Teaching Children to Think Strategically: Results from a Randomized Experiment

Donald Green
Yale University

Dan Gendelman
Eshcolot/Mind Lab

November 15, 2004

Strategic reasoning may be defined as "the art of outdoing an adversary" in a competitive setting (Dixit and Nalebuff (1991, p.ix). This art may take a variety of cognitive forms. It may involve the exhaustive assessment of all possible courses of action or more superficial analysis based on rules of thumb, called heuristics (Chi et al. 1988). Teaching exhaustive assessment is matter of showing students how to recognize and evaluate all of the possible options available in a game. However, in complex strategic situations where decision-makers face time constraints, exhaustive search is impossible, and heuristic reasoning, a necessity. Performance in these situations is a matter of making efficient shortcuts, examining in depth a narrowed set of options that look most promising.

Identifying this narrowed set of options properly is a matter of acquiring, applying, and adjusting strategic principles. Proficient chess players, for example, know to avoid moving their rooks' pawns at the start of the game. Some players understand this as a hard and fast injunction, whereas others think of it as part of a broader injunction to build strength in the center of the board while leaving lines of pawn on the outside as a defensive wall behind which the king may hide later in the game. Players with a deeper understanding of chess know when to abandon this heuristic in favor of other strategic imperatives. As players confront novel strategic situations, the application of strategic principles becomes more tentative, and the principles themselves more subject to change (Xia 1998).

Teaching people to think strategically is thus a matter of showing them how to search efficiently for solutions and to adjust their operating principles to fit the strategic situation at hand. Whether these skills can be taught is an empirical question. The wide variance in proficiency in games like chess may be regarded as a function of irreducible individual differences in cognitive ability, or instead as a function of practice in the application of heuristics. This essay provides evidence that at least some heuristic knowledge can be made explicit and represented in a teachable form.

This research report departs from existing work on strategic reasoning in several ways. First, our investigation is grounded in randomized experimentation. We investigate strategic performance by randomly assigning subjects to different interventions and examine post treatment differences. Although this methodology is common in psychology, it is surprisingly rare in the large literature on strategy games, especially chess. Instead, that literature has been content to describe individual differences in cognitive style using non-experimental research methods (Chase and Simon 1973, Gobet 2001, Waters et al. 2002). While the cognitive differences between accomplished chess players and novices are suggestive, it is by no means clear that these differences cause varying levels of performance. Moreover, since these cognitive differences are often difficult if not impossible to change, such studies offer limited insight into the question of how one might improve strategic performance.

Second, the substantive focus of our research differs from other studies of intellectual performance. The small but influential corpus of research that evaluates interventions designed to improve intellectual acuity (e.g., Wright 1991) has largely overlooked strategic reasoning. To be sure, researchers have examined ways to instruct students in cognitive strategies (Blagg 1991, Costa 1991, Jones & Idol 1990, Nickerson et al. 1985, Segal et al 1985, Whimbey et al. 1975), and some have attempted to impart experts' modes of thinking (Schoenfeld & Hermann 1982, Schoenfeld et al. 1982), but researchers have seldom examined ways of improving subjects' ability to survey the range of logical possibilities or to make judicious choices between alternative courses of action in a competitive environment.

Finally, the present research pays special attention to the pedagogic value of heuristic reasoning in game playing. Much of the literature on game playing, by contrast, examines the extent to which game-playing improves with practice or familiarity (Horgan & Morgan 1992) or imparts substantive knowledge (Koran and McLaughlin 1990, Henderson et al. 2000) or improves cognitive skills (Smith and Cage 2000, Thompson 2001). Our work is closest to that of Tubau and Alonso (2003), who show that subjects made better strategic choices in counterintuitive game settings when encouraged to form a useful mental representation of the game before playing; in their study of the Monte Hall Dilemma, practice alone did not improve performance.

The present study attempts to fill this gap by evaluating the effectiveness of a school-based curriculum called The Mind Lab that purports to improve students' ability to reason strategically. The Mind Lab is a program that provides instructors and game-based teaching materials to elementary schools. The instructional program is designed to impart strategic principles by way of analogies to real-life situations. For example, when teaching children to reason through games that present complex sequencing problems, the lesson draws an analogy to a formidable journey that seems overwhelming unless it is broken down into a series of more manageable steps. The idea behind the analogies is to provide easy-to-remember heuristics that have meaning both in games and in life.

The central empirical question is whether the analogy-based approach used by the Mind Lab in fact improves strategic reasoning, as evidenced by improved performance in strategy games. Although performance in abstract logic games represents just one domain where strategic thinking manifests itself, the experimental paradigm used here represents an important first step toward more nuanced understanding of strategic thinking.

This essay is organized as follows. We begin by providing an overview of the experimental design. After describing the population under study and the experimental intervention, we present the statistical model used to estimate the treatment effects. Next, we present results showing that the pedagogic approach used in the Mind Lab significantly improves performance in abstract reasoning. Not only do students in the treatment group perform better than the control group in the game used for instructional purposes; they also perform better than the control group when later presented with a new game that involves somewhat different tactics. Data on effort, as distinct from performance, reveal no difference between treatment and control groups. Taken as a whole, these findings suggest that aspects of strategic reasoning can be imparted through classroom instruction.

Experimental Design and Analysis

Subjects. Students were drawn from 8 classroom groups from 5 schools in 3 Israeli cities. The schools were chosen so as to represent a broad socioeconomic cross-section of Israeli society. Students from the most affluent school reside in neighborhoods with average incomes of $36,000 per year. Students from the mid-level schools reside in neighborhoods with a mean income of $20,000. The low income neighborhoods have average annual incomes of $12,000. Students were drawn from grades 3-6. One classroom group was drawn from the high socioeconomic stratum, 3 from the middle, and 4 from the bottom. Of the 195 children in the study, 5 were in grade 3, 24 in grade 4, 140 in grade 5, and 26 in grade 6.

Intervention. Students in each classroom were given a computer-administered pre-test, which introduced them to the rules of the solitaire strategy game Rush Hour® and coached them through some practice examples. After this introductory period, students were presented with a series of 15 Rush Hour® puzzles and encouraged to solve as many as possible in the 30 minutes allotted.

Each classroom group of students who completed this pretest was randomly divided into treatment and control groups. The randomization of each classroom was checked to ensure that the resulting experimental assignments were uncorrelated with the pre-test. Random assignments that did not satisfy this criterion were discarded, and new random assignments were conducted. Taking all of the classes combined, 100 students were assigned to the treatment group, and 95 were assigned to the control group.

A week later, each classroom was revisited. This time, the students were exposed to different types of instruction. The treatment group was presented with a 20 minute lesson concerning a strategic principle relevant to the game Rush Hour®. The lesson is summarized in Appendix I. The control group, on the other hand, was presented only with a series of examples of Rush Hour® puzzles and solutions, with no discussion of strategic principles. Thus, the factor that distinguishes the treatment and control conditions is the lesson plan, not the absolute amount of time spent examining the game.

Due to the vagaries of student attendance, some of the students tested in the pre-test phase of the experiment were absent during the post-test. The number of observations drops from 195 to 179 (92%), 85 in the control group and 94 in the treatment group. Although the rate of attrition is slightly higher in the control group than the treatment group, the difference is non-significant using a 2-tailed Fisher's exact test (p=.30).

A week following the post-test, students were presented with a new strategy game, Lunar Lockout. Students in the treatment and control group were treated similarly during this follow-up session, with the only difference being that students in the treatment group were encouraged to recall and implement the thinking methods from the second meeting. Follow-up testing was conducted in 6 of the 8 classroom groups, causing a reduction in the number of cases to 62 in the control group and 71 in the treatment group.

Outcome measures. Because the games were played on the computer, data on the quality and quantity of play were easily gathered for each student during all three tests. The quality of play was gauged by the number of cards solved. For example, in the pre-test, students answered an average of 7.0 cards correctly. This average increased to 8.1 when students were presented with a new set of puzzles during the first post-test. The mean in the second post-test was 5.0, reflecting the unintended difficulty of the puzzles created for this exercise.

Another outcome measure is the number of puzzles that each student attempted. Since many students failed to complete puzzles that they attempted, these scores have higher means. The pre-test mean was 11.5; the first post-test, 12.1; and the second post-test, 10.8. These scores provide useful measures of the effort that students invested in these puzzles. Not surprisingly, the numbers of puzzles attempted and completed are correlated (pre-test r=.40, p < .01; first post-test r=.24, p < .01; second post-test r=.11, n.s.). However, this correlation remains sufficiently weak that the predictors of attempts and completes turn out to be different, as we shall see.

Estimation. Because randomization was performed within each school and subject to stratification on pre-test scores, the appropriate regression model is one that introduces pre-test scores and dummy variable for school. Let Y represent a vector of post-test scores. Let X denote a dummy variable scored 1 if the student was assigned to the treatment group. Let S represent an n x 7 matrix of dummy variables marking each school. Let P represent pre-test scores. Let U represent a vector of disturbances. The regression model is thus:

$$Y = a + Xb + Sc + Pd + u$$

The central hypothesis of this study concerns the parameter *b*: if the treatment improves test performance, *b* is positive. Thus, a one-tailed test will be used to gauge the statistical significance of the result against the null hypothesis that the treatment did nothing to improve scores. The same model applies to the second pre-test. The predictors in this model are the same. While it may be tempting to add results from the first post-test as covariates, this model could produce biased estimates of *b*, as the first post-test is a manifestation of the treatment.

Results

Table 1 shows the results of a regression of the first post-test scores on the treatment, controlling for school and pre-test scores. The key finding is that the estimate of *b* is substantial, amounting to approximately one-fifth of a standard deviation in the post-test score distribution. This treatment effect is also statistically significant (*b*=.562, SE=.309, p=.035). Evidently, the instruction provided to the treatment group improved their post-test performance.

Was this improvement due to increased understanding of Rush Hour tactics or greater motivation to solve puzzles? If the latter, we should see students in the treatment group attempting more puzzles than students in the control group. As it turns out, no such relationship exists. As shown in Table 1, assignment to the treatment group does nothing to predict the number of cards attempted during the first post-test. The treatment increased students' success rate, not the number of puzzles they tried to solve.

We have seen that teaching strategic principles can have an immediate effect on student performance, but what about its enduring effects? Does the treatment group continue to dominate the control group a week later, when the classes are presented with a new game? Table 2 shows that the enduring effects of the treatment are surprisingly powerful (b=1.215, SE=.355, $p < .01$). Note that 1.215 is two-thirds of the standard deviation of scores observed in the control group.

In order to gauge whether these results reflected the special characteristics of students who were present at both post-tests, we recalculated Table 1 for the same sample (n=122) and found that the results in Table 1 remain qualitatively unchanged. The loss of observations increases the standard error associated with the treatment effect at the first post-test, but this treatment effect was no larger for the group present at the second-post test than for the entire sample of observations. In fact, as Table 2 shows, the treatment effect on the first post-test was slightly smaller among those who attended all three sessions than for those who took the first post-test but not the second. Evidently, the powerful results for the second post-test are not attributable to the idiosyncrasies of those who participated in all three tests.

We do not find evidence of significant interactions between the treatment and either gender, grade, or socioeconomic status. Nor do we find significant differences in treatment effects between those who scored above or below the median in the pre-test. Our inability to detect statistically reliable interactions may simply reflect the limitations of sample size, but it remains interesting that the treatment seems to improve performance across the spectrum of talents reflected in the pre-test.

Apart from suggesting that the treatment improved puzzle-solving performance, what can the data tell us about the quality of the children's play when solving puzzles?

The computer program used to administer the puzzles also gathered data on the number of unforced moves – that is, wasted moves that brought the players no closer to a solution. We calculated the number of unforced moves per solved puzzle. Note that dividing by the number of solved puzzles focuses attention solely on the puzzles that each child was able to master successfully, as opposed to penalizing the students who flailed about unsuccessfully on most of the puzzles. Restricting attention to the children who took all three tests (n=122) reveals significant negative relationships between the treatment and the average number of wasted moves. The mean number of unforced moves per successful puzzle was 33.1 in the first post-test with a standard deviation of 23. However, regression reveals that the treatment group made 9.586 fewer unforced moves (SE=4.095) on the first post-test. This effect is significant at the .05 level. After switching from Rush Hour® to Lunar Lockout on the second post-test, the average number of unforced moves declined to 9.4 with a standard deviation of 6.9. Again, the treatment produced a significant decline in the number of unforced moves ($b=-2.883$, $SE=1.079$, $p< .01$).

Discussion

The data presented above indicate that the Mind Lab curriculum had three statistically robust effects on puzzle-solving performance. First, exposure to analogies illustrating strategic principles increased the puzzle-solving performance of children in the treatment group. Second, the treatment group was able to apply this lesson beyond the confines of a single game; the treatment effect was even more pronounced when the children in treatment and control groups confronted a new game. Third, when solving puzzles in both games, children in the treatment group showed clear signs of improved efficiency in their search for solutions.

The experiment is also notable for what it did not show. The treatment had no discernible effect on how much effort students invested in solving the puzzles. There was no apparent relationship, for example, between the treatment and the number of puzzles that students attempted. We interpret this pattern to mean that the Mind Lab intervention did not enhance performance through motivation. And by extension, the pattern implies that the limiting factor in solving these puzzles is not motivation but rather understanding.

Among their many intriguing implications, these results suggest that strategic acumen is not a fixed trait. Rather, even relatively brief interventions, such as a short presentation of an analogy, can have substantial effects on the facility with which children grapple with puzzles. This finding opens up a variety of research trajectories. How much larger do the effects become when the intervention becomes more time-intensive? How long-lasting are the effects? How far does strategic performance in one domain travel into other domains, such as interpersonal negotiation or academic performance?

Appendix I: Rush Hour® Lesson Synopsis

First lesson (both control and treatment groups):

The lesson starts with an introduction about the puzzle game "Rush Hour", including a graphical demonstration of the rules.

The name of the game is "Rush Hour."

The goal in this game is to free the red car out of the traffic jam, through the opening on the right side of the board.

Each exercise has a different starting position on the board. You may drive the cars and trucks forward and backward in the direction they are facing, trying to clear the way for the red car out of the jam.

Let's see an example:

We will move the green car upwards one square.

Now we will move the green truck 2 squares to the left.

The third move will be to drag the blue truck downwards

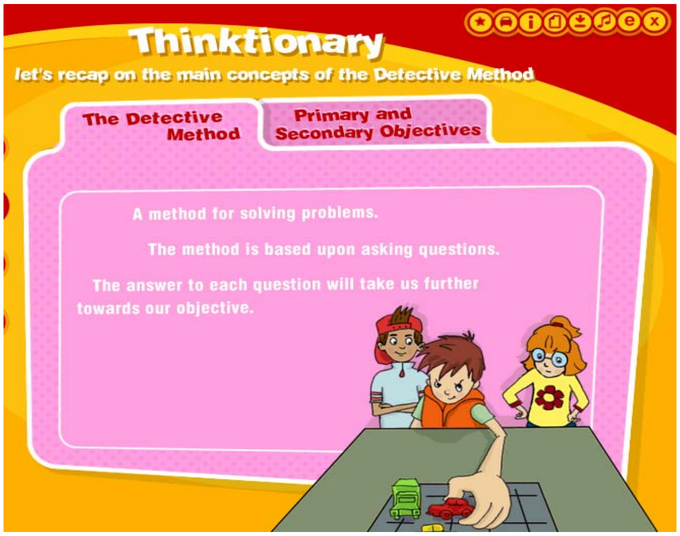
AHAA! Ready? You may now drive your way out to freedom!

Second Meeting:

- Treatment Group receives an oral presentation from a Mind Lab instructor covering the following material.

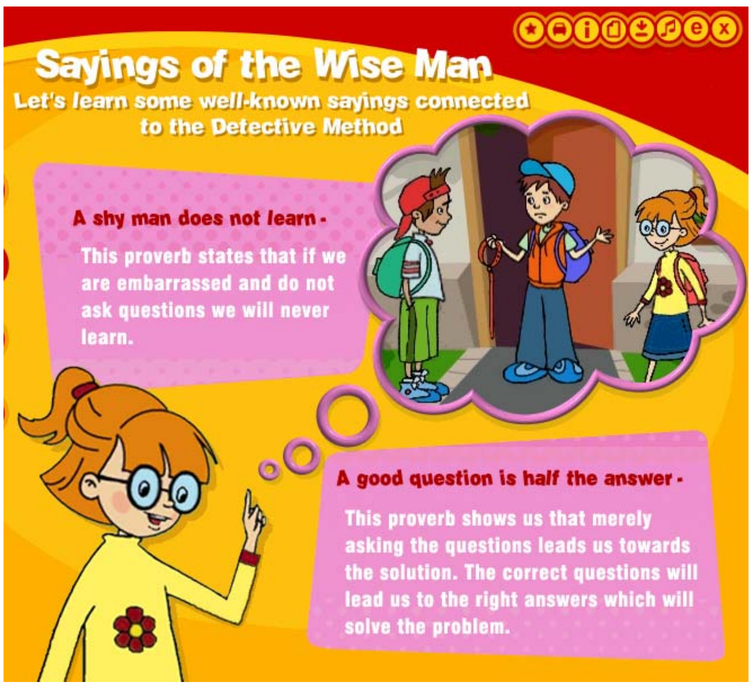
Two thinking methods:

1. The Detective Method – A method for solving problems. The method is based upon asking questions.

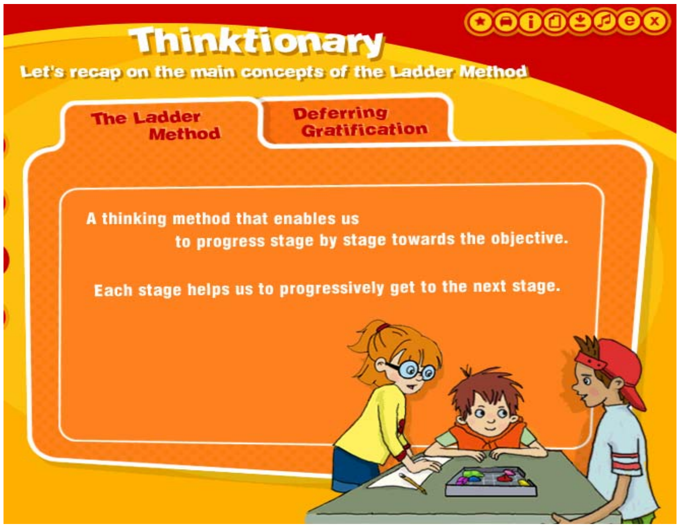


The children are taught two well known sayings that relate to the Detective Method. The sayings are designed to relate to their everyday life:

1. A shy man does not learn – This proverb states that if we are embarrassed and do not ask questions, we will never learn.
2. A good question is half the answer – This proverb shows us that merely asking the questions leads us towards the solution. The correct questions will lead us to the right answers, which will solve the problem.

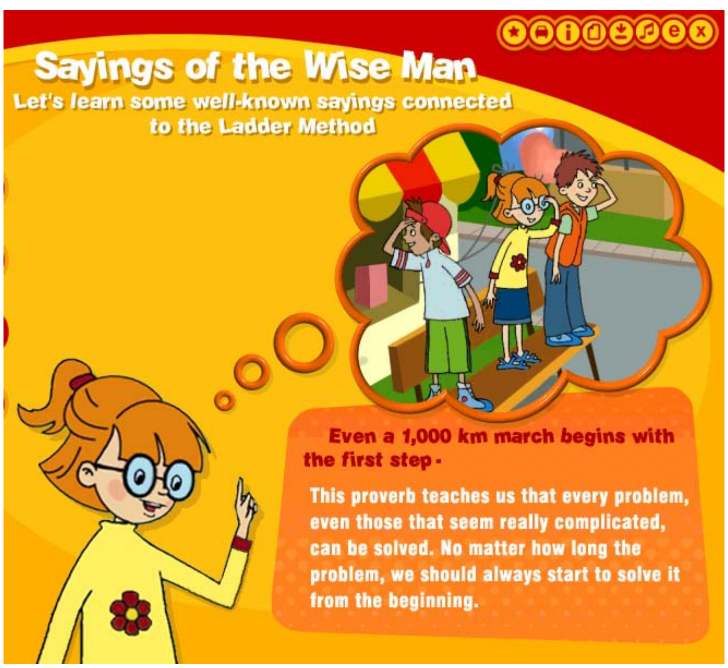


2. The Ladder Method – A thinking method that enables us to progress stage by stage towards the objective. Each stage helps us to progressively get to the next stage and complete our goal.



The children were also taught two well known sayings that relate to the Ladder Method:

1. Even a 1,000 kilometer march begins with the first step – This proverb teaches us that every problem, even those that seem really complicated, can be solved. No matter how long the problem, we should always not be intimidated; just start at the beginning and work through it gradually.



After completing this part of the lesson, the children begin the first post-test consisting of 15 exercises that are to be solved in 30 minutes.

- Control Group –

The activity of the control group included solving some general thinking riddles. Then two Rush Hour® exercises from previous meetings were reviewed and solved along with the children. Afterwards, children in the control group took the first post-test.

Third Meeting

- The treatment group activity included :

Teaching the rules of the Lunar Lockout Game
A review of the thinking models taught in the second meeting.

Solving together with the class an example exercise on the board using the thinking methods.

Cards Solving Time: 15 cards in 30 minutes

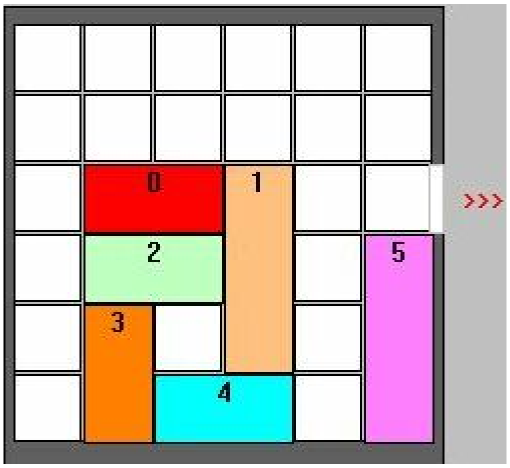
- The control group activity included:

Teaching the rules of the Lunar Lockout Game
Solving together with the class an example exercise on the board.

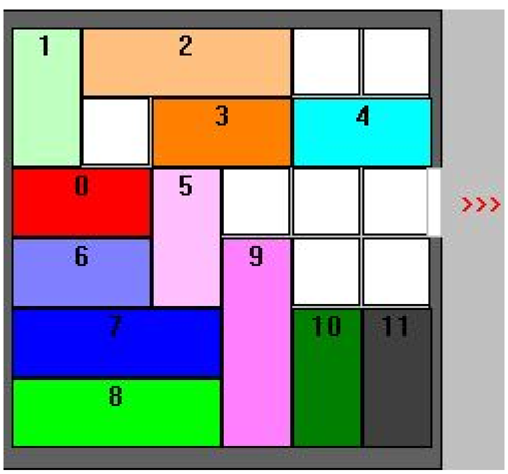
Cards Solving Time: 15 cards in 30 minutes

Appendix II: Example of the Two Puzzle Games: Rush Hour® and Lunar Lockout

Card 2:

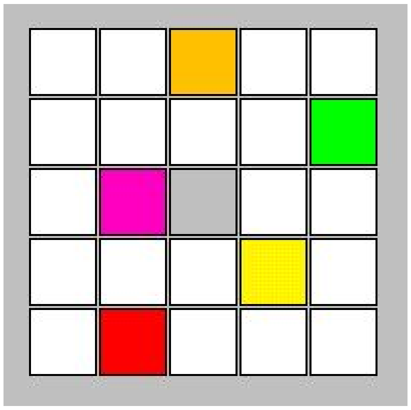


Card 12:

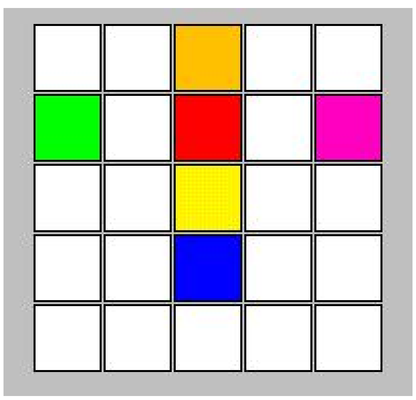


Example of Lunar Lockout cards:

Card 1



Card 12



References

Blagg, N .(1991) *Can we teach intelligence? A comprehensive evaluation of Feuerstein's Instructional enrichment program*. Hillsdale, NJ: LEA.

Chase,W.C. & Simon, H.A. (1973). Perception in chess. *Cognitive psychology* , 4, 55-81.

Chi, M.T.H., R. Glaser, and M.J. Farr. (1988). *The nature of expertise*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Costa, A. (Ed.). (1991). *Developing minds: programs for teaching thinking* (revised ed., vol. 2. Alexandria, VA: ASCD.

Dixit, A., & B. Nalebuff. 1991. *Thinking Strategically: The Competitive Edge in Business, Politics, and Everyday Life*. New York: W.W. Norton and Company.

Gobet, F. (2001) "Chess expertise." In *International encyclopedia of the social and behavioral sciences* (N.J. Smelser & P.B. Baltes, eds.). New York: Elsevier.

Horgan, D. D.; Morgan, D. (1990). Chess expertise in children. *Applied cognitive psychology*. Vol 4(2), 109-128.

Jones, B.F., & Lorna, I. (1990). *Dimensions of thinking and cognitive instruction*. Hillsdale, NJ: Erlbaum

Koran, L. J; McLaughlin, T. F. (1990). Games or drill: Increasing the multiplication skills of students. *Journal of instructional psychology*. Vol 17(4), 222-230.

Nickerson, R., Perkins, D. N., & Smith, E. (1985). *The teaching of thinking*. Hillsdale, NJ: Erlbaum.

Schoenfeld, A. H., & Herrmann, D. J. (1982). Problem perception and knowledge structure in expert and novice mathematical problem solvers. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 8:484-494.

Schoenfeld, A. H. (1982). Some thoughts on problem-solving research and mathematics education. In F. K. Lester & J. Garofalo (Eds.), *Mathematical problem solving: Issues in research* (pp. 27-37). Philadelphia: Franklin Institute Press

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

Tubau, E., & Alonso, D. (2003). Overcoming illusory inferences in a probabilistic counterintuitive problem: The role of explicit representations. *Memory & Cognition*. 31(4), 596-607.

Waters, A.J., Gobet, F., & Leyden, G. (2002). Visuospatial abilities of chess players. *British Journal of Psychology*. 93(4), 557-565.

Whimbey, A. (1975). *Intelligence can be taught*. New York; E. P. Dutton.

Wright, E.D. (1991). Odyssey: A curriculum for thinking In Costa, A. (Ed.). *Developing minds: programs for teaching thinking*. (rev. ed, Vol 2). Alexandria, VA: ASCD. Pp. 48-50.

Xia, Chenhong. (1998). Decision-making factors in Go expertise. Dissertation Abstracts International: Section B: the Sciences & Engineering. Vol 58(12-B), 6838, US: Univ Microfilms International.

Table 1: Distribution of Results for Pretest, First Post-test, and Second Post-test

Scores	Pre-Test	First Post-Test	Second Post-Test
0	0	1	0
1	2	2	3
2	8	2	11
3	11	2	22
4	19	4	19
5	27	13	26
6	40	25	26
7	16	35	18
8	17	22	2
9	13	18	2
10	13	20	1
11	7	19	1
12	9	8	1
13	4	3	0
14	4	0	0
15	5	5	1
Total	**195**	**179**	**133**
Mean	**7.0**	**8.1**	**5.0**
SD	**3.2**	**2.7**	**2.2**

Table 2: Regression Results for First Post-Test (N=179)

	Dependent Variable	
Predictors	Successfully Completed Puzzles	Attempted Puzzles
Treatment	.562	.001
	(.309)	(.337)
Pre-test Score	.457	
	(.058)	
Pre-test Attempts		.304
		(.061)
Classroom 1	-3.695	1.556
	(.684)	(.751)
Classroom 2	-.549	.756
	(.639)	(.701)
Classroom 3	-1.064	.176
	(.574)	(.621)
Classroom 4	-.133	1.596
	(.702)	(.698)
Classroom 5	-.406	2.471
	(.593)	(.667)
Classroom 6	-.029	2.198
	(.579)	(.636)
Classroom 7	-.169	1.941
	(.672)	(.736)
Constant	5.216	7.329
	(.587)	(.788)

Table 3: Regression Results for First and Second Post-Test Among Those Subjects who Took Both Tests (N=122)

	Dependent Variable	
Predictors	First Post-Test	Second Post-Test
Treatment	.444	1.215
	(.380)	(.355)
Pre-test Score	.435	.225
	(.071)	(.067)
Classroom 2	-.315	-1.341
	(.685)	(.641)
Classroom 3	-.918	-2.448
	(.605)	(.565)
Classroom 5	-.308	-.997
	(.624)	(.583)
Classroom 6	.010	-2.051
	(.634)	(.593)
Classroom 7	-.662	-.908
	(.750)	(.701)
Constant	5.393	4.159
	(.675)	(.631)