant correlation data were computed; the results are given in Table 2.

Table 2. Correlations (Pearson) between chess score and performance on the arithmetic, reading, and visuospatial tests by condition and number of years in the project. Number of subjects in parentheses.

	Time of test							
Condition and tests	Pretest		Post-test 1, after one year of teaching		Post-test 2, after two years of teaching		Post-test 3, after three years of teaching	
Experimental classrooms								
Reading	.21	(82)	.23*	(82)	-.01	(51)	.15	(36)
Visuospatial	.06	(84)	.24*	(79)	.29*	(52)	.22	(36)
Arithmetic1	.30**	(87)	.46***	(81)	.32*	(56)	.58***	(37)
Arithmetic2	.00	(87)	.23*	(79)	.31*	(54)	.62***	(37)
Numeral writing	.33*	(41)	.31**	(79)	.18	(56)	.48**	(37
Forward series	-.17	(21)	.42***	(79)	.27*	(54)	.45**	(37)
Backwards series	.40	(21)	.42***	(78)	.38**	(54)	.52***	(37)
Control classrooms								
Reading	.14	(82)	.40***	(76)	.37**	(57)	.16	(30)
Visuospatial	.14	(81)	.43***	(75)	.02	(57)	.08	(30)
Arithmetic1	.26*	(87)	.34**	(79)	.25*	(59)	.22	(33)
Arithmetic2	.13	(87)	.42***	(80)	.15	(58)	.21	(33)
Numeral writing	.57***	(43)	.36**	(76)	-.12	(60)	.02	(29)
Forward series	.10	(20)	.33**	(75)	.15	(57)	.16	(29)
Backwards series	.62***	(20)	.19	(74)	.22	(57)	.17	(29)

First, the reading and visuospatial tests data were scrutinized. As can be seen the correlations with chess in the experimental classrooms did not reach the $p<.01$ significance level at any point in time, whereas significant correlations were obtained in the control classrooms at post-test 1 (also after two years for reading). But these correlations dropped to insignificance at post-test 3. Thus, no consistent or lasting correlations between chess on the one hand and reading and visuospatial performance on the other could be observed. These results are in line with findings reported by Grabner, et. al., (2007) and indicate that both factors, at least as measured here, may be unrelated to the chess skill.

The results for the arithmetic measures paint a quite different picture, with substantially different trajectories in the experimental compared to the control classrooms. In the former, the correlations between chess and the arithmetic measures increased in strength with number of years of chess teaching. In the control classrooms the same correlations increased from pretest to post-test 1, and then dropped to insignificance.

It may be observed that the correlations decreased from the first to the second post-test in the intervention classrooms. Why this happened is difficult to explain by the data at hand. A possible contributing factor may have been that the staff person in the intervention recreation center took a new job during the second year of the project; as mentioned earlier this may have resulted in a decrease in the chess activities in that center.

That the correlations between chess scores and arithmetic performance were very strong at the final posttest indicate that a link between the skills of chess and arithmetic may have been formed for at least some students, most probably the students with the highest scores on the chess test. This interpretation can be tested directly by partitioning the students in each classroom into two groups: one composed of the students with low and medium scores on the final chess test and another with the students with high scores on the final chess test. This was accomplished by labelling the six students with the highest scores as a “high-score” group and the remaining students as a “low-score” group. The result was four groups of students: Control, high-score; Control, low-score; Experimental, high-score and Experimental, low-score.

Intervention effects, continued analyses

Partitioning the students into these four groups allows the study of whether the Experimental, high-score students outperform the Experimental, low-score and the Control, high-score students on the reading, visuospatial and arithmetic tests. An answer in the affirmative would indicate that chess teaching benefits some, but not all, students in the domains studied. Second, do the Control, low-score students outperform the Experimental, low-score students? Such a finding would indicate that replacing mathematics with the teaching of chess may have a detrimental effect for low performing chess students. The data, tested by analyses of variance, are presented in Table 3. The three right hand columns give the results of the statistical analyses.

Table 3. Percent correct (lsmeans, with number of subjects in parentheses) as a function of year in project, teaching condition, and high or low score on the final chess test.

Years in project and tests	Condition and chess score group: Intervention Low/Medium	Intervention High	Control Low/Medium	Control High	F-statistics LH	IC	LH*IC
Pre-test							
Reading	43 (62)	59 (20)	44 (62)	58 (21)	ns	ns	ns
Visuospatial IQ	37 (61)	43 (24)	36 (59)	36 (23)	ns	ns	ns
Numeral writing	-	-	-	-	-	-	-
Forward series	72 (16)	70 (5)	71 (17)	50 (4)	ns	ns	ns
Backwards series	25 (16)	80 (5)	35 (17)	50 (4)	ns	ns	ns
Arithmetic1	58 (64)	73 (24)	60 (66)	76 (22)	*	ns	ns
Arithmetic2	28 (64)	42 (24)	22 (66)	33 (22)	*	ns	ns
Chess	40 (63)	52 (24)	28 (65)	56 (24)	*	ns	ns
One year in project							
Reading	53 (63)	63 (24)	62 (62)	73 (23)	*	*	ns
Visuospatial IQ	42 (61)	44 (23)	50 (60)	53 (23)	**	ns	ns
Numeral writing	76 (60)	92 (24)	74 (63)	88 (22)	***	ns	ns
Forward series	58 (60)	87 (24)	59 (61)	71 (22)	***	ns	ns
Backwards series	55 (59)	83 (24)	54 (60)	64 (22)	**	ns	ns
Arithmetic1	60 (63)	77 (23)	58 (63)	66 (24)	**	ns	ns
Arithmetic2	67 (61)	73 (23)	56 (64)	71 (24)	**	ns	ns
Chess	35 (60)	59 (24)	13 (58)	26 (23)	***	***	ns
Two years in project							
Reading	74 (38)	75 (16)	75 (42)	80 (18)	ns	ns	ns
Visuospatial IQ	58 (41)	65 (14)	59 (44)	62 (17)	ns	ns	ns
Numeral writing	40 (43)	55 (17)	44 (45)	51 (18)	ns	ns	ns
Forward series	53 (41)	71 (16)	57 (41)	54 (18)	*	ns	ns
Backwards series	40 (41)	67 (16)	57 (41)	54 (18)	**	ns	*
Arithmetic1	70 (43)	86 (17)	71 (45)	74 (18)	ns	ns	*
Arithmetic2	50 (41)	70 (17)	51 (44)	57 (18)	**	ns	ns
Chess	51 (42)	70 (17)	23 (44)	31 (18)	***	***	ns
Three years in project							
Reading	79 (27)	76 (12)	78 (22)	83 (12)	ns	ns	ns
Visuospatial IQ	62 (27)	63 (12)	67 (21)	68 (12)	ns	ns	ns
Numeral writing	23 (28)	58 (12)	17 (21)	29 (12)	**	ns	ns
Forward series	59 (28)	76 (12)	63 (21)	69 (12)	*	ns	ns
Backwards series	41 (28)	60 (12)	35 (21)	40 (12)	*	ns	ns
Arithmetic1	70 (28)	85 (12)	75 (23)	81 (12)	*	ns	ns
Arithmetic2	53 (28)	70 (12)	54 (23)	61 (12)	ns	ns	ns
Chess	47 (26)	75 (12)	15 (21)	29 (12)	***	***	**

LH=The group variable, or low- versus the high-scoring students; IC= The teaching variable, or the intervention versus the control classrooms; LH*IC= the interaction between the group and the teaching variables.

Starting with the chess results; at pretest the teaching variable was not significant, in line with findings already presented. With the exception of the Grade 2 classroom entering the first year, this again shows that the initial differences between the experimental and control classrooms were marginal. At all post-tests, however, both the group and the teaching variables were significant. At the final post-test, also the group by teaching variable reached significance, indicating that the high-score intervention students outperformed their low-performing peers to a higher degree than the high-score control students did. Indeed, the increase in chess scores in the intervention classrooms seems to be concentrated to the high-score students.

This, however, does not mean that the high-score students also outperformed their peers as to reading and visuospatial performance. Table 3 clearly demonstrates that no consistent such results were found. Therefore, it must be again concluded that it has not been possible do demonstrate a relation between chess skill on the one hand and reading or visuospatial performance on the other.

The five arithmetic measures present a quite different picture. The difference between the high- and low-score groups did not reach the $p<.01$ level at pretest but did so at the post-tests, with a few exceptions. The differences between the experimental and the control classrooms were insignificant, which repeats the results already presented in Table 1 above. Surprisingly, no group by teaching interactions were significant at the $p<.01$ level. These findings mean that the students in the high-score groups outperformed their low-score peers, irrespective of whether they belonged to the intervention or the control group. The insignificant interactions also imply that it cannot be argued that the low-score control students performed at a higher level than the low-score experimental students. In other words, there was no evidence of detrimental effects of the substitution of 10 math lessons each year with chess lessons.

The results for the two arithmetic tests were inconsistent in that they were not significant at all post-tests. Therefore, it was decided to analyze these results further. In this follow-up only the data from the final post-test were included. The results for the two arithmetic tests were first collapsed, and then subdivided into two categories based on level of difficulty of the individual subtraction problems. Into an "easy" group were brought no-trade subtraction problems, such as "9-4", 17- 3", "38-2", etc, whereas problems requiring trades, for example "18 – 9", "21 – 19", and "82 – 23", or problems with large minuends and subtrahends, e.g., "58-43" were brought to the "difficult" group. Finally, the performance differences in each classroom between the high- and the low-score chess groups were t-tested, see Table 4.

Table 4. Percent of easy and difficult arithmetic problems solved by the low- and high-score groups at the final posttest. Number of students in parentheses.

Type of problem and grade	Intervention		**t**-statistics			Control		**t**-statistics		
	Low/Medium	High	**t**	df	p	Low/Medium	High	**t**	df	p
Easy problems solved										
Grade 1	76 (17)	83 (6)	ns	-	-	63 (18)	86 (7)	3.96	23	.001
Grade 2	87 (13)	100 (5)	3.01	21	.01	82 (17)	88 (6)	ns	-	-
Grade 3	85 (17)	95 (6)	ns	-	-	86 (13)	93 (6)	ns	-	-
Grade 4	93 (16)	91 (6)	ns	-	-	93 (11)	92 (6)	ns	-	-
Difficult problems solved										
Grade 1	33 (17)	53 (6)	ns	-	-	30 (18)	64 (7)	2.45	23	.05
Grade 2	63 (13)	84 (5)	ns	-	-	54 (17)	57 (6)	ns	-	-
Grade 3	44 (17)	73 (6)	3.07	21	.01	54 (13)	62 (6)	ns	-	-
Grade 4	61(16)	81 (6)	2.28	20	.05	58 (11)	69 (6)	ns	-	-

The analysis of main interest was whether the high-score students solved a greater number of difficult problems than the low/medium-score students. As can be seen from the table, this was the case in the Grade 3 and 4 experimental classrooms, as well as for the control high-score students in Grade 1. Thus, it seems that students with high scores on the chess test were exceptionally good at solving difficult subtraction problems. It should perhaps be pointed out that the Grade 3 and 4 classrooms in the above table entered the project as Grade 1 and 2 classrooms, respectively, the first year.

Analysis of the students' chess skill

Evidently level of chess skill is strongly related to performance on the arithmetic tests, which raises the question of how to characterize the chess skill acquired by the high-score-students. To obtain a preliminary answer to that question, performance on the final chess test was analyzed. The test had a total of 21 items, which were partitioned into the three categories already described in the Methods section.

The extent to which the students were able to solve the 9 most difficult items was focused, because these items can be seen as measures of the student's ability to analyze the positions as meaningful entities. The relevant data are presented in Table 5. When reading the table, it should be remembered that it presents only the final post-test chess data. As the results for the two classrooms entering the project the first year (as Grade 1 and Grade 2) differed substantially, the results for these two were presented separately.

Table 5. Percent of chess problems solved by teaching condition, years in project and low or high chess score.

	Years in project (grade at entrance), and chess score group							
Condition and chess problem category	One year in project (Grade 1)		Two years in project (Grade 1)		Three years in project (Grade 1)		Three years in project (Grade 2)	
	Low	High	Low	High	Low	High	Low	High
Experimental classrooms								
1 Facts	71	92	94	97	89	100	89	100
2 Simple	4	50	50	97	60	86	52	97
3 Advanced	0	3	12	31	18	31	17	61
No. of students	*17*	*6*	*14*	*6*	*16*	*6*	*14*	*6*
Control classrooms								
1 Facts	24	64	54	86	50	61	53	77
2 Simple	0	0	3	22	3	25	0	33
3 Advanced	0	0	0	0	0	6	0	0
No. of students	*18*	*6*	*17*	*6*	*12*	*6*	*11*	*6*

Because of the low number of students in each cell in the table, no statistical tests were run. However, the results clearly demonstrate that the control students solved almost none of the advanced chess problems. In contrast, many such problems were solved in the experimental classrooms, in particular by the high-score students in the classroom entering the project as a Grade 2. Thus, one aspect of the good arithmetic results in the high-score groups seems to be an outstanding ability to solve "Mate in two moves" problems.

It should be observed that the high- score Grade 2 students scored substantially higher than the other experimental, high-score groups. In particular, the results on the Category 3 items indicate that these students may have reached a critical threshold in the sense that they could solve fairly difficult chess problems.

In the control classrooms, the high-score students solved a substantially greater number of Simple problems than the low-score students. Is it this difference in chess skill that explains why these students solved a greater number of arithmetic problems than their low-score peers? If this is the case, a first threshold in the chess skill seems to be reached when the students develop the capacity to solve "Mate in one move" problems. It must be underlined that this chess skill evidently developed without any formal chess teaching.

In the experimental, low-score groups an increase in performance is found only from Grade 1 to Grade 2, but not to Grades 3 and 4 despite continued chess teaching. The developmental trajectories for both the control, high and the control, low students demonstrate a parallel trend, that is, an improvement from Grade 1 to Grade 2, but then a standstill. Is it this slowdown that explains why the correlations in the control classrooms between chess and the three domains increased from the first to the second post-test, but then decreased?

The slowdown in chess development for the low-score intervention groups seems puzzling, bearing in mind that they received the same amount of chess teaching as the intervention high-score groups. Perhaps a decreasing motivation for chess has contributed to the results? Unfortunately, no direct measures of motivation for chess and mathematics were collected, but indirect such information can be found in the answers to the questionnaire about chess activities given to Grade 2, 3 and 4 in the autumn the last year of the project. This questionnaire had items on chess activities in-school and after-school, and also questions on chess club membership. The questions and the answers are given in Table 6.

Table 6. Answers to the questionnaire on chess activities given in autumn, the final year of the project. The results are given as absolute numbers.

	Years in project, condition and high versus low chess score group (grade when entering the project)											
	Two years in project (Grade 1)				Three years in project (Grade 1)				Three years in project (Grade 2)			
Playing chess. . .	EL[a]	EH[a]	CL[a]	CH[a]	EL	EH	CL	CH	EL	EH	CL	CH
during breaks between lessons												
Never	14	6	13	5	14	6	14	6	12	2	11	6
Sometimes	0	0	0	0	1	0	0	0	3	0	0	0
Often	0	0	0	1	0	0	0	0	0	2	0	0
Always	0	0	0	0	0	0	0	0	0	2	0	0
in recreation center												
Never	4	1	6	1	7	1	6	2	14	2	11	6
Sometimes	7	3	6	4	3	1	2	1	0	1	0	0
Often	2	1	0	0	0	0	0	0	1	2	0	0
Always	1	1	1	1	1	0	0	0	0	1	0	0
with family members/peers												
Never	9	1	8	2	7	5	11	4	10	1	10	5
Sometimes	5	3	5	3	3	1	3	2	5	3	1	1
Often	0	2	0	0	0	0	0	0	0	2	0	0
Always	0	0	0	1	1	0	0	0	0	0	0	0
on computer												
Never	3	1	6	1	8	3	9	4	12	1	10	4
Sometimes	10	2	3	3	4	2	3	1	1	2	1	2
Often	1	3	2	1	2	0	1	1	2	1	0	0
Always	0	0	2	1	1	0	1	0	0	3	0	0
Membership in chess club												
Yes	1	1	0	0	1	0	0	0	0	5	0	0
No	13	5	13	6	14	6	13	6	15	1	14	6
Playing in chess club												
Never	13	5	13	6	14	6	13	6	14	1	14	6
Sometimes	0	1	0	0	0	0	0	0	1	2	0	0
Often	0	0	0	0	1	0	0	0	0	1	0	0
Always	1	0	0	0	0	0	0	0	0	3	0	0
Playing in chess tournaments												
Yes	1	1	0	0	1	0	0	0	0	3	0	0
No	13	5	13	6	14	6	17	6	6	3	14	6

[a]EL = experimental, low score; [a]EH = experimental, high score; [a]CL = Control, low score; [a]CH = control, high score. [a] Some students did not answer all questions or did not participate in the recreational center activities.

The most distinctive finding in the above table is the response profile for the experimental, high score students, who entered the project as a Grade 2. These students distinguished themselves with respect to a high rate of playing chess during breaks between lessons, when visiting the recreation centers, in the family, with peers, and on computer. In addition many were members of the local chess club, often played chess in the club and participated in tournaments. In short, in comparison with the other students, they "bathed" in chess. Their low-score peers were much more inactive, which is taken as an indirect indication of a split between the high- and low-score students in this classroom. This interpretation is supported by informal observations by the chess teachers. They reported that the majority of the students in this classroom voted “No” to participate in a national chess tournament for Grade 4 classrooms despite good chances of winning the tournament had they participated.

Looking at the other classrooms answering the questionnaire, no distinct differences were observed between the high- and low-score groups. It may be speculated that the split between the high- and low-score groups did not come about until the high-score students entered the local chess-club. Of course, these speculations require socio-metric and motivation measurements to be corroborated.

Discussion

Both Gobet and Campitelli (2006) as well as Nicotera and Stuit (2014) preferred a randomized over a quasi-experimental design. In an analysis of designs for studies of teaching Shadish and Cook (2009) observed that randomization may create more problems than it solves. They mentioned that participants may be unwilling to accept random assignment and differential attrition rates may disrupt the results. In the present case, a randomized design was not an option; the formation of the classrooms had to follow the local regulations. Only two classrooms were started each year, which excludes the alternative with one experimental and two control classrooms. The initial **t**-tests made of the differences between the intervention and control classrooms support the conclusion that the two Grade 1 classrooms entering the project each year were approximately equivalent.

It should be underlined that the number of students in some of the analyses is small; in particular when the classrooms are partitioned into high and low chess score groups. The results of these analyses should be taken as indicatory only.

Of main importance is the fact that all students in the classrooms participated throughout the project and that the attrition rate was small. This distinguishes the present study from many other chess studies which are beset by different types of selection biases (e.g., Nicotera and Stuit, 2014). In effect, the low attrition rate in the present study may explain the twin findings of no transfer between chess and arithmetic on the classroom level, but significant effects when the students were partitioned into high- and low-score groups. The low attrition rate may also be one of the reasons for the high correlation coefficients between chess and arithmetic. A drop-out of students with low chess scores would undoubtedly have decreased these coefficients.

The reliability data indicate that the quality of the measurements was acceptable. However, the ambition to adapt the tests to the students’ current skill level meant substantial changes in some cases. For example, the absence of measurements of the numeral writing and number series skills during the first year of the project, and the change from recognition to production arithmetic tests may be seen as threats to validity. Finally, a complete analysis of the results would have been facilitated by tests of motivation for chess and mathematics and by socio-metric tests throughout the project period. Despite these shortcomings, however, the results display a high level of consistency, which allows a discussion of the findings obtained.

As to the substantial results, no significant differences between the experimental and control classrooms were found with respect to the reading and visuospatial tests. These results recurred in the correlation analyses and in the low- versus high-chess score student comparisons. Usually significant correlations are found between chess performance and results on visuospatial test for young (e.g., Frydman & Lynn, 1992) but not for old subjects.

The present results argue for a reconsideration of how to define and measure the visuospatial skill. The result for reading is in agreement with findings in other studies; that is, small or insignificant correlations between chess and reading scores (Doll & Mayr, 1987; Grabner et. al., 2007).

The results of central interest in the present study are the strong correlations between the chess scores and the results on the number series and arithmetic tests, together with the finding that the high-score students outperformed their low-score peers. Strong correlations between chess scores and the results on tests of arithmetic have also been reported by other researchers. These results have usually not attracted much attention, perhaps because no convincing explanation of these findings has been offered. The suggestion by Doll and Mayr (1987) that there is a transfer from defining the squares on the chessboard by letter-digit coordinates to mathematics has, so far, not been subjected to empirical tests.

An alternative hypothesis, advanced in the introductory section of this report, says that transfer may take place between chess and arithmetic when both skills have reached given threshold levels allowing instant recognition of meaningful relationships. In the case of arithmetic a distinction can be made between the prerequisite mental number sequence skill and the ability to use it when solving arithmetic problems. In students, aged five to six years, the former capacity can be measured by tasks such as solving backwards counting problems ("Count from thirteen to eight") and doubles problems ("What is two plus two?; three plus three=?"; etc) without taking help of finger counting (see Johansson, 2013 for details). For children having started school, the ability to solve, in particular, backward jump problems ("Continue the series 35, 32, 29, __, __, __) measures the same capacity. In addition to the number word sequence skill, earlier studies (Johansson, 2005b) have shown that also the skill of numeral writing is an important prerequisite, which was the case in the present study too.

In the case of chess, no corresponding distinction between prerequisite capacities and the access of advanced solution procedures has been identified. But the results obtained point to two possible threshold. First, the absence, see Table 3, of significant interactions between the high/low and the experiment/control variables indicate that in both the experimental and the control classrooms the high-score students were superior to their low-score peers to about the same degree. The results in Table 5 demonstrate that the control high-score students differed from their low-score peers in their ability to solve simple problems such as "Mate in one move". Therefore, already this initial capacity of perceiving meaningful chess relationship may constitute a first basis for the formation of a link between chess and arithmetic.

The results for Advanced problems, do not give clear evidence for a second threshold. The high-score experimental students, who entered the project as a Grade 2 classroom, were outstanding in their ability to solve this type of chess problems (see Table 5). But they performed at about the same level on the five arithmetic measures as the high-score experimental students, who entered the project as a Grade 1 classroom, despite the fact that the letter group of students scored substantially lower on the Advanced problems.

An analysis of the interview results for the Grade 2 classroom, however, indicates the presence of a second threshold. Comparing the results for the Experimental, high score students with the other three Grade 2 groups (Experimental, low score, Control, high score, and Control low score), it was found that the former group often reported the jump strategy (on 33 % of the problems), but seldom the counting strategy (on 17 % of the problems). For the other three groups the mean frequencies of these strategies were 24 % and 28 %

respectively. The former group solved 84 % of the subtraction problems correctly, compared to a mean of 68 % for the other three groups. However, when considering these results, it should be remembered the number of students is small. These results, although based on only a few subjects, suggest that children having reached the second threshold are able to view and solve both arithmetic and chess problems as meaningful entities.

Given that the above results can be replicated in a study with a greater number of students, it may be concluded that a first goal for chess teaching in school might be to stimulate the students to develop their chess skills to a Simple “Mate in one move” level.

It may be asked why some students in the control classrooms developed a capacity to solve some of the Simple problems. A near-at-hand answer is the presence of leaks from the intervention to the control classrooms about the chess project. The project was well known in the school, for example by the letters of permission sent to the control children’s parents. The findings of an increase in chess activities from Grade 1 to Grade 2 that then levelled off indicate that the game of chess may have attracted an initial interest in the control classrooms.

Continuing with the chess questionnaire results in Table 6, it may be observed that in Grade 4 there was a great difference between the high- and low-score groups. The low-score students in this classroom reported no in-school chess activities and very few after-school ones, which is in glaring contrast to the reports by the high-score students. It seems that a split did occur between the high- and low-score students. This interpretation is supported by what has already been reported about this classroom voting “No” to the participation in a national chess tournament.

The differences between the high- and low-score chess groups in Grade 2 and 3, see Table 6, were much smaller, indicating that the split may have taken place during the third or the fourth grade. Perhaps the recruitment to the chess club may have contributed in bringing forth the split? Whatever the case, it seems necessary to invest substantial effort into studying whether splits occur, and if that is the case, what are the causes and how to counteract them.

As to the causal relationship between the skills of chess and arithmetic; it may be the case that the chess skill transfers to arithmetic, that the skill of arithmetic transfers to chess or that the relationship is best characterized as iterative. An argument in favor of the first alternative is the finding that students with high chess skills also improve their arithmetic skills to a greater extent than students with low chess skills and the finding that the high-score students in the classroom entering the project as a Grade 2 often reported the jump strategy, but seldom the counting strategy. If the main causal relation is from chess to arithmetic, it seems of outmost importance to maintain the motivation for chess in all students as long as the whole-class teaching of chess goes on. The chess teachers have suggested that whole-class chess teaching should take place only during the two first school years. Then it may continue in voluntary forms.

If the skill of chess is dependent on the teaching of arithmetic, the results obtained point to the importance of number word sequence and numeral writing practices together with the teaching of the jump method solution procedure. However, such a change in teaching methods may be difficult to accomplish considering the strong impact of the constructivist theory on present day math teaching. Present-day math teaching, at least in Sweden, focus on using concrete manipulatives which supports the formation of a unit conception of number which, in its turn, forms the basis for counting solution procedures. This teaching tradition obstructs the development of understanding arithmetic problems meaningfully, and hence, the formation of a connection between chess and arithmetic.

In conclusion, the results suggest that the learning of chess has a positive impact on the learning of arithmetic. In addition, the results offer ideas on how the practice of teaching chess can be improved and how evaluation studies can be designed for a comprehensive measurement of the effects of teaching.

References

Chabris, C. F., & Hearst, E. S. (2003). Visualization, pattern recognition, and forward search: effects of playing speed and sight of the position on grandmaster chess errors. *Cognitive Science, 27*, 637 – 648.

Charness, N., Reingold, E.M., Pomplun, M. and Stampe, D.M. (2001). The perceptual aspect of skilled performance in chess: Evidence from eye movements. *Memory & Cognition, 29,* 1146 – 1152.

Christiaen, J. (1976). *Chess and cognitive development* (Epstein, S., Trans). Unpublished master's thesis, Gent, Belgium. (In Nicotera and Stuit, 2014.)

Chase, W. G., & Simon, H. A. (1973a). Perception in chess. *Cognitive Psychology, 4*, 55 – 81.

Chase, W. G., & Simon, H. A. (1973b). The mind's eye in chess. In W. G. Chase (Ed.), *Visual information processing* (pp 215 – 281). New York: Academic Press.

De Groot, A. D. (1965). *Thought and choice in chess.* The Hauge: Mouton.

Doll, J., & Mayr, U. (1987). Intelligenz und Schachleistung—eine Untersuchung an Schachexperten [Intelligence and performance in chess—A study of chess experts]. *Psychologische Beiträge, 29,* 270-289. (In Grabner, et. al., 2007)

Gobet, F., & Campitelli, G. (2006). Education and chess: A critical review. In T. Redman (Ed.), *Chess and education: Selected essays from the Koltanowski conference* (pp. 124-143). Dallas: Chess Program at the University of Texas at Dallas.

Grabner, R. H., Stern, E., and Neubauer, A. C. (2007). Individual differences in chess expertise: A psychometric investigation. *Acta Psychologica, 124*, 398 – 420.

Frank, A., & D'Hondt, W. (1979). Aptitudes et apprentissage du jeu d'e'checs au Zaire (aptitudes and learning of the game of chess in Zaire). *Psychopathologie Africaine, 15*, 81–98. (InGrabner, et al, 2007)

Frydman, M., & Lynn, R. (1992). The general intelligence and spatial abilities of gifted young Belgian chess players. *British Journal of Psychology, 83*, 233-235.

Holding, D. H. (1992). Theories of chess skill. *Psychological Research, 54*, 10 – 16.

Horgan, R. W., & Morgan, D. (1990). Chess expertise in young children. *Applied Cognitive Psychology, 4,* 109 – 128.

Johansson, B. S. (2005a). Number-word sequence skill and arithmetic performance. *Scandinavian Journal of Psychology, 46*, 157 – 167.

Johansson, B. S. (2005b). Numeral writing skill and elementary arithmetic menal calculations. *Scandinavian Journal of Educational Research, 49*, 3 – 25.

Johansson, B. S. (2013). *Matematik i förskola och förskoleklass* [Mathematics in preschool and preschool class. Uppsala: Kunskapsföretaget AB.

Järpsten, B. (2004). *DLS bas: för skolår 1 och 2. Handledning*. Stockholm: Psykologiförlaget.

Järpsten, B. (2009). *DLS: handledning: för klasserna 2 och 3*. Hägersten: Psykologiförlaget.

Linhares, A. (2005). An active symbols theory of chess intuition. *Minds and Machines, 15,* 131 – 181.

McDonald, P. S. (n.d.). *The benefits of chess in education: A collection of studies and papers on chess and education*. Unpublished manuscript. Retrieved from http://www.psmcd.net/otherfiles/BenefitsOfChessInEdScreen2.pdf.

Nicotera, A., & Stuit, D. (2014). *Literature review of Chess Studies*. Basis Policy Research. Saint Louis: Chess Club and Scholastic Center of Saint Louis.

Saariluoma, P., & Kalakoski, V. (1998). Skilled imagery and long-term working memory. *American Journal of Psychology, 110*, 177-201.

Scholz, M., Niesch, H., Steffen, O., Ernst, B., Loeffler, M., Witruk, E., & Schwarz, H. (2008). Inpact of chess training on mathematical performance and concentration ability of children with learning disabilities. *International Journal of Special Education, 23* (3), 138-148. (InGrabner, et al, 2007).

Shadish, W. R. & Cook, T. D. (2009). The renaissance of field experimentation in evaluating interventions. *Annual Review of Psychology, 60,* 607 – 629.

Unterrainer, J. M., Kaller, C. P., Halsband, U., and Rahm, B. (2006). Planning abilities and chess: a comparison of chess and non-chess players on the tower of London task. *British Journal of Psychology, 97*, 299 – 311.

Wechsler, D. (1999*). Wechsler Intelligence Scale for Children: WISC III: manual*. Stockholm: Psychologiförlaget.