[To appear in T. Redman (in press), Education and chess]

Educational benefits of chess instruction: A critical review

Fernand Gobet & Guillermo Campitelli

University of Nottingham

Address correspondence to:

Dr. Fernand Gobet

ESRC Centre for Research in Development, Instruction and Training

School of Psychology

University of Nottingham

Nottingham NG7 2RD

United Kingdom

email: frg@psyc.nott.ac.uk

+ 44 (115) 951 5402 (phone)

+ 44 (115) 951 5324 (fax)

1 Introduction

"Chess playing makes kids smarter." "Chess increases mathematical abilities." "Chess improves academic performance." Numerous similar claims have been made about the efficacy of using chess to foster education (see, for example, several papers on the USCF site for education).[1] Indeed, schools in various countries (e.g., USA, France, Argentina) offer chess as an optional subject, and some even propose compulsory classes. There is clearly a strong interest worldwide in the potential advantages of chess in education, and the conference from which this book stems is just another example of this interest.

Implicit in all these activities is the belief that skills acquired playing chess can transfer to other domains. Is this belief based on well-substantiated evidence? Is the educational value of chess a well-established empirical fact? Or have chess players been blinded by their love of the game into thinking that it offers instructional advantages? In this chapter, we attempt, as objectively as possible, to tackle the question of whether chess is advantageous for general education. To do so, we subject research into the educational benefits of chess to the same rigorous criteria commonly used in academia for evaluating educational research.

First, we briefly review how psychology and education deal with the question of transfer of skills from one area of knowledge to another. We then look at whether chessplayers as a group differ from the population at large—if playing chess promotes transferable skills, some differences should be visible, for example in IQ, visuo-spatial abilities, or planning behavior. Further, we consider what type of experiment should ideally be carried out to establish the presence of transfer. Having dealt with these preliminaries, we will be in a position to discuss the results from previous experimental studies that attempted to establish transfer from chess abilities to general abilities. We shall see that, in contrast to the strong claims often found in chess literature, the evidence is rather inconclusive, and, most importantly, somewhat rare.[2]

---

[1] http://www.uschess.org/scholastic/sc-research.html

[2] Given these negative conclusions, it may be appropriate to state that the authors do not suffer from disinterest nor have an axe to grind with regard to chess. The first author, an international master, has taught chess for several years in high school. The second author, an expert, has also been active in scholastic chess and in coaching top-level Argentinean players.

We end this chapter with a few proposals on how research into the role of chess in education may be improved in the future.

1.1 The question of transfer

The question addressed in this chapter can be summarized as follows: Can a set of skills acquired in a specific domain (in our case, chess) generalize to other domains (e.g., mathematics, reading) or to general abilities (e.g., reasoning, memory)? This is an old question, which, for a long time, was answered positively; for example, for centuries, it was accepted without dispute that learning Latin or geometry would train the mind and prepare it to cope with other topics. However, when, for the first time at the beginning of the 20th century, the question was studied scientifically, the conclusions were rather different. Thorndike's "identical element theory" proposed that there would be transfer from one domain to another only to the extent that there is overlap between the components of both skills (Thorndike & Woodworth, 1901). For example, geometry would be useful in studying higher mathematics, as some geometrical concepts are used in, say, calculus, but would not help the study of history. Overall, empirical evidence supports the idea that transfer is a function of the degree to which the tasks share cognitive elements (Anderson, 1990; Singley & Anderson, 1989; Travers, 1978), which suggests that transfer from one task to another is often limited.

A different view of transfer emerges from the psychological study of intelligence. Researchers in this field believe that one or a few transferable abilities form the basis of intelligence. These abilities are seen as general, at least within verbal or visuo-spatial domains, and are supposed to apply to a variety of domains (see Sternberg, 2000, for an overview). However, these basic abilities are also seen as innate, and thus not amenable to improvement through practice. Finally, other researchers have recently proposed that the best way to train transferable abilities is to teach generic skills, such as strategies for learning, methods for solving problems, and techniques for reasoning. This approach has achieved some limited success (Grotzer & Perkins, 2000).

In spite of these disagreements about the nature of transfer, some results are clear. In particular, recent research into expertise has clearly indicated that, the higher the level of expertise in a domain, the more limited the transfer will be (Ericsson & Charness, 1994). Moreover, reaching a high level of skill in domains such as chess, music or mathematics requires large amounts of practice to acquire the domain-specific knowledge which determines expert performance (Bloom, 1985, Simon & Chase, 1973; de Groot & Gobet, 1996; Ericsson, Krampe & Tesch-Römer, 1993; Saariluoma, 1995). Inevitably, the time spent in developing such skills will impair the acquisition of other skills.

1.2 The psychology of chess skill

Apart from educational interest, a large amount of research has been carried out into the psychology of chess playing, and quite a lot is known about the cognitive underpinnings of chess ability (Charness, 1992; Gobet, De Voogt & Retschitzki, in press). In general, the empirical evidence suggests that practice, as opposed to sheer talent, is essential for high levels of performance. Expertise in chess requires the acquisition of specialized knowledge, including memory of a large number of chess-specific patterns that may elicit appropriate moves, evaluations, or plans (Simon & Chase, 1973; Gobet & Simon, 1996). It also requires the ability to search through the space of possible positions effectively; this ability is to a large extent made possible by knowledge, which enables players to selectively generate and accurately evaluate chess positions (Simon & Chase, 1973; Gobet & Simon, 1998). Computational models of chess expertise can reproduce the perceptual and memory development of skill from novice to master (De Groot & Gobet, 1996; Gobet & Simon, 1996, 2000); typically, these models emphasize the role of practice and study for acquiring the large number of domain-specific perceptual patterns (more than 100,000) necessary for expert performance. As mentioned above, specificity of knowledge points to difficulties in transfer.

That chess skill is specific has been nicely illustrated in studies with children. Chi (1978) as well as Schneider, Gruber, Gold and Opwis (1993) examined the role of chess expertise in memory recall in both children and adults, and both within and outside the domain of expertise. They found that chessplaying children vastly outperformed non-playing adults in the recall of briefly-presented chess positions, but that, in all cases, adults were much better at remembering lists of digits. This was taken as evidence that chess skill shows little transfer to other domains, and certainly not to the memory of digits.

Thus, although there is still some debate about the question of transfer of skills, the "null hypothesis" for many psychologists, educationalists, and policy makers in education, and also the default position taken in this chapter, is that transfer is minimal. This is only a hypothesis, open to refutation by the empirical data. A first source of evidence is to consider whether the chess population as a group differs from the general population.[3]

1.3 Do chess players have special cognitive abilities?

If chess has a positive influence on intelligence, one should expect chessplayers to be more intelligent than the general population. Studies aimed at answering this question have produced mixed results. In an early study, Djakow, Petrowski, and Rudik (1927), who studied a group of eight of the best grandmasters of the time, did not find differences with a control sample on general intelligence or visuo-spatial memory. The only exceptions were recall tasks where the material was related to chess. Grandmasters were slightly better in memorizing moving spots in an 8x8 matrix, and showed a clear superiority when chess pieces were used. More encouraging results were found by Doll and Mayr (1987), who compared chess masters with a non-chessplaying control group on the "Berlin Structural Model of Intelligence" test.[4] They found that the masters performed significantly better on general intelligence and in tasks related to "information processing capacity for complex information," "working speed," and "numerical thinking"; however, they did not do better in a visuo-spatial task. Similar results were found with children. Frydman and Lynn (1992) studied the mental abilities of young Belgian chess players (about 11 years old) using the French version of the Wechsler Intelligence Scale for Children, a widely-used IQ test consisting of a verbal and a performance scale. They found that their sample had a higher general IQ than the population mean, as well as a higher performance IQ and a higher verbal IQ (the performance IQ was higher than the verbal IQ). Finally, the stronger players had higher performance IQ scores than the

---

[3] Obviously, any effect potentially useful for education should be visible even at low levels of skill. Positive transfer that could be observed only at the master level, say, would be of little practical interest.

[4] For additional information about most of the tests mentioned in this chapter, the reader is referred to textbooks such as Cronbach (1960) or Sternberg (2000).

weaker players. In another study with children, Horgan and Morgan (1990) found that the best chessplayers in their sample (mean age around 11) scored higher than the age-relevant norms on the Raven's Progressive Matrices (an intelligence test measuring reasoning and "pure" intelligence) and on the Piagetian plant task (a task aimed at measuring children's ability to use combinatoric logic in formal operations; Kuhn & Brannock, 1977). In a different study, Horgan (1992) was interested in how people, given a sequence of hypothetical results in a chess tournament, could predict their performance against new opponents. She found that chessplaying children were better at this task than non-chessplaying parents or statistics students. This ability transferred to making predictions in a non-chess domain (tennis), which lay outside the domain of expertise of these children.

Given that chess is a visuo-spatial skill, a plausible hypothesis is that its practice helps develop visuo-spatial abilities. However, the existing data are not conclusive. As we have just seen, Frydman and Lynn (1992) found a correlation between skill at chess and performance IQ. As this scale includes measures of visuo-spatial abilities, this result suggests that high-level chess playing may require strong visuo-spatial abilities. This possibility has to be qualified by the fact that performance IQ includes measures not related to visuo-spatial ability, such as alertness to essential detail, concentration, logical thinking, and the ability to work under time pressure. In addition, other studies have reached a different conclusion. An unpublished study by Lane (mentioned in Cranberg & Albert, 1988, p. 161), did not find any reliable correlation between chess skill and performance on a visuo-spatial task (the Guilford-Zimmerman Spatial Visualization Subtest). Lane used a sample of players from novices to strong amateurs, which did not include masters. Similarly, Waters et al. (in press) found no evidence for a correlation between chess skill and visual memory ability in a group of British adult chess players, which ranged from Class D players to grandmasters. Moreover, chessplayers did not differ from non-chessplayers.

In sum, the available empirical evidence suggests that chessplayers tend to be more intelligent than non-chessplayers, and that, at least with children, there is some correlation between chess skill and intelligence. Surprisingly, there is little to support a link between chess skill and visuo-spatial ability, as assessed by psychometric tests. All these results are based on correlational data, which makes any conclusion about the direction of causality highly tentative. A possibility is of course that practicing chess develops intellectual ability. However, these results could also be explained by selection processes: smarter people are more likely to choose, and to excel in, an intellectual activity such as chess. Finally, there is an indefinite number of more complex causal models accounting for these correlations, for example: both chess skill and IQ are improved by higher motivation. In order to settle the question, we need more powerful experimental designs, where the presence or absence of chess instruction is directly manipulated. This will be the topic of the next section.

2 The ideal experiment

Education science, as well as psychology and medicine, has developed a variety of techniques for establishing whether a given treatment (in our case, chess instruction) positively affects some behavioral target, such as school performance, cognitive ability, or attitude to school (e.g., Travers, 1978; Keppel, 1982). At the very least, the treatment group must be compared to a control group, and possible differences evaluated by a posttest measuring the variable(s) of interest. However, it has been well established in scientific research that simply belonging to a treatment group may affect behavior (the "placebo effect"). A stronger design is therefore to use two control groups; the first (the placebo group) receives an alternative treatment, and the second is given no treatment at all. If the treatment group, but not the placebo group, shows improvement, it can be concluded that the effect is specific to some features of the treatment group, and not due to unspecific factors such as participation in an experiment.

However, the designs discussed so far are not enough to establish that the treatment causally affects behavior. One must also show that there were no differences at the outset. This can be done by randomly allocating participants to treatment and control groups. An alternative way, which can be beneficially combined with the use of randomization, is to submit the participants to a pretest at the beginning of the experiment. This pretest, which can include several tasks, should measure the same variables as those measured during the posttest.

Several additional precautions should be taken. The participants should be blind to the goal of the experiment, or, even better, to the fact that they belong to an experiment, as knowing that one is participating in an experiment may change one's behavior, irrespective of the treatment received. The same applies to the experimenter. Another precaution is to have different people carry out the treatment, the pretest, and posttest, to avoid knowledge of previous results or group assignment contaminating data collection.

To summarize, the “ideal experiment” should meet the following requirements: random assignment of the participants to the various groups; presence of a pretest to insure that there is no initial difference between the groups; presence of a posttest to measure potential differences due to the treatment; presence of an experimental group and of two control groups, one for eliminating the possibility of a placebo effect; provision of different people for carrying out the treatment, the pretest, and posttest; experimenter’s and tester’s ignorance of the nature of the group assignment; and participants’ ignorance of the purpose of the experiment, and even of the fact that they are participating in a study.

Unfortunately, this ideal experiment is difficult to conduct, for a number of practical, administrative, and ethical reasons. For example, school authorities and parents may object to the random assignment of participants; it is difficult to conceal from the participants that they belong to a special group; and, the presence of a pretest may cue participants to the fact that they belong to an experiment. Indeed, a large number of studies in education use a weaker version of the ideal experiment, where one has to satisfy oneself with a much simplified form of investigation, called the “quasi-experiment.” In this case, one does not manipulate the assignment of participants to a given group, but one uses groups already formed (e.g., children attending a chess club vs. children not attending). The problem with quasi-experiments is that the conclusions which can be drawn are severely limited, mainly because the direction of causality is no longer under the experimenter’s control (e.g., Travers, 1978).[5] This theme will recur in the analysis of several studies below. For example, an experiment finds that the chess group is more intelligent than the control group. Is it because of chess instruction? Is it due to the fact that intelligent children are more likely to play chess? Or, is it due to a third variable? For instance, it is reasonable to assume that children coping well with time pressure play chess better, as thinking time is limited by a clock, and obtain higher scores in intelligence tests, as some of the subtests are carried out under time restrictions.

---

[5] Statistical techniques exist that allow one to infer causal structures from quasi-experimental designs (e.g., Glymour et al., 1987). These techniques typically require very large samples and a substantial number of variables. We will take up this topic in the Discussion.

3 Education and chess: Empirical evidence

We are now in a position to evaluate research aimed at demonstrating the general benefits of chess instruction. Before discussing the results, however, we must briefly describe the methods we used to select and analyze the studies.

3.1 Methods of selection

3.1.1 Selection criteria

We compiled a list of publications from several sources: material from the USCF site on research into scholastic chess, previous reviews (Bönsch, 1987; Dextreit & Engle, 1981), scientific databases (Web of Science, PsycInfo), and our own archives.[6] The main criteria for selection were as follows:

- The publication contains an *empirical* investigation into the effect of chess instruction upon some target ability or behavior (such as intelligence, attitude towards school, reading performance, etc.);
- The putative effects were measured in an objective way; and
- At least a modicum of detail was given about the methodology used.

These criteria eliminated most articles, which were either reviews of past work, descriptions of teaching methods, observation studies reporting anecdotes, and statements of opinions. We also decided to perform our analysis on the available publications, without contacting the authors for additional detail. Although this would have led to useful clarifications in several studies, we took this decision because (a) this would have been extremely time consuming; (b) several authors may not be accessible anymore; and, most importantly, (c) school authorities or decision-making bodies on education are likely to base their judgments only on the information available.

3.1.2 Evaluation criteria

Our main interest in this review involves the methodology used, and the extent to which it justifies claims of transfer from chess to other domains. With a few

[6] It was decided not to include the material contained in this book, unless it had been published previously.

exceptions, we will not say much about the statistical techniques used to analyze the results, which were generally appropriate. When dealing with statistical tests of significance, we will use the level of $p < .05$ as the minimum requirement for a statistically significant result (this is current practice in the behavioral sciences), though more lenient levels were used in some of the studies (e.g., $p < .10$).

3.2 Description of the main studies

We now present the experiments that passed our selection criteria. A summary is given in Table 1, which compares the selected studies with the ideal experiment along seven criteria, and in Table 2, which provides a list of the psychological and educational measures used. The first and most disappointing finding was that we could identify only a handful of studies. In particular, we were unable to find scientific reports about well-publicized chess experiments, such as that conducted between 1976 and 1979 in Switzerland and that conducted between 1980 and 1984 in Venezuela (for informal reports, see Dextreit & Engel, 1981, and Ferguson, undated-a, respectively).

3.2.1 Chess and cognitive development; Christiaen (1976), Christiaen & Verhofstadt-Denève (1981)

The goal of this study was to investigate the effect of chess instruction on children's cognitive development, and more specifically, on the appearance of the stages as described by Piaget's theory (see Flavell, 1963, for an introduction). In particular, the interest lay in the transition between the stages of "concrete operational thought" and the next stage, called the "formal-operation" stage. According to Piaget's theory, this transition occurs at around 14 to 15 years of age.

	Random allocation	Pretest	Posttest	Control group I (placebo)	Control group II (do-nothing)	Different teacher and tester	Subjects blind to the experiment
Ideal experiment	√	√	√	√	√	√	√
Christiaen (1976)	√		√		√	√	√
Frank (1979)	√	√	√		√	√	?
Liptrap (1998)		√	√		√	√	√
Ferguson 1		√	√		√	?	?
Ferguson 2		√	√		√	?	?
Margulies		√	√		√	√	√
Fried & Ginsburg	√		√	√	√	√	?

Table 1. Comparison of the experimental design used in the seven selected studies with that of the ideal experiment.

Christiaen used a "posttest-only control group" design. Twenty fifth-grade students (average age at the beginning of the experiment: 10 years 7 months) from two classes of a Belgian boys school were randomly assigned to the chess group, and 20 to the non-chess group. The chess group received one hour a week of chess instruction on Fridays after school, while the control group did not perform any activity and simply went home; this lasted for 42 weeks during term, spread over one and a half years. The chess course was made compulsory by the teacher and the school management. Chess instruction consisted of theory, games, and tournaments. No pretest was given—to prevent the children suspecting that they were part of an experiment. The main posttests consisted of two standard Piagetian tests (the balance-beam test and the liquid test). The study also took advantage of the fact that a series of aptitude tests for orientation purposes were given annually to the children of the sixth grade, and one of these tests (the "PMS") was used as a dependent variable. The school results at the end of the year were also considered as dependent variables. Finally, chess level in the treatment group was evaluated by a seven-round tournament played using the Swiss system.

Study	Measures
Christiaen (1976)	Balance test (Piaget) Liquid test (Piaget) PMS (aptitude tests for orientation purposes) School results
Frank (1979)	PMA (Primary Mental Abilities test) DAT (Differential Aptitude Tests) GATB (General Aptitude Tests Battery) D2 test of Brieckenkamp (test of attention) Rorschach test (projective test)
Liptrap (1998)	TAAS (Texas Assessment of Academic Skills)
Ferguson 1	CTA (Watson-Glasser Critical Thinking Appraisal) Torrance tests of creative thinking
Ferguson 2	TCS (Test of Cognitive Skills; Memory and Verbal Reasoning subtests
Margulies	DRP (Degree of Reading Power test)
Fried & Ginsburg	WISC-R (Wechsler Intelligence scale – Revised) Block design subtest Picture completion subtest Survey of school attitudes

Table 2. Psychological and educational measures used in the seven studies under review.

No reliable effect was found in either of the two Piagetian tasks, nor in any of the 5 subtests of the PMS test or in its aggregate score, although, in all cases, the chess group did better than the control group (one subtest and the aggregate score were marginally significant at $p < .10$). Statistically significant effects were found for the school scores, both after 5 months of chess instruction, and at the end of the sixth grade.

This study has several good features: random allocation of children to treatment and control groups; the fact that the tests were administered by people other than the researcher teaching chess; and the number and variety of tests used.

However, some weaknesses must also be mentioned. The lack of a second control group raises the possibility that any observed differences are due to placebo effects. Given the lack of a pretest, it is possible that there were differences between the two groups at the outset of the experiment. In his cautious conclusions, Christiaen (1976, p. 61) mentions another possible source of contamination: “However, we must take into consideration a possible influence of the teachers, who were aware of the trial and thus consciously or subconsciously could act favorably or unfavorably in their relation with the pupils.”

3.2.2 Chess and aptitudes; Frank and D’Hondt (1979), Frank (1981)

In a study conducted in Zaire (Frank & D’Hondt, 1979; Frank, 1981), 92 teenagers (16 to 18 years old) were randomly allocated either to a compulsory chess group or to a control group. The chess group met for an hour twice a week over a one-year period. Instruction included lectures, tests, simultaneous games, and practice. The control group did not perform any special activity. Two psychometric tests (the Primary Mental Abilities Test, and the General Aptitude Tests Battery), totaling 12 subtests, were administered, both before and after the intervention. Three other tests were given only before the intervention: the Differential Aptitude Test, the D2 test (a test of attention), totaling together 6 subtests, and the Rorschach test (a projective test using inkblots). For the experimental group, chess skill at the end of the year was estimated using chess tests.

The study aimed at testing two hypotheses. First, efficacy in learning chess is a function of several cognitive aptitudes, including spatial ability, perceptual ability, reasoning, creativity, and general intelligence. Second, learning chess in turn influences the development of these aptitudes.

In line with the first hypothesis, some of the sub-tests were found to correlate with chess skill after one year of instruction (“spatial aptitude” and “numeric ability” from the Primary Mental Abilities test; “administrative sense” and “numeric aptitude” from the General Aptitude Tests Battery; and “office work” from the Differential Aptitude test). No reliable correlation was found with the variables extracted from the Rorschach test.

With respect to the second hypothesis, a comparison of the scores on the posttest showed that the chess group performed better than the control group on “numerical aptitude” and “verbal ability.” Thus, it would seem that skills acquired during chess instruction generalized to other domains. However, the interpretation of the results for "numerical aptitude" is complicated by two factors. First, as we have just seen, this measure predicts chess skill. To address this possible confound, we conducted additional analysis on the posttest, using the same variable in the pretest to statistically control for the variability present before the experiment (a technique called "analysis of covariance"; e.g., Keppel, 1982). Even after this adjustment, the chess group outperformed the control group. So, in this respect, the result is robust. The second factor is more worrying. So far, we have dealt with the difference between the two groups in the posttest, but the real question is whether chess instruction *improves* performance. In fact, the progress of the chess group for "numerical aptitude" was small (from 10.4 to 12.2), and most of the differential effect between the two groups is explained by the fact that the control group, for unknown reasons, *dropped* from 12.0 in the pretest to 8.9 in the posttest. Additional statistical analysis showed that the chess group did not improve reliably for this measure. Thus, we conclude that only "verbal ability" was reliably enhanced by chess instruction.

In summary, this study has several strengths, including: random allocation of students to the experimental and control group; use of pretest and posttest with a number of standardized measures; and presence of different persons doing the teaching and testing. However, several weaknesses should also be mentioned: there was only one "non-treatment" control group, and no correction was used for the number of statistical tests carried out (see the Discussion section for more detail). Moreover, it could be argued that most of the tests used were invalid, as they were not designed for an African culture (e.g., Cronbach, 1960); though, it is unclear specifically how this last point would have affected the hypotheses tested by Frank and D'Hondt. Finally, Frank and D'Hondt mention two additional weaknesses which may limit the generalizability of their study: first, the lack of motivation and interest of most students, and, second, the low test results obtained. In particular, the lack of motivation suggests caution before organizing compulsory chess classes.

3.2.3 Chess and standard test scores; Liptrap (1998)

The aim of this study was to investigate the extent to which elementary students' participation in a chess club affects their standardized test scores. The study, which was conducted in four elementary schools in a large suburban school district (serving middle-class and affluent neighborhoods) near Houston, Texas, included 571 students, and compared their third and fifth grade scores on the Texas Assessment of Academic Skills (TAAS). An advantage of this method of assessment is that a related measure, the Texas Learning Index (TLI), makes it possible to compare students across years and across grades. The comparison was between students who participated in a school chess club in fourth and/or fifth grade and students who did not. The chess sample included 67 students (74.6% male), and the non-chess sample 504 students (50.8% male). The samples were further divided into special-education students, regular students, academically-able students, and gifted-and-talented students. Both for the chess and control group, most of the sample consisted of regular students (34.3% and 53.4%, respectively).

In the 3rd grade (that is, before chess instruction), the chess group was somewhat better (3.4 TLI points) than the control group with respect to reading, but not reliably so, and was marginally better for mathematics ($p = .051$). By the fifth grade, the chess group clearly outperformed the control group for reading ($p < .001$) and mathematics ($p < .005$). The differences were particularly clear with the sub-group of regular-track students: chessplayers scored 4.3 TLI points higher than non-chessplayers in reading ($p < .01$) and 6.4 points higher in mathematics ($p < .00001$). Both groups improved over the two years for the two scores, but the improvement was stronger for the chess group (about twice the improvement of non-chess players in both reading and mathematics), although no information is provided on the statistical reliability of this difference.

Strengths of the study include a large sample, pretest and posttest, and the fact that the chess teacher did not collect the test scores himself. Unfortunately, several weaknesses are also present. In particular, the non-random assignment of subjects to the two groups leaves open the possibility that the two groups were self-selected; this suspicion is supported by the fact that the two groups differed at the beginning (although not reliably so), the chess group being overall better in both mathematics and reading. Another weakness is that statistical tests are reported selectively: t-tests comparing the chess group and the non-chess groups are given for the regular students at the 3rd and 5th grades, and for the special-education students at the 5th grade, but not for the other subgroups or for the entire sample. This leads to the unfortunate (and perhaps incorrect) impression that the claims of the study are overstated. Finally, there were obvious differences between the two samples, which strongly suggests that some self-selection mechanism was present: the chess group was predominantly male, and 52.3% of the chess group belonged to the subgroups of academically-able students or gifted-and-talented students, as compared to 33.5% of the control group. Liptrap anticipated this criticism by noting that "the chess group is 3:1 male and has a higher percentage of AA and GT students. The objection that this constitutes a self selective elite group is answered by considering the four tracks separately." However, the data presented in his article are not sufficient to rule out the possibility of self selection.

3.2.4 Ferguson's studies

The studies carried out by Robert Ferguson are often cited. Here, we discuss two of them (a third study, the "Tri-State area school pilot study," cited in Ferguson (undated-a), did not provide enough detail to allow us to evaluate it).

3.2.4.1 Teaching the fourth "R" (Reasoning) through chess; (Ferguson, undated-b)[7]

This project, which took place from 1979 to 1983, aimed at providing stimulating experiences fostering the development of critical and creative thinking. Participants were gifted students (with an IQ equal to or higher than 130) in grades 7 through 9, in the Bradford (PA) area school district. They chose among a variety of special activities such as chess, dungeons and dragons, Olympics of mind, problem solving with computers, creative writing, and independent study. Each group met once a week for 32 weeks.

Participants were tested with alternate forms of the Watson-Glaser Critical Thinking Appraisal test (CTA) and of the Torrance test of creative thinking, both at the beginning and at the end of the year. Results for the CTA showed that the chess group significantly outperformed the non-chess groups ($p < .001$), the computer group ($p < .003$), and the non-participants ($p < .025$). With the Torrance test of creative thinking, the chess group showed statistically significant improvement in "fluency," "flexibility" and "originality" when they were compared to the population norms and the non-chess groups. There was also a significant difference in "fluency" and

[7] This study is referred to as "Developing critical and creative thinking through chess" in Ferguson (undated-a).

"originality" (but not in "flexibility") for the chess group compared to the computer group.

Ferguson used a pretest and posttest design and used more than one control group, each carrying out activities other than chess. However, this study has an important weakness that rules out any interpretation of the results in terms of the contribution of chess training to critical thinking and creativity: students switched activities either quarterly or semi-annually, and chess players participated in other activities as well. Therefore, we do not know whether the improvement is due to chess treatment or to the other treatments. Another limitation of this study is that it investigated a gifted population; hence, the claims cannot be generalized to the entire population of school students. Finally, the sample was rather small (15 students in the school chess club).

3.2.4.2 Developing of reasoning and memory through chess (Ferguson, undated-a)

This study looked at skills in Memory and Verbal reasoning. Fourteen 6th graders (9 boys and 5 girls; mean IQ=104.6) from a rural school in Pennsylvania, received chess lessons two or three times a week and played chess daily from September 1987 to May 1988. The design consisted of a pretest and posttest, using the "memory" and "verbal reasoning" subtests from the California Achievement Test battery. The differences between the posttest and the pretest were compared to the national norms. The chess group improved more than the general population in the "memory" subtest ($p < 0.001$). With "verbal reasoning," the chess group outperformed the general population only marginally ($p < 0.10$). Ferguson also reports tests about gains within the treatment group, although it is not clear how these results were computed.

Strengths of the study include a pretest and posttest design, and the use of well-standardized measures. However, there are also several weaknesses in this study. There was no random allocation of participants to groups, the sample was very small, and there was no control group carrying out an activity different than chess. Therefore, it is difficult to identify what are the causes of the improvement. For example, the results could be explained by the fact that the group was performing an intellectual task daily, in which case any intellectual task could have done as well as chess. An alternative explanation could draw on the motivation that students were receiving daily from their tutor.

3.2.5 The effect of chess on reading scores; (Margulies, undated)

This study was interested in changes in reading scores after chess instruction. Mid-elementary school children in the South Bronx, New York City, joined chess clubs at school. In the first year, they received chess instruction by chess masters; in the second year, instruction was enhanced by computer-supported chess activities. Participation was voluntary. All subjects had taken a "Degree of Reading Power Test" (DRP test) at the end of the school year and in the prior year (students who scored below the 10^{th} percentile were not included in the study). These results were used as pretest and posttest. The results of the two years were combined, yielding 53 subjects (it is unclear from the report whether some students participated in both years). The control groups consisted of the national norm for the same grade and the average student in the school district.

Students in the chess group made reliably greater improvements than the average student in the country and the average student in the school district, who both showed no gain. However, it also turned out that the chess group had higher entry-level scores, on average, which could point to a selection bias. To address this possible confound, Margulies compared the chess group with a non-chess control group made up of children with high entry-level reading scores, comparable to those of the chess group. Again, more pupils in the chess group showed gains than in the control group.

These results sound impressive, and it is fair to say that the study has several strengths (including: use of pretest and posttest; and tests carried out independently from chess instruction). Unfortunately, it also has several weaknesses. The use of a quasi-experimental design makes conclusions about the effect of the treatment highly tentative. This is acknowledged by Margulies (p. 10): "[…] chess participants form a pool of intellectually gifted and talented students. Students who join this group make contact with a core of high achievers and thereby develop more academic interests, speak at higher levels of standard American speech and take on the values of achievement. […] Our research does indicate that although some chessplayers began the year as poor readers, the chess program attracts a higher percentage of excellent readers than are found in the general District Nine population. This supports the possibility that chess participation does function as an Intellectually Gifted and Talented Program." Another weakness is that the effect of playing chess is confounded by the use of computers in the second year. Finally, the study did not use a control group engaged in another activity not related to chess. As a consequence, it is difficult to identify the reasons for the improvement. For example, we cannot rule out the possibility that the effects were caused by participation in an extra-curricular activity or involvement in an intellectual task.

3.2.6 The effect of learning to play chess on cognitive, perceptual, and emotional development in children (Fried & Ginsburg, undated)

This experiment was interested in the effects of chess instruction on the development of perceptual ability, visuo-spatial ability, and attitude towards school. The sample consisted of children with mild learning and behavior problems, referred for counseling by classroom teachers because of behavior problems. Thirty fourth and fifth graders (15 males and 15 females) from Brooklyn, New York, were randomly assigned to one of three groups: chess, counseling (which was used as a placebo group), and no-contact. Each subject was blind to the treatment condition. Chess instruction consisted of lectures, demonstrations and games. After 18 weeks, three tests were given: (a) the "figure completion" subtest of the revised version of the Wechsler Intelligence Scale for Children, measuring perceptual ability, and in particular visual awareness to detail; (b) the "block design" subtest of the same test, measuring visuo-spatial ability; and (c) a survey of school attitudes. No difference was found between the three groups. Additional analysis showed the presence of an interaction between gender and treatment in the block design task and the school attitude test. In the block design task, girls scored significantly lower than boys in the chess group, while no gender difference was seen in the other groups. In the survey of school attitudes, girls obtained higher scores than boys in the chess group, while there were no gender differences in the two other groups. These two interactions were not predicted by Fried and Ginsburg.

The strengths of this study are a clean design, with random assignment to the groups, an experimenter blind to the treatment condition for each subject, and the presence of two control groups, one controlling for placebo effects. A relatively minor weakness is the absence of a pretest, which is counterbalanced by the random group assignment.

3.3 Methodological remarks

As mentioned earlier, the design that allows us to be most confident about assigning effects to chess instruction is the "ideal experiment," which randomly assigns participants to treatment and control groups. For a variety of reasons discussed above, this type of experiment is very hard to conduct in education, and researchers often have to be content with weaker designs. The seven empirical studies reviewed in this section are no exception. While most of these studies reported some positive effect, all had methodological weaknesses that limit their generalizability to some extent.

Only three studies (Frank & D'Hondt, 1979; Christiaen & Verhofstadt-Denève, 1981; and Fried & Ginsburg, undated) used a design with random assignment. The other studies, which relied on quasi-experimental designs, used statistical techniques too weak to infer the causal structure of the variables under study. In particular, there were often no safeguards against the placebo effect (belonging to a group under study affects one's behavior) and the (non-conscious) influences of the teacher/tester.

In some cases, a large number of variables were examined, either in the pretest or the posttest, or in both. Just by chance, some of the statistical tests may have turned out to be significant. To control for the inflation of false positives (i.e., significant differences due to chance), it is necessary to incorporate statistical corrections for multiple tests. None of the studies included such corrections. The possibility of spurious results is clear in Frank and D'Hondt's (1979) and in Christiaen and Verhofstadt-Denève's (1981) studies, where a number of statistical tests were carried out on a number of different variables. In both cases, effects that were directly predicted (chess improves a number of aptitudes, including visuo-spatial abilities, in Frank and D'Hondt's study; and chess speeds up transition between Piaget's stages in Christiaen and Verhofstadt-Denève's study) were not observed. In the second case, non-predicted differences (positive influence of chess on school results) were observed. This pattern of results suggests the need for caution in interpreting the outcome of the experiment.

Three other methodological weaknesses should also be mentioned. First, no study has examined the long-term benefits of instruction. This is an important omission, as chess instruction would be of little value if its presumed benefits disappeared after one or two years. Second, the characteristics of the teacher were rarely, if ever, controlled. Typically, the teacher is a motivated chess player convinced that chessplaying has considerable benefits for children. A possible confound is that these special features, rather than chess itself, influence children. Finally, a very unsatisfactory feature of this research is the total lack of replication. None of the studies under review was repeated with the same experimental design and under similar conditions.

4 Discussion

4.1 Evaluation of past and current research

As mentioned earlier, research in psychology and education suggests that cognitive skills do not transfer well from one domain to another. Thus, the default position for most education experts will be that skills developed during chess study and practice will not transfer to other domains. Do the empirical data on chess research refute this position? Unfortunately, the answer is: no.

In our review of chessplayers' cognitive abilities, we presented some evidence that adult players are more intelligent than the average population. However, some specific predictions were not supported; in particular, that chessplayers should have better visuo-spatial skills. More importantly, these correlational data are in themselves insufficient to establish the causal role of chess practice on intelligence; as noted several times, such data are consistent with a variety of other explanations, including the possibility that chess selects more intelligent individuals. What is needed are studies where chess instruction is directly manipulated, in order to allow one to evaluate the specific effect of chess instruction. Our review uncovered three studies meeting this criterion, with a random assignment of participants to a chess group and to control group(s). Results only weakly supported transfer from chess instruction: in Frank and D'Hondt's (1979) experiment, only "verbal ability" was convincingly influenced by chess instruction; in Christiaen and Verhofstadt-Denève (1981), an effect was found only with school scores; however, Christiaen (1976) warns us of possible unwanted contamination from the teachers; finally, Fried and Ginsburg (undated) did not find any main effect of instruction.

We have already commented on the methodology used in these studies, and mentioned a number of weaknesses. This critical evaluation was qualified by the fact that a number of organizational, ethical, and practical constraints hamper research into education. Here, we give some more general considerations about the quality of the studies we have reviewed, using criteria commonly used in scientific research.

Many of these studies, and most of the studies that did not satisfy our selection criteria, did not give enough detail about the methods used and the results obtained. Such a level of detail would be necessary for enabling proper evaluation, making replication possible (a fundamental aspect of scientific research), or carrying out additional analysis (e.g., information about variability should be given by providing standard deviations). In this respect, the two published studies fare better, in part because the original research publication was available (PhD thesis for Frank, 1981, and "licence" (roughly equivalent to a Masters' thesis), for Christiaen, 1976). Indeed, out of the seven studies we reviewed, only two were published in peer-reviewed journals. Producing research that can pass the barrier of refereed journals is a *sine qua non* condition for gaining acceptance within the educational community. Furthermore, none of the studies clearly answers the question of "what is specific to chess instruction?", either because of weaknesses in the design or because no effect was found. A number of alternative explanations can be provided for many of the studies reviewed; two examples will suffice: any intervention would produce the same result (the placebo explanation); or, the motivation and talent of the teacher played the crucial role.

There is also a tendency to highlight the data supporting positive effects of chess instruction, while downplaying data supporting self-selection effects, the role of innate talents, or data that fail to support the hypothesis that chess instruction provides the expected benefits. While this tendency is understandable, given the goals of most researchers, it clearly backfires when the studies are read critically.

In most studies, no well-established theory was used to predict results or justify why chess should help; we believe that this weakens the case for chess instruction, mainly because the default position of most psychologists and educationalists is that transfer is unlikely. Typically, "common-sense" theories are used— a rather weak approach. For example, Frank (1981, p. 72), in an otherwise objectively reported study, concludes that "skill in chessplaying requires possession of a large number of aptitudes to a greater or lesser degree, but all of them necessary." This claim is often used in the literature as support for chess education. However, in its weak form, this is an empty statement, in the sense that most human activities are likely to tap multiple abilities (e.g., mathematics, sports, music), and, in its strong form, it is empirically incorrect (e.g., the results of Djakow et al., 1927, or even Frank's own data, where a number of abilities did not correlate with chess skill). Similarly, Margulies (undated) explains his result that reading scores are positively affected by chess instruction as follows: "Chess players combine high-level processes —knowledge and information about the position—and an interactive approach in which each 'candidate move' is considered much like a word or phrase in reading. The cognition processes are very similar. Both chess and reading are decision-making activities, and some transfer of training from one to the other may be expected (p. 11)." Again, the link is at best suggestive, and most human activities (even watching a movie!) would engage similar processes.

4.2 Recommendations for future research

We realize that our conclusions are likely to disappoint many chessplayers, in particular those who have invested considerable amounts of time and energy in promoting chess in schools, and those who have actually collected data about the effect of chess instruction. To counterbalance this negative impression, we would like to reiterate that some studies (Frank & D'Hondt, 1979; Christiaen & Verhofstadt-Denève, 1981; and Fried & Ginsburg, undated) were very well conducted, in spite of the enormous logistical difficulties that their authors are likely to have met. We would also like to provide some advice about how future research should be carried out. This is obviously not an exhaustive list.

Some specific recommendations follow directly from our analysis:

- More studies should be carried out, using a methodology as close as possible to what we have called the "ideal experiment."
- In particular, several factors should be better controlled, such as non-specific factors (placebo effects) and effects due to teachers' personality or style.

- It is necessary to carry out studies that reach the publication requirements of peer-reviewed journals, and, of course, to have the studies published in well-respected journals in education or psychology.
- Authors should avoid drawing conclusions that are biased or selective. In general, they should provide a more objective and less enthusiastic analysis of the data.

We have so far emphasized the ideal experiment, which would identify effects of chess instruction most crisply. However, there are important practical difficulties with this design, in particular with random assignment of participants to groups. With the growing concern over ethical issues in the behavioral sciences, in particular in the United States, it is likely that the realization of experiments using this design will meet with increasing difficulties. It may therefore be useful to point out alternative approaches, which, so far, have seldom been used in chess education research.

A first approach is to use statistical means, as opposed to direct manipulation as in the ideal experiment, to control for variability in group allocation. Several techniques, essentially based on the analysis of correlations, have been developed, such as structural equation modeling, causal modeling, Bayesian networks (e.g., Glymour et al., 1987; Tabachnik & Fidell, 1996). These techniques will provide a more detailed causal picture than the classical experiment, typically providing a network showing the causal links between several variables. For example, for chess and education, they could show the links between reading scores at time 1, level of motivation, parental support, presence or absence of chess practice, age, and reading scores at time 2. Unfortunately, these methods are harder to understand and use than those typically used in the experiments we have reviewed, and, to be reliable, require large samples (at least several hundred participants) and a number of variables, ideally measured over a long period of time.

A second approach is to carry out detailed analyses of what is being learnt during chess instruction, and to relate this newly acquired knowledge to characteristics of other domains, thus creating an empirical link between changes occurring during chess instruction and possible transfer to other domains. Ideally, this should be combined with a theoretical analysis of the cognitive processes involved. Such empirical and theoretical studies are badly needed if one wants to promote chess instruction as a means of developing general cognitive abilities, but such studies are also difficult to carry out. Singley and Anderson (1989) may be consulted for an example of such an analysis in the domain of computer programming.

These recommendations only cover studies aimed at investigating the specific question of transfer. Additional work may compare various methods of teaching chess (a neglected domain of research) or study the detail of the learning processes underlying the early stages of the acquisition of chess skill.

5 Conclusion

As shown in the documents collected by the USCF, chess teachers and chess masters are sanguine about the benefits of chess instruction, proposing that chess develops, among other things, general intelligence, ability to concentrate, ego strength, self-control, analytical skills, and reading skills. De Groot (1977) is more specific and has suggested that chess instruction may provide two types of gain: first, "low-level gains," such as improvement in concentration, learning to lose, learning that improvement comes with learning, or interest in school in underprivileged environments; and second, "high-level gains," such as increase in intelligence, creativity, and school performance. Our review indicates that research has mostly explored the possibility of high-level gains, and this, with mixed results.

As argued in this chapter, there is a huge chasm between the strong claims often found in chess literature and the rather inconclusive findings of a limited number of studies. The extant evidence seems to indicate that (a) the possible effects of optional chess instruction are still an open question; (b) compulsory instruction is not to be recommended, as it seems to lead to motivational problems; and (c) while chess instruction may be beneficial at the beginning, the benefits seem to decrease as chess skill improves, because of the amount of practice necessary and the specificity of the knowledge that is acquired.

This chapter has critically reviewed the extant literature, and has proposed avenues for further research. We hope that the somewhat negative conclusions we have reached will stimulate the next wave of empirical studies. While chess may not "make kids smarter," it may offer what De Groot calls "low-level gains" for our society, and it would be a pity not to exploit this opportunity.

6 Reference list

Anderson, J. R. (1990). *Cognitive psychology and its implications* (3rd ed.). New York: Freeman.

Bloom, B. S. (1985). *Developing talent in young people*. New York, NY: Ballantine Books.

Bönsch, E. (1987). *Schachlehre für Lehrende und Lernende* [Chess teaching for teachers and learners]. (2 ed.). Berlin: Sportverlag.

Charness, N. (1992). The impact of chess research on cognitive science. *Psychological Research, 54,* 4-9.

Chi, M. T. H. (1978). Knowledge structures and memory development. In R. S. Siegler (Ed.), *Children's thinking: What develops?* Hillsdale, N.J: Erlbaum.

** Christiaen, J. (1976). *Chess and cognitive development.* Unpublished Master's thesis, Gent, Belgium.

Christiaen, J., & Verhofstadt-Denève, L. (1981). Schaken en cognitieve ontwikkeling [Chess and cognitive development]. *Nederlands Tijdschrift voor de Psychologie, 36*, 561-582.

Cranberg, L., & Albert, M. L. (1988). The chess mind. In L. K. Obler & D. Fein (Eds.), *The exceptional brain. Neuropsychology of talent and special abilities*, pp. 156-190. New York: Guilford press.

Cronbach, L. J. (1960). *Essentials of psychology testing (3rd edition)*. New York: Harper International Editions.

** De Groot, A. D. (1977). Memorandum: Chess instruction in school? A few arguments and counterarguments. In H. Lyman, *Chess in the classroom. An answer to NIE* (1981).

De Groot, A. D., & Gobet, F. (1996). *Perception and memory in chess. Heuristics of the professional eye*. Assen: Van Gorcum.

Dextreit, J., & Engel, N. (1981). *Jeu d'échecs et sciences humaines* [Chess and human sciences]. Paris: Payot.

Djakow, I. N., Petrowski, N. W., & Rudik, P. A. (1927). *Psychologie des Schachspiels* [Psychology of the game of chess]. Berlin: de Gruyter.

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.

Ericsson, K. A., & Charness, N. (1994). Expert performance: Its structure and acquisition. *American Psychologist, 49*, 725-747.

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100,* 363-406.

** Ferguson, R., Jr. (undated-a). Chess in education: Research summary. A review of key chess research studies for the BMCC Chess in education "A wise move "conference.

** Ferguson, R., Jr. (undated-b). Teaching the fourth "R" (Reasoning) through chess.

Flavell, J. H. (1963). *The developmental psychology of Jean Piaget*. Princeton, NJ: Van Nostrand Company.

** Frank, A. (1981). *Chess and aptitudes* (English translation of Frank's unpublished PhD thesis in French). Saugus, MA: American Chess Foundation.

Frank, A., & D'Hondt, W. (1979). Aptitudes et apprentissage du jeu d'échecs au Zaire [Aptitudes and learning of the game of chess in Zaire]. *Psychopathologie Africaine, 15,* 81-98.

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

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

Glymour, C., Scheines, R., Spirtes, P., & Kelly, K. (1987). *Discovering causal structure. Artificial intelligence, philosophy of science and statistical modeling*. Orlando, FL: Academic Press.

Gobet, F., De Voogt, A., & Retschitzki, J. (in press). *Moves in mind*. Hove: Routledge/Psychological Press.

Gobet, F., & Simon, H. A. (1996). Templates in chess memory: A mechanism for recalling several boards. *Cognitive Psychology, 31,* 1-40.

Gobet, F., & Simon, H. A. (1998). Pattern recognition makes search possible: Comments on Holding (1992). *Psychological Research*, *61,* 204-208.

Gobet, F. & Simon, H. A. (2000). Five seconds or sixty? Presentation time in expert memory. *Cognitive Science, 24,* 651-682.

Grotzer, T. A., & Perkins, D. N. (2000). Teaching intelligence. In R. J. Sternberg (Ed.), *Handbook of Intelligence* (pp. 492-515). Cambridge: Cambridge University Press.

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

Horgan, D. D. (1992). Children and chess expertise: The role of calibration. *Psychological Research, 54*, 44-50.

Keppel, G. (1982). *Design and analysis: A researcher's handbook* (2nd Ed.). Englewood Cliffs, NJ: Prentice-Hall.

Kuhn, D., & Brannock, J. (1977). Development of the isolation of variables scheme in experimental and "natural experiment" contexts. *Developmental Psychology, 13*, 9-14.

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

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

Saariluoma, P. (1995). *Chess players' thinking: A cognitive psychological approach.* London: Routledge.

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

Simon, H. A., & Chase, W. G. (1973). Skill in chess. *American Scientist, 61*, 393-403.

Singley, M. K., & Anderson, J. R. (1989). *Transfer of cognitive skill*. Cambridge, MA: Harvard University Press.

Sternberg, R. J. (Ed.). (2000). *Handbook of Intelligence*. Cambridge: Cambridge University Press.

Tabachnik, B. G., & Fidell, L. S. (1996). *Using multivariate statistics*. New York, NY: HarperCollins.

Thorndike, E. L, & Woodworth, R. S. (1901). The influence of improvement in one mental function upon the efficiency of other functions. *Psychological Review, 9,* 374-382.

Travers, M. W. (1978). *An introduction to educational research.* New York, NY: Macmillan.
Waters, A. J., Gobet, F., & Leyden, G. (in press). Visuo-spatial abilities in chess players. *British Journal of Psychology*.

Note: Manuscripts marked with ** are available from the USCF (http://www.uschess.org/scholastic/sc-research.html)

7 Acknowledgments

We thank Dr. Ronald E. Batchelor, Dr. Peter C. Lane, and Dr. Julian M. Pine for useful comments on this chapter.