basis policy research

Literature Review of Chess Studies

November, 2014

Anna Nicotera

David Stuit

Acknowledgements

This literature review was commissioned by the Chess Club and Scholastic Center of Saint Louis (CCSCSL). We would like to acknowledge the constructive feedback and support provided by Tony Rich, Joy Bray, Bill Thompson, and Matt Barrett of CCSCSL.

This study benefited from input at various stages from Michael Podgursky of the University of Missouri-Columbia.

At Basis Policy Research, Sy Doan, Claire Graves, Heather Price, and Lauren Shaw helped to search for chess studies for inclusion in the review. Additionally, Claire Graves and Lauren Shaw provided research assistance and careful technical review.

We would like to thank the researchers who provided supplementary information we needed to incorporate their papers into this literature review. Joseph DuCette, Sigrun-Heide Filipp, Anthony Glendinning, Jim Liptrap, Stuart Margulies, Christine Palm, Murray Thompson, Roberto Trinchero, and Anna van Zyl all provided key data we needed for our analyses. Malcolm Pein helped connect us with difficult to reach researchers.

Finally, we are grateful to Stefan Löffler, John Foley, and three anonymous reviewers for their thoughtful reviews.

Table of Contents

Executive Summary

Many students in the United States participate in after-school chess clubs. While students join chess clubs for competitive play, there is a growing trend to develop and implement scholastic chess curriculum that targets students' academic outcomes through after-school and in-school initiatives. Scholastic chess instruction uses chess as a springboard to work on cognitive and academic skills that are critical to student performance, such as logical and spatial thinking, reasoning, long-term planning, assessment, decision-making, memory, judgment, and strategizing. The research base that explores whether chess programs impact student cognitive, academic, and behavioral outcomes is growing. The over-arching goal of this literature review is to identify the degree to which existing empirical evidence supports the theory that participation in chess programs, whether designed as in-school or after-school programs, will lead to improved academic, cognitive, and/or behavioral outcomes for school-aged children.

This literature review identified 51 studies of chess. Twenty-four of the 51 studies met a set of pre-determined criteria for eligibility and were included in analyses. Results from the literature review were categorized by the quality of the study design and organized by whether the studies examined after-school or in-school chess programs. The main findings from this literature review are:

1. After-school chess programs had a positive and statistically significant impact on student mathematics outcomes.
2. In-school chess interventions had a positive and statistically significant impact on student mathematics and cognitive outcomes.

While the two primary outcomes listed above are based on studies that used rigorous research design methodologies, the results should be interpreted cautiously given the small number of eligible studies that the pooled results encompass (two high-quality after-school studies and six high-quality in-school studies).

The after-school chess studies examined competitive chess clubs and provided very little detail about how the programs were implemented. On the other hand, the in-school chess studies examined scholastic chess programs and provided some details about the programmatic components. Taken as a whole, the positive mathematics and cognitive outcome results from in-school chess studies may be explained by the chess programs being incorporated into students' weekly academic schedules, instruction during the school day leading to higher attendance rates and lower attrition, administering the program for an extended period of time, and connecting the intervention with math instruction and curriculum.

Introduction

Many students in the United States participate in after-school chess clubs. While students join chess clubs for competitive play, there is a growing trend to develop and implement scholastic chess curriculum that targets students' academic outcomes through after-school and in-school initiatives. Scholastic chess instruction uses chess as a springboard to work on cognitive and academic skills that are critical to student performance, such as logical and spatial thinking, reasoning, long-term planning, assessment, decision-making, memory, judgment, and strategizing. The research base that explores whether chess programs impact student cognitive, academic, and behavioral outcomes is growing. The over-arching goal of this literature review is to identify the degree to which existing empirical evidence supports the theory that participation in chess programs, whether designed as in-school or after-school programs, will lead to improved academic, cognitive, and/or behavioral outcomes for school-aged children.

There have been a number of research reviews of chess studies (see Appendix B for a list of previous reviews). This literature review differs from the previous studies by using rigorous search, coding, and analytic strategies. Specifically, this literature review follows the protocols and quality standards used by the U.S. Department of Education's Institute of Education Sciences' (IES) to identify studies for inclusion in the What Works Clearinghouse (WWC).[1] It also relies on the methods recommended by the Campbell Collaboration for systematic reviews of education research.[2] Standardized effect sizes are generated for each of the eligible studies so that the magnitude and statistical significance of results can be compared across studies.

This literature review is organized in the following manner. First, the literature review describes study inclusion criteria, search protocols, coding processes, classification system for study designs, and procedures for calculating standardized effect sizes. Second, the literature review presents results from eligible chess studies, organized by the type of chess intervention: after-school and in-school programs. Third, the results of the literature review are discussed. Finally, recommendations for future research are provided.

[1] http://ies.ed.gov/ncee/wwc/default.aspx
[2] http://www.campbellcollaboration.org/resources/training/The_Introductory_Methods.php

Literature review methodology

Study inclusion criteria

For the purposes of this literature review, studies had to meet a set of inclusion criteria. First, the study must have examined the impact of a defined intervention that incorporated chess as a major feature. The chess intervention could take place during the school day or after school. To expand the pool of potential studies, studies of interventions that used other spatial or strategy games could be eligible if they met all of the other inclusion criteria.

Second, the study must have examined the impact of the chess intervention on academic, cognitive, non-cognitive, and/or behavioral outcome measures. Studies were deemed ineligible if the only outcome measures for the intervention were chess skills and/or chess rankings.

Table 1. Inclusion criteria for eligible chess studies

Study Element	Inclusion Criteria
Intervention	Primary analysis designed to estimate the effects of an intervention that incorporates chess or game(s) similar to chess as a major feature
Outcome	Must use student-level outcome measures (academic, cognitive, non-cognitive, or behavioral) with evidence of validity and reliability. Examples of assessments or indicators include: standardized test scores, end-of-course grades, high school graduation, intelligence, memory, concentration, problem-solving, attention span, spatial reasoning skills, self-confidence, self-efficacy, self-esteem, critical thinking, creative thinking, grit, persistence, school-day attendance, study habits (planning), attitudes toward school
Design	Study must be designed to compare participants in the chess intervention with a comparison group of non-participants. Examples of eligible study designs include: experimental (e.g., random assignment) and quasi-experimental (e.g., regression discontinuity, propensity-matched pre-post comparison)
Sample	School-aged children (ages 4-18 or US equivalent grades PreK-12)
Year	Study conducted between 1970 and July 2014
Language	Available in English

Third, the study design must have compared students who participated in the chess intervention to a comparison group of students who did not participate in the chess intervention. Qualitative case studies without a comparison group and anecdotal descriptions of chess programs were ineligible. Additionally, the studies must have used the same outcome measure for both the participating students and the comparison group.

Fourth, the sample must have been composed of school-aged children participating in a chess intervention (ages 4-18 or US equivalent grades of PreK-12).

Finally, this literature review included only studies conducted between 1970 and July 2014 that were available in English. Table 1 summarizes the inclusion criteria this literature review used to establish eligibility.

Literature searches

An exhaustive search of research on chess-based interventions was conducted in order to identify and gather studies that met the inclusion criteria in Table 1. Using "chess" as the primary search term, multiple databases of peer-reviewed published research were searched, including the following:

Academic Search Premier	Google Scholar
EconLit	ProQuest Dissertations & Theses
Education Research Complete	PsycINFO
E-Journals	Web of Science – Social Science Citation Index (SSCI)
ERIC	WorldCat

The websites of the following organizations, programs, and curriculum related to chess were also reviewed to identify eligible studies:

Berkeley Chess School	Ho Math Chess
Chess at Three	International Society for Chess Research
Chess for Success	It's Our Move
Chess in Schools and Communities	Kasparov Chess Foundation
Chess-in-the-Schools	National Scholastic Chess Foundation
Chesskid.com	Success Chess
Chess Magnet School Curriculum	Susan Polgar Foundation
Chess Palace Program	Think Like a King
Chess Program – Univ. of Texas, Dallas	The US Chess Federation
FirstMove	

After studies or previous literature reviews of chess studies were identified and obtained, the bibliographies were reviewed to identify additional studies that may not have emerged through search databases or chess organization websites. Once the search process began to detect only studies that had already been identified, a list of eligible studies was shared with several chess researchers to determine if any studies were missed through the search process.

Based on the inclusion criteria established in Table 1, studies that examined the impact of game-based interventions similar to chess (e.g., spatial or strategy games) could have been eligible for this literature review. However, after searching for these types of studies, none were identified that met the full set of inclusion criteria. As a result this literature review will focus entirely on chess intervention studies.

Additionally, the search process did not produce studies that examined non-cognitive outcome measures, such as grit and persistence. The literature review will discuss results for academic, cognitive, and behavioral outcome measures.

The literature search resulted in 51 studies on chess. Twenty-four of the 51 studies met the inclusion criteria for eligibility to be included in this literature review (see Appendix A). Of the studies not included, seven were reviews of research and did not include original results and 20 were deemed ineligible based on not meeting the inclusion criteria in Table 1 (see Appendix B).

Coding eligible chess studies

Each of the eligible chess studies listed in Appendix A was reviewed and coded using a process based on the WWC Study Review Guide.[3] Table 2 summarizes the information that was collected and coded from each eligible study. If relevant information was not presented in the published report (e.g., number of subjects, pretest and posttest statistics), study authors were contacted and asked to provide additional information.

[3] http://ies.ed.gov/ncee/wwc/DownloadSRG.aspx

Table 2. Information collected from eligible studies

Study Detail	Categories
Chess intervention	After-school or in-school Name of chess curriculum, if applicable Chess club or scholastic chess program Duration of chess intervention (i.e., number of weeks) Frequency of chess intervention (e.g., daily, twice per week, once per week, less than once per week) Amount of time per meeting (e.g., 60 minutes)
Study design	Comparison of chess participants and non-participants, with no controls for differences in groups Quasi-experiment, control for differences in groups by matching on student characteristics Experiment, control for differences by random assignment at student, classroom, or school-level
Sample	Number of participants (chess intervention and comparison group) Characteristics
Age of sample	Age range Grade levels
Location of study	City State Country
Outcome measures	Academic Behavioral Cognitive Non-cognitive
Assessment	Name of assessment Construct validity
Pretest statistics	Mean and standard deviation *t*-test statistic ANOVA F-test *p*-value Test for group equivalence at pretest
Posttest statistics	Mean and standard deviation *t*-test statistic ANOVA F-test Odds ratio Regression coefficient *p*-value

Quality of study design

The purpose of this literature review is to examine whether the body of research on chess interventions shows that chess has an impact on student outcomes. In order for a study to measure the impact of chess, the study must show that outcomes for individuals in the chess intervention were a result of participating in the program. The eligible studies varied in the quality of their research design and in turn, their ability to link findings with participation in the chess intervention. A majority of the studies used research designs that did not control for group equivalence when comparing chess participants and non-participants. As a result, the findings from these studies should be examined cautiously because the differences in outcomes between chess participants and the comparison group could be a result of differences in individual student characteristics, rather than the impact of the chess intervention.

Table 3 presents a study design quality classification system that will be used throughout this literature review to provide context when interpreting findings.

Table 3. Classification by quality of study design

Tier I	**Experiment that controls for differences by random assignment at student, classroom, or school-level; OR** **Quasi-experiment that controls for differences in groups by matching on student characteristics AND reports group equivalence on pretest results**
Tier II	**Quasi-experiment that controls for differences in groups by matching on student characteristics BUT does not show group equivalence on pretest results**
Tier III	**Comparison of chess participants and non-participants, with no controls for differences in groups on pretest results**

Three of the eligible studies randomly assigned classrooms or schools to the chess intervention and control groups (Romano, 2011; Sallon, 2013; Scholz et al., 2008).[4] Random assignment of classrooms or schools can be used as a strategy to include more students in the study since may be easier to randomly assign groups than individuals.

[4] While Forrest et al. (2005) randomly assigned two classrooms in the study to chess instruction and control group, the pretest indicated that students in the two classrooms were not equivalent. Therefore, the study was not considered a Tier I study for this literature review.

Romano (2011) randomly assigned 123 classrooms in 33 schools and included over 1,700 students. Sallon (2013) randomly assigned 14 schools and included nearly 500 students. While these two studies are the largest studies of chess interventions, both are dissertations that have not been published in peer-reviewed journals.

Studies that used quasi-experimental research designs, such as matching chess participants and comparison group individuals on relevant characteristics, were coded as Tier I studies if they provided information about the equivalence of the treatment and comparison groups. One study matched chess and comparison group students on IQ and concluded that there were no statistically significant differences between groups on the pretest (Van Zyl, 1991). This study met the requirements to be classified as a Tier I study.[5] The other two chess studies that used matching strategies did not provide evidence that there were no statistically significant differences between the matched students on a pretest (DuCette, 2009; Hermelin, 2004). Even when students were matched on demographic characteristics (e.g., gender, age, race/ethnicity), if the students differed on the outcome measure at pretest or information was not presented about whether there was group equivalence, it cannot be concluded that differences in the outcome measure after the intervention were the result of the program. As a result, these two studies were coded as Tier II studies.

The final classification, Tier III, was given to studies that compared a chess intervention group and a comparison group, but did not use a research design that controlled for differences between participating and non-participating students[6]. The biggest concern in interpreting the results of these studies is that students who participated in the chess intervention may have systematically differed from their non-participating peers in ways that impacted the differences in outcome measures. For example, students who perform better in math may be more likely to participate in chess clubs. Consequently, if a study showed that students in the chess club performed better in math, the higher performance of chess club participants may be the result of higher performing students joining chess rather than the intervention improving math scores among participants. These studies are included in this literature review because they make up the majority of eligible chess studies. However, the results from these studies should be reviewed with the understanding that differences may be the result of student selection into the chess intervention, rather than the impact of the program on participating students.

---

[5] The Van Zyl (1991) study tested group equivalence on an IQ test, a cognitive outcome measure, but reported differences in group outcomes with an academic performance measure.

[6] WWC would classify Tier III studies as "not meeting evidence standards." The Tier III studies have been included in this review to provide information from the body of research on chess interventions because the majority of studies fall into this study design category.

Calculating eligible chess study effect sizes

After the information listed in Table 2 was extracted and coded from the eligible chess studies and each study was categorized as Tier I, II, or III, effect sizes were calculated for each of the outcome measures. Similar to other fields of research, the eligible chess studies used a variety of assessments to measure the impact of the intervention (e.g., standardized tests from different states or countries) and reported the results in many different ways (e.g., gains from pretest to posttest, differences in means at posttest, analyses that controlled for student demographics, etc.). Converting the results from eligible chess studies into effect sizes standardizes the results and allows for the comparison of the magnitude and statistical significance of findings across studies.

Many studies presented more than one of the outcome measures of interest: academic – mathematics, academic – reading, cognitive, and behavioral. Additionally, many studies presented more than one finding per outcome measures. For example, studies presented academic results by grade level or presented academic results from different types of assessments (e.g., standardized assessments and end-of-course grades). Effect sizes were first calculated for every result presented in the studies, outcome measure by finding. For studies that reported more than one finding by outcome measure (academic, cognitive, and behavioral), the findings were pooled to generate one effect size per outcome measure per study.

Standardized mean difference effect size

All of the eligible chess studies reported posttest results for the chess intervention participants and comparison group. However, there were eligible studies that did not use a pretest. If all studies had presented pretest and posttest data, a standardized mean gain effect size could have been calculated. Given the nature of the way that results were presented, a standardized mean difference was calculated for all outcome measures in the eligible studies. The standardized mean difference effect size represents the difference in outcome measures at posttest between chess intervention participants and the comparison group. The standardized mean difference effect size is calculated using the formula: [7]

$$ES = \frac{\overline{M}_1 - \overline{M}_2}{s_p},$$

[7] Lipsey, M.W., & Wilson, D.B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: SAGE Publications.

where $\overline{M}_1$ is the mean of the outcome measure for the chess intervention participants, $\overline{M}_2$ is the mean of the outcome measure for the comparison group, and s_p is the pooled standard deviation of the outcome measure. The pooled standard deviation, s_p, is calculated using the formula: [8]

$$s_p = \sqrt{\frac{(n_1 - 1){s_1}^2 + (n_2 - 1){s_2}^2}{n_1 + n_2 - 2}},$$

where s_1 is the standard deviation for the chess intervention participants, s_2 is the standard deviation for the comparison group and n_1 and n_2 indicate the sample size for the chess participants and comparison group, respectively.

The standardized mean difference effect sizes present the estimate of the difference in outcome measure means between participants in the chess intervention and the comparison group in terms of standard deviations. For example, a positive effect size of 0.300 would indicate that participants in the chess program scored 0.300 standard deviations higher than the comparison group on the outcome measure.

Table 4. Converting statistics to standardized mean difference effect size

Statistic	Conversion Formula
t-test statistic	$ES = t\sqrt{\frac{n_1 + n_2}{n_1 n_2}}$
ANOVA F-test	$ES = \sqrt{\frac{F(n_1 + n_2)}{n_1 n_2}}$
Odds ratio	$ES = (Log\ (Odds\ ratio))\ \frac{\sqrt{3}}{\pi}$
Regression coefficient	Unstandardized regression coefficient = $(\overline{M}_1 - \overline{M}_2)$
p-value[9]	Determine *t*-value based on the *p*-value and degrees of freedom using a two-sided *t*-distribution table, where t-value^2 = F. Use the ANOVA F-test equation above to calculate *ES*.

[8] *Ibid.*

[9] It is possible to calculate effect sizes from an exact *p*-value (e.g., $p = 0.040$) and categorical *p*-values (e.g., use $p = 0.05$ if $p < 0.05$ is reported) that indicate the statistical significance of the difference between chess intervention participants and control group on a given outcome measure. Exact *p*-values are better for

Ten of the eligible chess studies did not report means and standard deviations for the outcome measures. However, the standardized mean difference effect size can still be calculated from other types of statistics. Table 4 presents conversion equations used to generate standardized mean difference effect sizes when means and standard deviations were unavailable.

The standardized mean difference effect size has been shown to be upwardly biased for small sample sizes, which is the case for many of the eligible chess studies. To correct for the upward bias, the standardized mean difference effect sizes were converted to Hedge's *g* effect sizes using the formula:[10]

$$Hedge's\ \mathrm{g} = ES\left(1 - \frac{3}{4N - 9}\right),$$

where N is the total sample size ($n_1 + n_2$) and ES is the biased standardized mean difference effect size.

It is also important to know the precision of the Hedge's *g* effect size. Effect sizes based on larger sample sizes are more precise than effect sizes based on smaller sample sizes. The precision is calculated by the standard error (SE) and the inverse variance weight (w). The SE for each Hedge's *g* effect size was calculated using the formula:[11]

$$SE = \sqrt{\frac{n_1 + n_2}{n_1 n_2} + \frac{(Hedge's\ \mathrm{g})^2}{2(n_1 + n_2)}},$$

where n_1 and n_2 indicate the sample size for the chess participants and comparison group, respectively. And the inverse variance weight for the Hedge's *g* effect size is:[12]

$$w = \frac{1}{SE^2}$$

calculating effect sizes. For some of the eligible chess studies, categorical *p*-values were used because they were the only statistic available.

[10] Lipsey & Wilson (2001)

[11] *Ibid.*

[12] *Ibid.*

Confidence intervals for Hedge's g effect size were also calculated to determine the lower and upper limits of the effect size and indicate whether the effect size was statistically significant. If zero is contained within the confidence interval band, the effect size is not statistically significant. The statistical significance of the effect sizes with 95% confidence was calculated using the formula:[13]

$$C.I.\,Lower\ Limit = Hedge's\ \mathrm{g} - 1.96(SE),$$

$$C.I.\,Upper\ Limit = Hedge's\ \mathrm{g} + 1.96(SE),$$

where SE is the standard error of Hedge's g effect size.

Weighted mean effect sizes

After Hedge's g effect sizes were calculated, the effect sizes were pooled by the chess intervention design (after-school or in-school) and by type (academic, cognitive, and behavioral). The academic performance outcome measures were further categorized by subject area (mathematics, reading, other). Pooling individual findings by intervention design and outcome measure category will allow this literature review to consolidate results and provide a more accurate estimate of the effect of chess interventions.

With the outcome measures organized by intervention design and type, weighted mean effect sizes and pooled confidence intervals were calculated to estimate the overall effect of chess interventions using the formulas:[14]

$$\overline{Hedge's\ \mathrm{g}} = \frac{\sum(w_i Hedge's\ \mathrm{g}_i)}{\sum w_i},$$

$$SE_{\overline{\mathrm{g}}} = \sqrt{\frac{1}{\sum w_i}},$$

$$Pooled\ C.I.\,Lower\ Limit = \overline{Hedge's\ \mathrm{g}} - 1.96\left(SE_{\overline{\mathrm{g}}}\right),$$

$$Pooled\ C.I.\,Upper\ Limit = \overline{Hedge's\ \mathrm{g}} + 1.96\left(SE_{\overline{\mathrm{g}}}\right),$$

[13] *Ibid.*
[14] *Ibid.*

where i indicates an effect size equal to 1 to k and w_i is the inverse variance weight for the Hedge's g effect size i.

Interpreting effect sizes

The direction, magnitude, and statistical significance of the standardized mean effect size and weighted mean effect size matter. For the purposes of this literature review, positive effect sizes indicate that students who participated in the chess intervention scored higher on the outcome measure than students in the comparison group. Negative effect sizes indicate that chess participants scored lower than the comparison group.

Deciding whether the magnitude of effect sizes in this report are substantively meaningful can be informed by comparing with results from high-quality education research studies. From a report that examined the results of over 100 education studies, average effect sizes that can be expected based on study characteristics that are presented in Table 5.[15] Depending on the type of assessment, type of intervention, and target recipients, the range of effect sizes for randomized studies is between .28 and .53.

Table 5.

Study Characteristic	Average Effect Size
Assessment Type	
Specialized or researcher developed	.53
Standardized test, narrow scope	.40
Standardized test, broad scope	.28
Type of Intervention	
Instructional format	.36
Teaching technique	.47
Instructional component or skill training	.50
Curriculum or broad instructional program	.32
Whole school program	.31
Target Recipients	
Individual students	.53
Small group	.40
Classroom	.41
Whole school	.30

Source: Lipsey et al (2012)

[15] Lipsey, M.W., Puzio, K., Yun, C., Hebert, M.A., Steinka-Fry, K., Cole, M.W., Roberts, M., Anthony, K.S., & Busick, M.D. (2012). *Translating the statistical representation of the effects of education interventions into more readily interpretable forms* (NCSER 2013=3000). Washington, DC: National Center for Special Education Research, Institute of Education Sciences, U.S. Department of Education.

For an effect size to indicate the difference between chess participants and the comparison group, it must be statistically significant. Throughout this literature review effect sizes and their confidence intervals are presented. If the confidence interval range includes zero, then the effect size is not statistically significant. Statistically insignificant results indicate that the study was unable to measure a difference between students in the chess intervention and students in the control group.

Summary of data coding

Table 6 presents all of the data collected and coded from the 24 eligible chess studies in this literature review, sorted by the study design classification Tier. The table includes each study's intervention type, sample size, age of children in the sample, study location, outcome measures (with name of assessment in parenthesis), and the Hedge's *g* effect sizes with confidence intervals for each of the outcome measures reported in the study.

Table 6. Eligible chess studies

Study	Chess Intervention	Intervention Details	Study Design	Sample	Sample Gender	Age of Sample	Location	Outcome Measures	Hedge's *g* Effect Size (C.I.)[16]
Christiaen (1976)	After-school	42 lessons, over two years	Tier I	Chess: 20 Comparison: 17	Male	5th-6th grade	Belgium	Academic, Math & Reading (DGB)	Academic, Math: 0.280 (-0.370, 0.930) Academic, Reading: 0.410 (-0.243, 1.063)
Fried & Ginsburg (n.d.)	In-school	18 lessons, one academic year	Tier I	Chess: 10 Comparison: 10	Male & Female	4th-5th grade	New York	Cognitive (WISC-R) Behavioral (Survey of School Attitudes)	Cognitive: 0.070 (-0.560, 0.700) Behavioral: 0.103 (-0.774, 0.980)
Hong & Bart (2007)	In-school	12 weekly lessons, one academic year	Tier I	Chess: 18 Comparison: 20	Male & Female	8-12 years old	South Korea	Cognitive (TONI-3 & RPM)	Cognitive: 0.172 (-0.280, 0.624)
Kakemi, Yektayar, & Abad (2012)	In-school	6 months	Tier I	Chess: 86 Comparison: 94	Male	5th, 8th, 9th grade	Iran	Academic, Math (Author) Cognitive (Unknown)	Academic, Math: 0.686 (0.385, 0.987) Cognitive: 0.819 (0.514, 1.123)
Romano (2011)	In-school	20-30 hours, one academic year	Tier I	Chess: 950 Comparison: 806	Male & Female	3rd grade	Italy	Academic, Math (Author)	Academic, Math: 0.340 (0.245, 0.434)
Sallon (2013)	In-school	30 hours, one academic year	Tier I	Chess: 201 Comparison: 282	Male & Female	2nd grade	England	Academic, Math (Author)	Academic, Math: 0.515 (0.331, 0.699)
Scholz, Niesch, Steffen, Ernst, Loeffler, Witruk, & Schwarz (2008)	In-school	Weekly, one academic year	Tier I	Chess: 31 Comparison: 22	Male & Female	Elem school	Germany	Academic, Math (Author)	Academic, Math: 0.204 (-0.344, 0.752)
Van Zyl (1991)	After-school	Weekly	Tier I	Chess: 80 Comparison: 80	Male & Female	5th-10th grade	South Africa	Academic, Math (Unknown)	Academic, Math: 0.640 (0.322, 0.958)

[16] Effect sizes by outcome measure may include results pooled from multiple findings.

Study	Chess Intervention	Intervention Details	Study Design	Sample	Sample Gender	Age of Sample	Location	Outcome Measures	Hedge's *g* Effect Size (C.I.)[16]
DuCette (2009)	After-school	Unknown	Tier II	Chess: 151 Comparison: 151	Male & Female	3rd-8th grade	Pennsylvania	Academic, Math & Reading (PSSA)	Academic, Math: 0.358 (0.131, 0.585) Academic, Reading: 0.249 (0.023, 0.475)
Hermelin (2004)	After-school	Unknown	Tier II	Chess: 38 Comparison: 38	Male & Female	5th-7th grade	South Africa	Academic, Math (End-of-Course Grades)	Academic, Math: 0.840 (0.371, 1.309)
Aciego, Garcia, & Betancort (2012)	After-school	Weekly, academic year	Tier III	Chess: 170 Comparison: 60	Male & Female	6-16 years old	Spain	Cognitive (WISC-R) Behavioral (TAMAI)	Cognitive: 0.388 (0.478, 0.299) Behavioral: -0.471 (-0.371, -0.570)
Barrett & Fish (2011)	In-school	Weekly, 30 weeks	Tier III	Chess: 15 Comparison: 16	Male & Female	6th-8th grade	Texas	Academic, Math (TAKS & End-of-Course Grades)	Academic, Math: 1.428 (0.867, 1.989)
Eberhard (2003)	In-school	Daily, semester	Tier III	Chess: 60 Comparison: 77	Male & Female	7th-8th grade	Texas	Cognitive (CogAT & NNAT)	Cognitive: -0.085 (-0.251, 0.081)
Ferguson (n.d.)	In-school	Weekly, 32 weeks	Tier III	Chess: 15 Comparison: 79	Male & Female	7th-9th grade	Pennsylvania	Cognitive (Watson-Glaser & Torrence Tests)	Cognitive: 0.782 (0.384, 1.181)
Forrest, Davidson, Shucksmith, & Glendinning (2005)	After-school	One academic year	Tier III	Chess: 18 Comparison: 18	Male & Female	3rd grade	Scotland	Academic, Reading (Neale) Cognitive (WISC-R) Behavioral (Bristol)	Academic, Reading: -0.004 (-0.466, 0.458) Cognitive: 0.613 (-0.055, 1.281) Behavioral: 0.400 (-0.260, 1.060)
Garcia (2008)	After-school	Weekly, one academic year	Tier III	Chess: 27 Comparison: 27	Male & Female	5th grade	Texas	Academic, Math & Reading (TAKS)	Academic, Math: 1.455 (0.855, 2.055) Academic, Reading: 1.436 (0.838, 2.034)

Study	Chess Intervention	Intervention Details	Study Design	Sample	Sample Gender	Age of Sample	Location	Outcome Measures	Hedge's *g* Effect Size (C.I.)[16]
Kramer & Filipp (n.d.)	In-school	Weekly, four academic years	Tier III	Chess: 84 Comparison: 83	Male & Female	Elem school	Germany	Cognitive (Unknown) Behavioral (Unknown)	Cognitive: 0.673 (0.452, 0.894) Behavioral: 0.267 (0.088, 0.447)
Liptrap (1998)	After-school	Weekly, unknown duration	Tier III	Chess: 23 Comparison: 269	Male & Female	5th grade	Texas	Academic, Math & Reading (TAAS)	Academic, Math: 1.134 (0.698, 1.570) Academic, Reading: 0.609 (0.180, 1.038)
Margulies (1992)	After-school	Two academic years	Tier III	Chess: 22 Comparison: 1,118	Male & Female	Elem school	New York	Academic, Reading (DRP)	Academic, Reading: 0.422 (0.000, 0.844)
Rifner (1992)	In-school	Weekly, one academic year	Tier III	Chess: 8 Comparison: 10	Male	7th grade	Indiana	Academic, Math & Reading (CTBS/4)	Academic, Math: 0.169 (-0.762, 1.100) Academic, Reading: 0.143 (-0.788, 1.074)
Sigirtmac (2012)	In-school	Unknown	Tier III	Chess: 50 Comparison: 50	Male & Female	6 years old	Turkey	Cognitive (Unknown)	Cognitive: 1.600 (1.150, 2.050)
Thompson (2003)	After-school	Weekly, one academic year	Tier III	Chess: 64 Comparison: 444	Male	6th-12th grade	Australia	Academic, Science (Author)	Academic, Science: 0.128 (-0.134, 0.390)
Trinchero (n.d.)	In-school	10-15 hours, one academic year	Tier III	Chess: 412 Comparison: 156	Male & Female	3rd-7th grade	Italy	Academic, Math (Author)	Academic, Math: 0.421 (0.228, 0.613)
Yap (2006)	After-school	30 lesson plans, two academic years	Tier III	Chess: 233 Comparison: 88	Male & Female	3rd-5th grade	Oregon	Academic, Math & Reading (OR) Behavioral (Coopersmith & Student Behavior)	Academic, Math: 0.276 (0.030, 0.522) Academic, Reading: 0.152 (-0.093, 0.397) Behavioral: -0.018 (-0.191, 0.155)

Literature review results

The following sections present weighted mean Hedge's *g* effect sizes and confidence intervals by chess intervention type (after-school and in-school) and outcome measures (academic – mathematics, academic – reading, cognitive, and behavioral). For each of the categories, the results are pooled by study design quality classifications. The presentation of results in this manner shows how results may differ based on the rigor of the study methodology.

For example, results that include all of the studies (Tiers I, II, and III) or results from Tier II and Tier III studies should be interpreted with more caution than results presented for the Tier I studies alone. Based on research design, results from the Tier I studies estimate the impact of the chess intervention on differences in student outcomes between participants and non-participants, whereas results that include Tier II and Tier III studies may be biased and reflect differences between participants and non-participants that are not due to the chess intervention.

After-school chess programs

In total, there were 11 eligible chess studies that looked at the impact of after-school chess programs on student outcomes (see Table 6). Of the 11 studies, two were classified as Tier I, two were classified as Tier II, and seven were classified as Tier III. [17]

Out of the 11 studies that examined after-school chess, seven used mathematics performance as the outcome measure. Table 7 presents weighted mean Hedge's *g* effect sizes, with confidence intervals, for the chess studies that examined the impact of after-school chess programs on mathematics performance. Overall, the results from the after-school chess studies on mathematics are positive and statistically significant irrespective of study design classification. The pooled effect size for all three Tiers of studies is 0.532. The standardized mean difference effect size for the Tier I studies is 0.570. The results suggest that the chess interventions analyzed by the after-school chess studies improved the math performance of chess participants compared to the comparison group.

[17] The outcome variable for one of the Tier III after-school studies (Thompson, 2003) is science academic performance. Since it was the only study that examined science, it is not included in the presentation of results. The outcome for chess participants from this study was statistically insignificant.

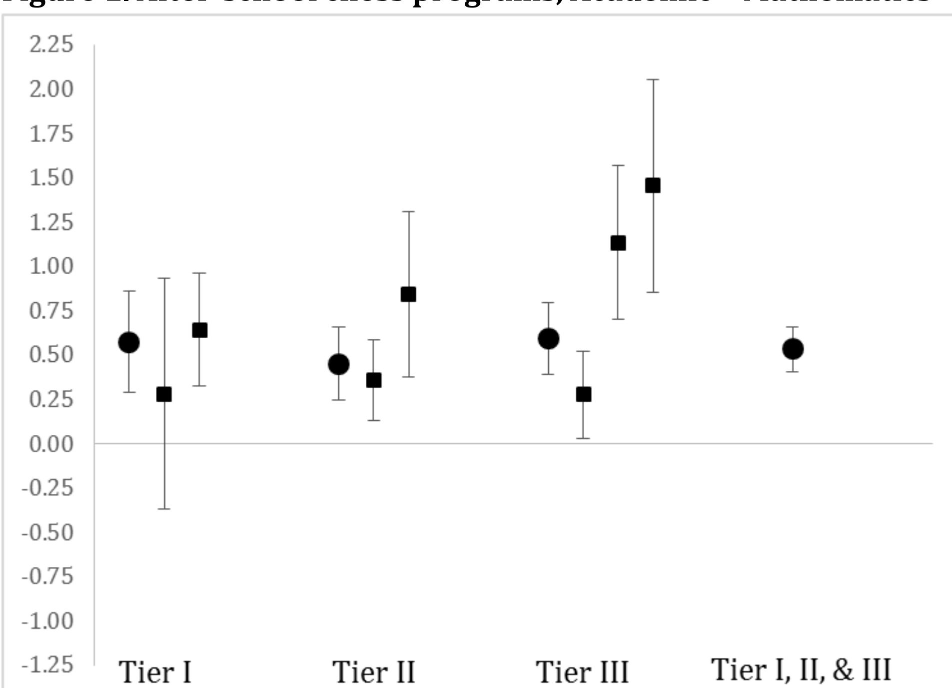
Table 7. After-school chess programs, Academic – Mathematics

Tier I, II, & III	0.532 (0.404, 0.661)
Tier III	0.594 (0.392, 0.795)
Tier II	0.450 (0.245, 0.654)
Tier I	0.570 (0.284 , 0.856)

Note: Weighted mean effect sizes (95 percent confidence interval)

Figure 1 shows the weighted mean effect sizes (circles) and individual mean effect sizes for each study (squares), by study classification type. The figure displays the range of effect sizes from the after-school chess studies and indicates that two of the three Tier III effect sizes were much larger in magnitude than findings from the Tier I and Tier II studies, studies that were conducted with more rigorous research design methodologies. Of the two Tier I studies, one of the effect sizes was statistically insignificant (Christiaen, 1976). Hence, the positive and statistically significant weighted mean effect size for Tier I studies (0.570) is driven by one study (Van Zyl, 1991).

Figure 1. After-school chess programs, Academic – Mathematics



Note: Circles indicate weighted mean effect sizes. Squares indicate standardized mean difference effect sizes from individual studies. Whiskers indicate 95 percent confidence interval.

Seven of the after-school chess studies looked at the impact of chess on reading outcome measures. Table 8 presents the pooled mean effect sizes for reading. The results for Tier II, Tier III, and the results for all eligible studies combined are positive and statistically significant. The reading results from the one Tier I study—the most rigorously designed—is statistically insignificant.

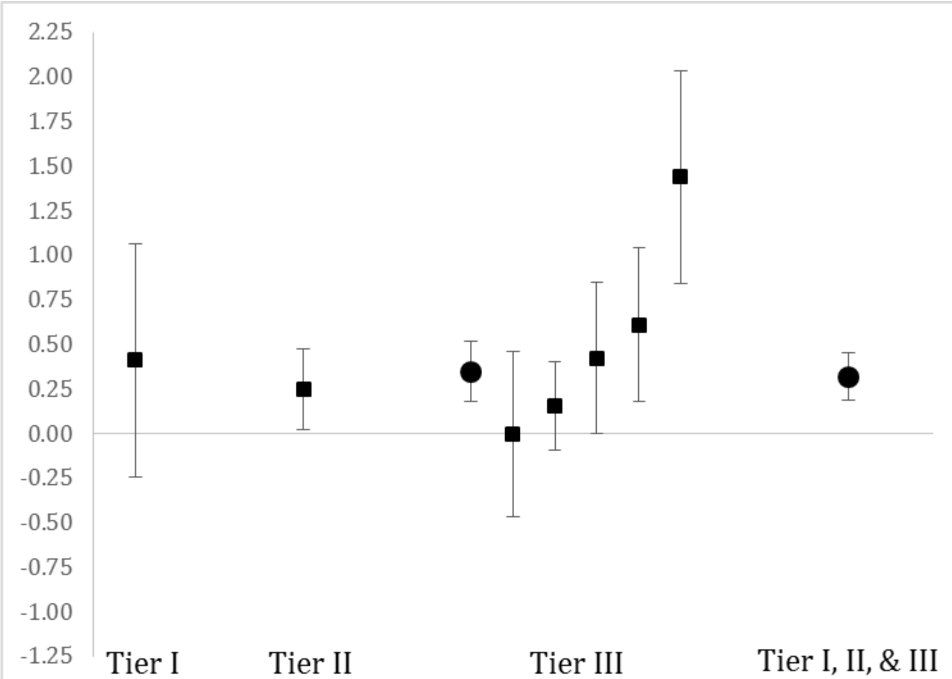
Table 8. After-school chess programs, Academic – Reading

Tier I, II, & III	0.316 (0.184, 0.449)
Tier III	0.347 (0.178, 0.516)
Tier II	0.249 (0.022, 0.475)
Tier I	0.410 (-0.243, 1.063)

Note: Weighted mean effect sizes (95 percent confidence interval)

Figure 2 displays the effect sizes by study classification type. In reading, the majority of studies are Tier III studies. Of the five eligible Tier III studies, two of the effect sizes were statistically insignificant (Forrest et al., 2005; Yap, 2006) and one was nearly insignificant (Marguiles, 1992). There was one Tier I study (Christiaen, 1976) and one Tier II study (DuCette, 2009). The positive and statistically significant weighted mean effect size from all of the eligible studies (0.316) is driven by the Tier III studies.

Figure 2. After-school chess programs, Academic – Reading



Note: Circles indicate weighted mean effect sizes. Squares indicate standardized mean difference effect sizes from individual studies. Whiskers indicate 95 percent confidence interval.

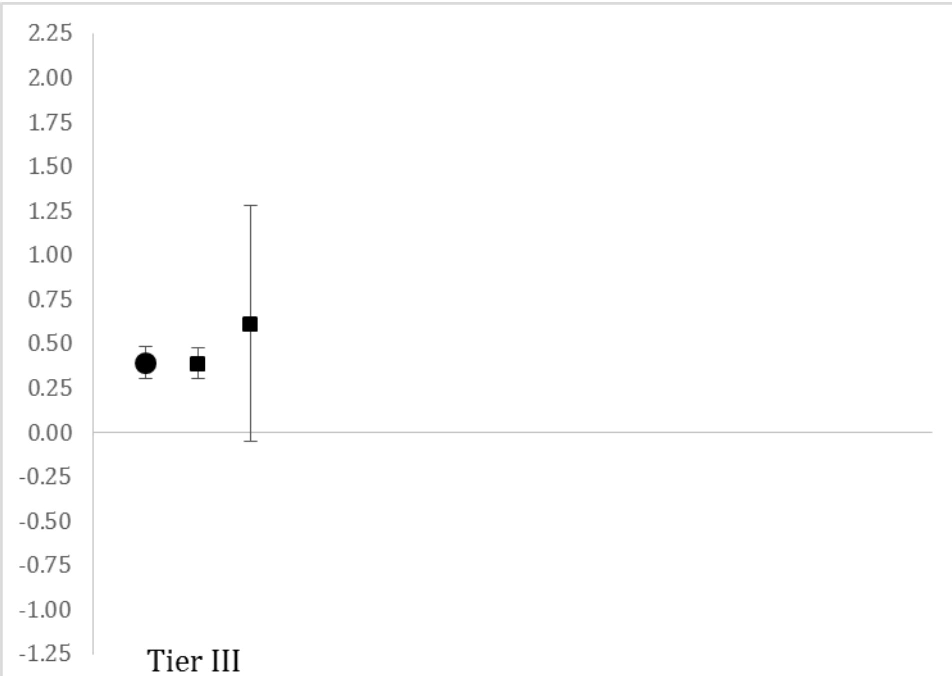
Table 9 and Figure 3 present results from eligible studies that examined after-school chess programs on cognitive outcome measures. There were only two after-school studies that used cognitive outcome measures and they were classified as Tier III studies (Aciego et al., 2012; Forrest et al., 2005). The weighted mean effect size in Table 9 is positive and statistically significant (0.474), but it is being driven by one Tier III study. Figure 3 shows that the finding from one study is statistically insignificant. Moreover, the Tier III studies were not designed to control for differences between chess program participants and comparison groups. Consequently, the results should be interpreted with caution; the differences between chess program participants and comparison groups may be attributable to differences between groups rather than the impact of the chess intervention.

Table 9. After-school chess programs, Cognitive

Tier III	0.474 (0.375, 0.572)

Note: Weighted mean effect sizes (95 percent confidence interval)

Figure 3. After-school chess programs, Cognitive



Note: Circles indicate weighted mean effect sizes. Squares indicate standardized mean difference effect sizes from individual studies. Whiskers indicate 95 percent confidence interval.

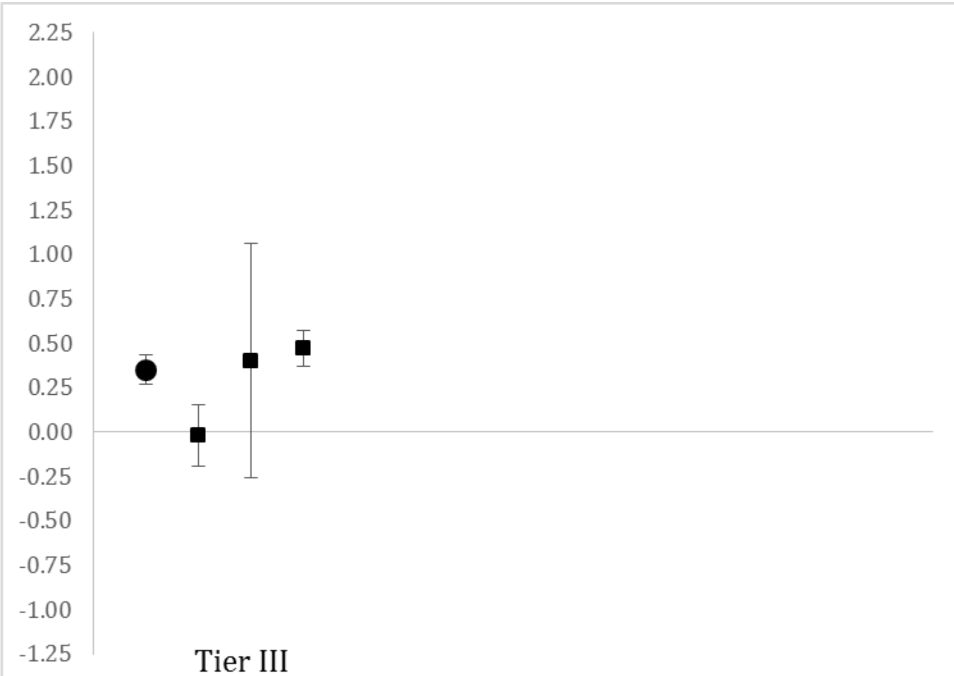
Table 10 and Figure 4 present results from three after-school studies that examined behavioral outcome measures (Aciego et al., 2012; Forrest et al., 2005; Yap, 2006). The research design methodologies of the three studies were classified as Tier III. The weighted mean effect size for the three studies is positive and statistically significant (0.351). However, similar to the cognitive outcome measures, the result should be interpreted cautiously. Not only are the three studies Tier III studies, but Figure 4 shows that the results from two of the studies were statistically insignificant.

Table 10. After-school chess programs, Behavioral

Tier III	**0.351 (0.265, 0.436)**

Note: Weighted mean effect sizes (95 percent confidence interval)

Figure 4. After-school chess programs, Behavioral



Note: Circles indicate weighted mean effect sizes. Squares indicate standardized mean difference effect sizes from individual studies. Whiskers indicate 95 percent confidence interval.

In-school chess programs

There were 13 eligible studies that investigated the impact of in-school chess programs on student outcome measures (see Table 6). Of the 13 studies, six were classified as Tier I studies, none were classified as Tier II studies, and seven were classified as Tier III studies.

Of the 13 in-school chess studies, seven examined the impact of chess programs on mathematics performance outcome measures. Table 11 presents weighted mean Hedge's *g* effect sizes, with confidence intervals, for the chess studies that examined in-school chess programs on mathematics performance. The results for Tier I, Tier III, and combined Tier I and III studies are positive and statistically significant. The Tier I weighted mean effect size suggests that chess interventions conducted during the school day had an estimated positive impact on student math performance of 0.395 standard deviations with statistical significance.

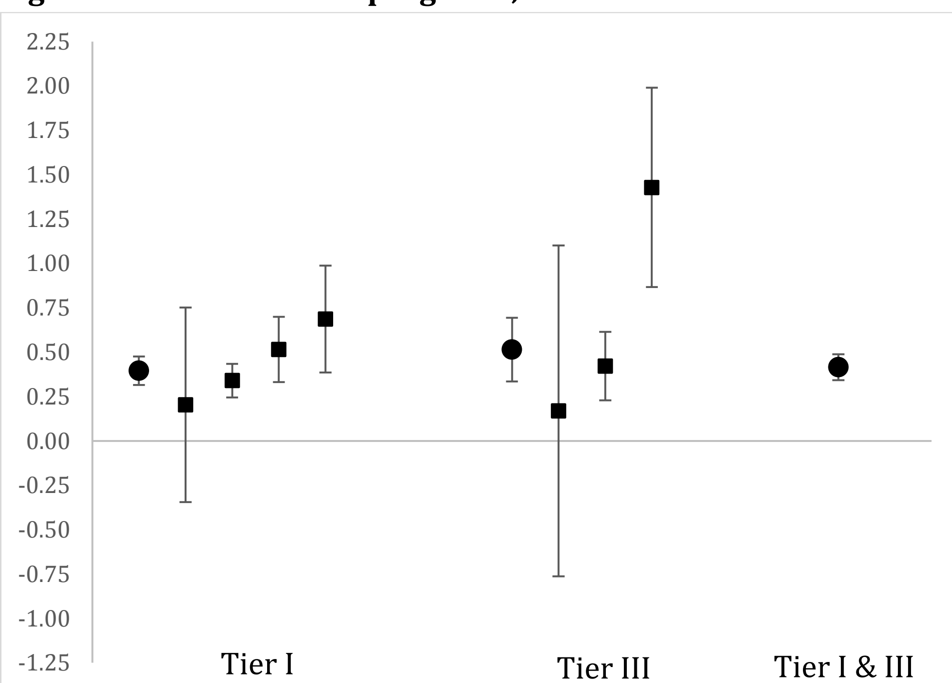
Table 11. In-school chess programs, Academic – Mathematics

Tier I & III	0.415 (0.342, 0.488)
Tier III	0.514 (0.335, 0.692)
Tier I	0.395 (0.315, 0.475)

Note: Weighted mean effect sizes (95 percent confidence interval)

Figure 5 shows the weighted mean effect sizes (circles) and individual standardized mean difference effect sizes (squares) for the in-school studies looking at mathematics. There were four Tier I studies. Three of the Tier I studies indicated positive and statistically significant results (Kazemi et al., 2012; Romano, 2011; Sallon, 2013), whereas the results from one of the studies was statistically insignificant (Scholz et al., 2008). Findings from two of the Tier III studies were statistically significant (Barrett & Fish, 2011; Trinchero, n.d.), while one was statistically insignificant (Rifner, 1992).

Figure 5. In-school chess programs, Academic – Mathematics



Note: Circles indicate weighted mean effect sizes. Squares indicate standardized mean difference effect sizes from individual studies. Whiskers indicate 95 percent confidence interval.

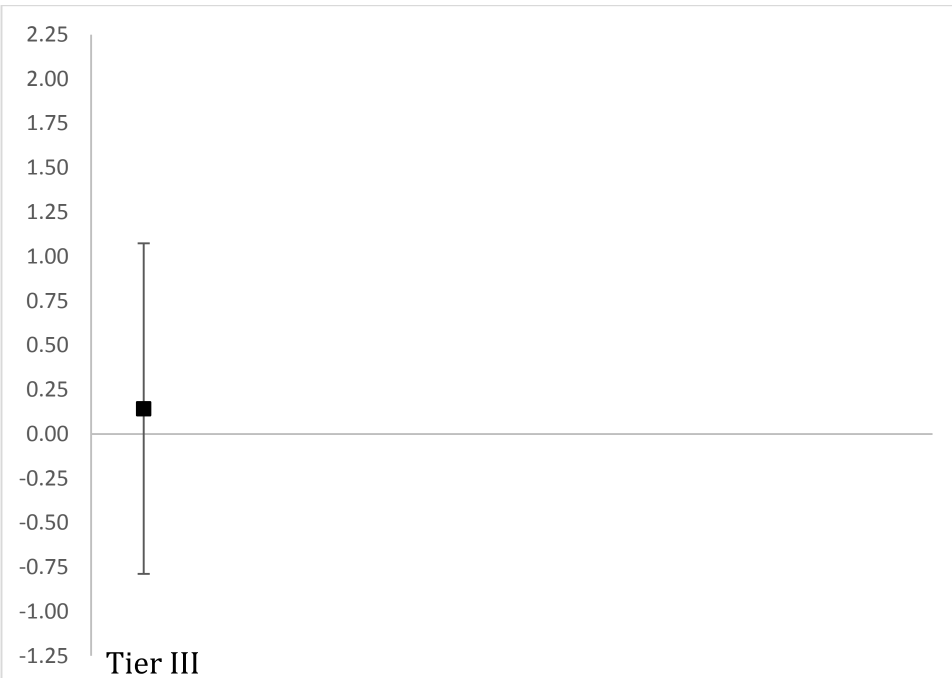
Table 12 and Figure 6 present results from the single Tier III study that focused on the impact of in-school chess programs on reading performance (Rifner, 1992). No Tier I or Tier II studies considered this outcome. Results from this one study indicate that in-school chess programs did not have a statistically significant effect on the reading performance of participants compared with non-participants.

Table 12. In-school chess programs, Academic – Reading

Tier I	**0.143 (-0.788, 1.074)**

Note: Weighted mean effect sizes (95 percent confidence interval)

Figure 6. In-school chess programs, Academic – Reading



Note: Circles indicate weighted mean effect sizes. Squares indicate standardized mean difference effect sizes from individual studies. Whiskers indicate 95 percent confidence interval.

There were seven in-school chess studies that examined cognitive outcome measures. Table 13 presents weighted mean effect size results from the studies. The results are positive and statistically significant for Tier I, Tier III, and the combination of Tier I and III studies. The Tier I studies had a pooled effect size of 0.541 with statistical significance.

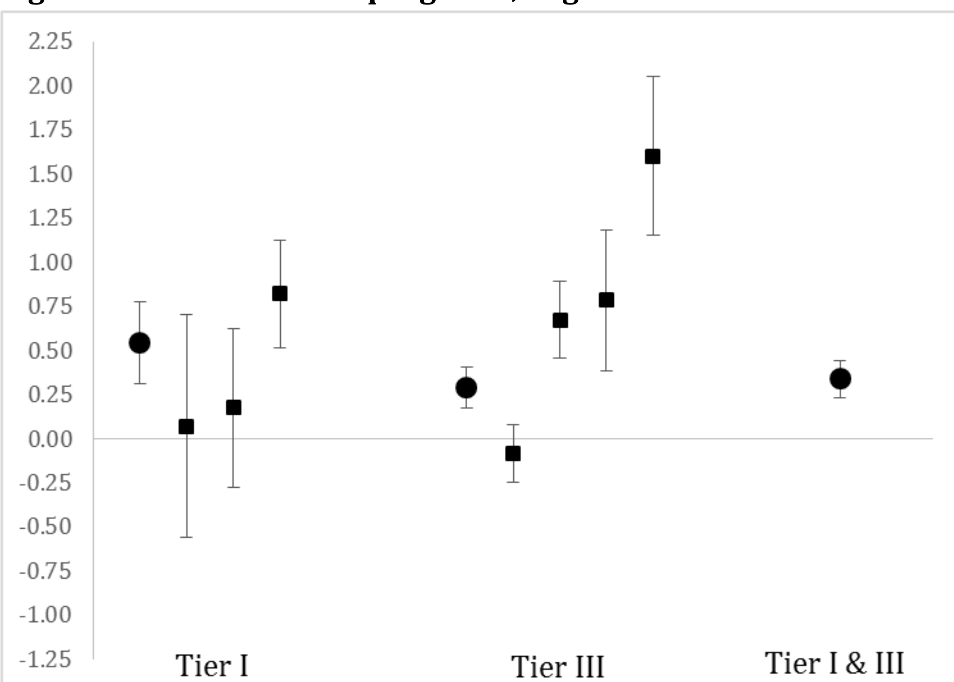
Table 13. In-school chess programs, Cognitive

Tier I & III	0.387 (0.280, 0.495)
Tier III	0.346 (0.225, 0.467)
Tier I	0.541 (0.307, 0.776)

Note: Weighted mean effect sizes (95 percent confidence interval)

Figure 7 displays the effect sizes for the eligible in-school chess studies. Of the three Tier I studies, one of the studies had a statistically significant finding (Kazemi et al., 2012), and the findings from two of the studies were statistically insignificant (Fried & Ginsburg, n.d.; Hong & Bart, 2007).

Figure 7. In-school chess programs, Cognitive



Note: Circles indicate weighted mean effect sizes. Squares indicate standardized mean difference effect sizes from individual studies. Whiskers indicate 95 percent confidence interval.

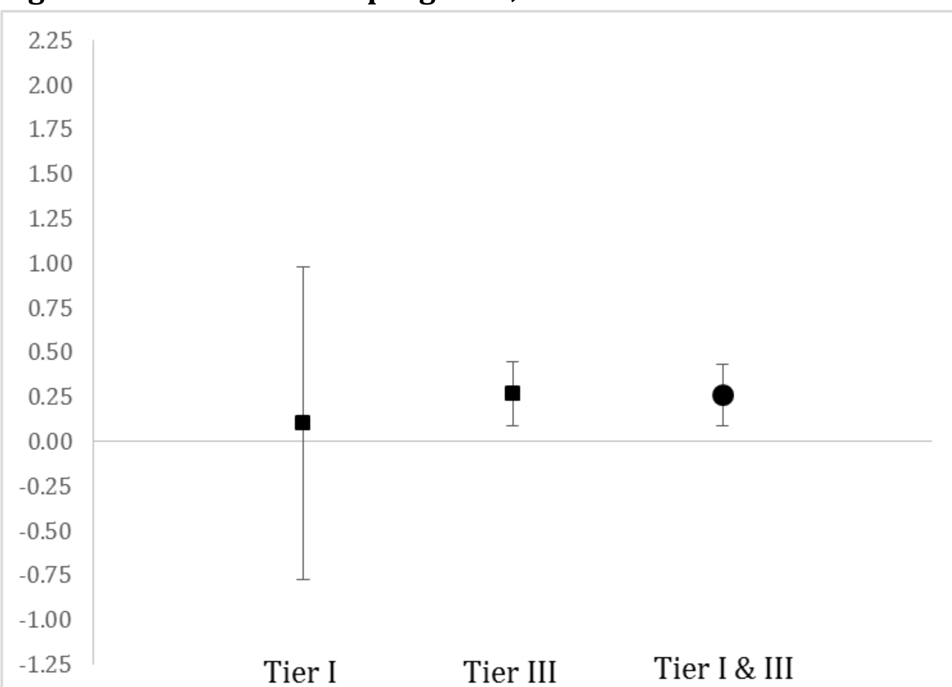
The final set of results for behavioral outcome measures from the in-school chess studies are presented in Table 14 and Figure 8. There were two in-school chess studies that looked at behavioral outcome measures. One of the studies was classified as Tier I, with a positive but statistically insignificant effect (Fried & Ginsburg, n.d.). The other study was classified as Tier III and had a positive and statistically significant result (Kramer & Filipp, n.d.).

Table 14. In-school chess programs, Behavioral

Tier I & III	**0.261 (0.086, 0.436)**
Tier III	**0.267 (0.088, 0.447)**
Tier I	**0.103 (-0.774, 0.980)**

Note: Weighted mean effect sizes (95 percent confidence interval)

Figure 8. In-school chess programs, Behavioral



Note: Circles indicate weighted mean effect sizes. Squares indicate standardized mean difference effect sizes from individual studies. Whiskers indicate 95 percent confidence interval.

Discussion of literature review results

Based on the set of chess studies eligible for this literature review, the pooled effect sizes from rigorous Tier I chess studies showed:

1. After-school chess programs had a positive and statistically significant impact on student mathematics outcomes (Table 7 and Figure 1).

2. In-school chess interventions had a positive and statistically significant impact on student mathematics and cognitive outcomes (Tables 11 & 13 and Figures 5 & 7).

The Tier I studies did not show statistically significant findings for cognitive outcomes in the after-school chess programs or for reading and behavioral outcomes in either after-school or in-school chess programs. While the two primary outcomes listed above are based on Tier I studies that used rigorous research design methodologies, the results should be interpreted cautiously given the small number of eligible studies that the pooled results encompass. However, the findings from the larger literature review (i.e., Tier II and Tier III studies) were generally consistent with those of the more rigorous Tier I studies. This suggests that the overall findings may have some general validity and robustness in that the Tier II and Tier III findings were congruent and add to the generalizability of the results by examining additional locales.

For the after-school chess studies, there were two Tier I studies that examined mathematics performance. The pooled results suggested that chess had an effect size of .57 on the math performance of chess participants. One of the studies did not find a statistically significant effect of after-school chess on math (Christiaen, 1976). This Tier I study was conducted in 1976 in Belgium and included 37 subjects. The second Tier I study showed a positive and statistically significant effect size for math (Van Zyl, 1991). The three year study was completed in 1990 and included 180 students in South Africa. The study did not implement a formal after-school chess program. Rather, students who participated in their schools' after-school chess club were matched using IQ scores with students who did not play chess.

The two Tier I after-school chess studies did not provide significant detail about the programmatic focus of the respective chess programs. However, the programs in the two studies appear to have been after-school chess clubs with a focus on competitive play, rather than interventions that used chess as part of a scholastic program to improve student outcomes. Moreover, the findings from the Tier I after-school chess studies are not generalizable beyond the location and samples of the two studies (i.e., early elementary grade students in countries outside of the United States). With only two Tier I after-school studies, the research base is too limited to conclude whether after-school chess clubs positively impact student outcomes.

There were four Tier I studies that examined the impact of in-school chess programs on student mathematics performance. The results from one of the four studies were statistically insignificant (Scholz et al., 2008). Of the three Tier I studies that indicated positive and statistically significant results, one study included 180 students in Iran (Kazemi et al., 2012), while the other two contained the largest samples of all of the chess studies. An in-school chess study from Italy included over 1,756 students in 123 classrooms that were randomly assigned to receive chess instruction (Romano, 2011). And an in-school chess study from England included 483 students 14 schools (Sallon, 2013).

The pooled effect size from the Tier I in-school chess studies on cognitive outcomes was positive and statistically significant. There were three studies that examined cognitive outcomes measures and qualified as Tier I. Two studies reported results that were statistically insignificant (Fried & Ginsburg, n.d.; Hong & Bart, 2007), and one study from Iran demonstrated positive and statistically significant results (Kazemi et al., 2012).

Among the Tier I in-school chess studies, the chess interventions were administered as a substitute for math classes during the week or as an element of an existing math class. Instead of receiving scholastic chess instruction, comparison group students in the in-school chess studies received additional math instruction. The in-school chess interventions did not focus on chess as a competitive play game, like the chess clubs studied by the after-school chess studies. Instead, the in-school chess studies examined chess programs that reflect the trend towards using chess as a component of an academic curriculum, whether the program is administered during the school day or after-school. There are several large studies underway in England[18], Germany, Spain[19], and Sweden that use random assignment to examine in-school chess programs. These studies will add to the literature on the impact of in-school scholastic chess interventions and perhaps deepen what is known about the components of these interventions that may have a differential impact on student outcomes.

---

[18] The study in England is an evaluation of *Chess in Schools* that will randomly assign 100 elementary schools to treatment and control groups, with a total sample size of approximately 3,000 students. 30 weeks of chess instruction will take place during the school day. The preliminary study will be available in the fall of 2015.

[19] There are three studies underway in Spain:

- Researchers from the University of Girona and the University of Lleida are studying in-school chess in Catalonia. The study includes approximately 100 elementary schools and 3,000 students.
- The Cantabria government is studying in-school chess in 31 schools with roughly 1,700 students.
- The City of Rendondela in Galicia is studying in-school chess in 13 schools with approximately 150 students.

Taken as a whole, the positive mathematics and cognitive outcome results from in-school chess studies may be explained by the chess programs being incorporated into students' weekly academic schedules, instruction during the school day leading to higher attendance rates and lower attrition, administering the program for an extended period of time, and connecting the intervention with math instruction and curriculum.

Theoretically these in-school scholastic chess programmatic components could be implemented in an after-school chess program format. In fact, research on effective after-school programs suggest characteristics that are similar to the components of the in-school chess programs: specific goals, structured content based on sound instructional techniques, and high student attendance.[20] An after-school scholastic chess program that was designed to mimic in-school chess (e.g., a program with a focus on chess skills related to academic performance or a program aligned to the school's math curriculum, using strategies to ensure high attendance) could potentially demonstrate large effect sizes. The pooled effect size in mathematics performance in this literature review for in-school studies was 0.395, which would be comparable to large-scale educational interventions in the United States.

Recommendations for future research

Given that all of the Tier I studies in this literature review were conducted in countries outside of the United States, additional research is needed to determine whether scholastic chess programs in the United States can have the same positive impact as the in-school chess programs in other countries. A rigorous study of a scholastic chess program would need to consider the following:

- Study design
- Sample size needed to estimate statistically significant effect size
- Details of the activities of students in the chess intervention and comparison group
- Outcome measures

Based on the study design quality classifications described earlier in this literature review, Tier I studies that are able to estimate the impact of a scholastic chess program on student outcomes would need to be designed to randomly assign students to the chess program and comparison group. Alternatively, the study could match students in a chess program with students in a comparison group using the same instrument for pretest and posttest.

---

[20] For example: Apsler, R. (2009). After-school programs for adolescents: A review of evaluation research. *Adolescence, 44*(173);

Power analysis can be used to determine the study sample size needed to generate an estimated effect size. If a new study was to produce the magnitude of the mathematics pooled effect size from the Tier I after-school chess studies (roughly 0.500), the new study would need to include a minimum of 126 students (63 in the chess intervention and 63 in the comparison group).[21] To estimate the effect size produced in the in-school chess studies (roughly 0.300), the new study would need to include a minimum of 350 students (175 in the chess intervention and 175 in the comparison group).[22] Depending on the size of the school, a new study may need to include more than one school to obtain a minimum of 126 (or 350) students who can be randomly assigned to chess programs or the comparison group.

When designing a new study of the impact of scholastic chess programs, significant consideration would also need to be given to the content of the chess intervention and the activities of the comparison group. For example, if the students assigned to the comparison group participate in an academic after-school activity like math tutoring or math instruction during the school day, it may be more difficult to discern a difference between the chess intervention and the comparison group if the student outcome measure is math performance.

A more complicated study design could use more than one type of treatment and comparison groups, such as randomly assigning students to two after-school chess intervention groups where one focuses solely on chess skills and the second aligns the chess instruction to the school's curriculum. In this scenario, the study could use a comparison group that participates in an after-school activity unrelated to chess. With this design, the study could compare the results for the two treatment groups, as well as the treatment groups compared with the comparison group.

A new study would have a number of options for outcome measures. This literature review revealed that previous studies of chess interventions used a wide variety of academic, cognitive, and behavioral outcome measures, largely because there are many hypotheses regarding how chess may impact students. A new study of after-school chess could choose to focus on academic measures, such as standardized test scores. However, the study may also want to include malleable cognitive measures, such as critical thinking, reasoning, and problem solving skills.

---

[21] Power analysis assumes power of 0.8 (e.g., the study would be able to detect statistically significant differences between groups 80% of the time) and statistical significance of 95%. If the power is increased to 0.9, the sample would need to include a minimum of 170 students (85 students in each group).
[22] Increasing the power to 0.9, the sample would need to include a minimum of 468 students (234 in each group).

Appendix A: Eligible chess studies

Aciego, R., Garcia, L., & Betancort, M. (2012). The benefits of chess for the intellectual and social-emotional enrichment in schoolchildren. *The Spanish Journal of Psychology, 15*(2), 551-559.

Barrett, D. C. & Fish, W. W. (2011). Our move: Using chess to improve math achievement for students who receive special education services. *International Journal of Special Education, 26*(3), 181-193.

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

DuCette, J. (2009). *An evaluation of the Chess Challenge Program of ASAP/After School Activities Partnerships*. Philadelphia, PA: After School Activities Partnerships. Retrieved from: http://www.sp2.upenn.edu/ostrc/docs/document_library/reqi/Research%20and%20Evaluation/An%20Evaluation%20of%20the%20Chess%20Challenge%20Program%20of%20ASAP.pdf

Eberhard, J.W. (2003). *The relationship between chess instruction and verbal, quantitative, and nonverbal reasoning abilities of economically disadvantaged students*. Unpublished doctoral dissertation, Texas A&M University - Corpus Christi.

Ferguson, R. (n.d.). *Educational benefits of chess: Summary based on research and articles*. Retrieved from: http://uschesstrust.com/wp-content/uploads/2007/08/educational-benefits-of-chess-summary-based-on-research-articles-by-dr-robert-ferguson.pdf.

Forrest, D., Davidson, I., Shucksmith, J., & Glendinning, T. (2005). *Chess development in Aberdeen's primary schools: A study of literacy and social capital*. Aberdeen, Scotland: University of Aberdeen. Retrieved from: http://www.scotland.gov.uk/Resource/Doc/930/0009711.pdf.

Fried, S., & Ginsburg, N. (n.d.). *The effect of learning to play chess on cognitive, perceptual and emotional development in children*. Unpublished manuscript.

Garcia, N.V. (2008). *Scholastic chess club participation and the academic achievement of Hispanic fifth grade students in south Texas*. Unpublished doctoral dissertation, University of Houston.

Hermelin, R. (2004). *The effect of playing chess on the mathematics achievement of primary school learners in two schools in KZN*. Unpublished master's thesis, University of KwaZulu-Natal, Durban, South Africa.

Hong, S & Bart, W.M. (2007). Cognitive effects of chess instruction on students at risk for academic failure. *International Journal of Special Education, 22*(3).

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

Kramer, A & Filipp, S. (n.d.) *Chess at Trier-Olewig Primary School: Summary and evaluation of the outcomes of the German School Chess Foundation (short version)*. Retrieved from: http://www.chessinschools.co.uk/download/Summary%20and%20Evaluation%20of%20the%20Outcomes%20of%20the%20German%20School%20Chess%20Foundation%20(English)%20%20.pdf.

Liptrap, J.M. (1998). Chess and standard test scores. *Chess Life, 41-43*.

Margulies, S. (1992). *The effect of chess on reading scores: District Nine chess program; Second year report*. New York, NY: The American Chess Foundation.

Rifner, P. J. (1992). *Playing chess: A study of the transfer of problem-solving skills in students with average and above average intelligence*. Unpublished doctoral dissertation, Purdue University.

Romano, B. (2011). *Does playing chess improve math learning? Promising (and inexpensive) results from Italy*. Unpublished doctoral dissertation, University of Pennsylvania.

Sallon, S. (2013). *An investigation into whether learning to play chess at a young age increases cognitive ability*. Unpublished dissertation, Manchester Metropolitan University.

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

Sigirtmac, A. D. (2012). Does chess training affect conceptual development of six-year-old children in Turkey? *Early Child Development and Care, 182*(6).

Thompson, M. (2003). Does the playing of chess lead to improved scholastic achievement? *Issues in Educational Research, 13*.

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

Van Zyle, A.S.A.J. (1991). *The significance of playing chess in improving a child's intellectual actualisation*. Unpublished doctoral dissertation, University of Pretoria, South Africa.

Yap, K.O. (2006). *Chess for Success evaluation: Final report*. Portland, Oregon: Northwest Regional Educational Laboratory. Retrieved from: http://media.oregonlive.com/beaverton_news/other/Chess%20for%20Success%20Final%20Evaluation%20Report%2010-31-06%20(2).pdf.

Appendix B: Ineligible chess studies

Study does not compare students in chess intervention with a comparison group

Adams, T. C. (2012). Chess from square a1: Incorporating chess into the gifted class. *Gifted Child Today, 35*(4), 243-251.

Argentin, G., Martini, A., & Romano, B. (2012). *Standing on the shoulders of chess masters: Using RCTs to produce evidence on teaching chess in schools*. Paper presented at the INVALSI-APPAM_UMD international conference "Improving education through accountability and evaluation" Rome, 3-5 October 2012.

Berkman, R. M. (2004). The chess and mathematics connection: More than just a game. *Mathematics Teaching in the Middle School, 9*(5), 246-250.

Ferreira, D. & Palhares, P. (2008). Chess and problem solving involving patterns. *The Montana Mathematics Enthusiast, 5*(2&3).

Korenman, M., Korenman, T., & Lyutykh, E. (2009). Checkmate: A chess program for African-American male adolescents. *International Journal of Multicultural Education, 11*(1).

Smith, J. & Sullivan, M. (1997, November). *The effects of chess instruction on students' level of field dependence/independence*. Paper presented at the Annual Meeting of the Mid-South Educational Research Association, Memphis, TN.

Outcome measure is chess skills

Bilalić, M., McLeod, P., & Gobet, F. (2007). Does chess need intelligence?—A study with young chess players. *Intelligence, 35*(5), 457-470.

de Bruin, A. B., Rikers, R. M., & Schmidt, H. G. (2007). The effect of self-explanation and prediction on the development of principled understanding of chess in novices. *Contemporary Educational Psychology, 32*(2), 188-205.

Gobet, F., & Campitelli, G. (2007). The role of domain-specific practice, handedness, and starting age in chess. *Developmental Psychology, 43*(1), 159.

Horgan, D. D. and Morgan, D. (1990) Chess expertise in children. *Applied Cognitive Psychology, 4*, 109-128.

Kiesel, A., Kunde, W., Pohl, C., Berner, M. P., & Hoffmann, J. (2009). Playing chess unconsciously. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(1), 292.

Schiff, S.M. (1991). Chess strategies: A course of study designed as an introduction to chess thinking. (Unpublished doctoral dissertation). Teachers College, Columbia University, New York City, NY.

Waters, A. J., Gobet, F. & Leyden, G. (2002). Visuospatial abilities of chess players. *British Journal of Psychology, 93*, 557–565.

Review of research

Berkley, D. (n.d.). *Chess instruction to improve algorithmic computation skills, mathematics, and life*. Retrieved from: http://mathandchessconsulting.com/docs/Chess_Position_paper.pdf.

Costello. P.J.M. (2013). The gymnasium of the mind: teaching chess in early childhood. *Early Child Development and Care, 183*(8), 1133-1146.

Ferguson, R. (1995). *Chess in education: Research summary. A review of key chess research studies*. Report presented at the BMCC Chess in Education "A Wise Move" Conference, New York City, NY.

Ferguson, R. C. (2006). *Teacher's guide: Research and benefits of chess*. New Windsor, NY: United States Chess Federation.

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.

McDonald, 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.

Niro, F.A. (2014). On the shoulders of Adriaan de Groot: Status of chess research in the early 21st century. *Journal of Chess Research, 1*(1).

Study unavailable

Bankauskas, D. (2000). Teaching chess to young children. *Young Children, 55*(4), 33-34.

Frank, A. (1978). *Chess and aptitudes* (Epstein, S., Trans.). Unpublished doctoral dissertation.

Galitis, I. (2002). Stalemate: girls and a mixed-gender chess club. *Gender and Education, 14*(1), 71-83. [Outcome measure is students' perception of chess]

Gaudreau, L. (1992). *Etude comparative sur les apprentissages en mathematiques 5e annee*. Unpublished manuscript.

Palm, C. (1990). *New York City schools chess program*. Unpublished manuscript.

Smith, J. P. & Cage, B. N. (2000). The effects of chess instruction on the mathematics achievements of southern, rural, black secondary students. *Research in the Schools, 7*(1), 19-26.