



# Um currículo que ensina habilidades de raciocínio pode melhorar o desempenho em testes padronizados?

Donald Green  
Yale University

Dan Gendelman  
Eshcolot

*Can a curriculum that teaches abstract reasoning skills  
improve standardized test scores?*

**Um currículo que ensina habilidades de raciocínio  
pode melhorar o desempenho em testes padronizados?**

*Donald Green*  
Yale University

*Dan Gendelman*  
Eshcolot

August 23, 2004

**Abstract:** *A pilot study was conducted in order to determine whether a curriculum that teaches strategic principles improves performance on standardized tests. Two classrooms in a low-SES Israeli school participated in an experiment. The more academically advanced classroom was exposed to a series of strategy games; the less advanced classroom was both exposed to games and taught a series of principles designed to guide their thinking in complex choice situations. These principles were taught with special reference to strategy games and then reinforced in the course of ordinary schoolwork. We find that in the wake of this intervention, the less advanced class outperformed the more advanced class on computerized tests of game-playing performance and on paper-and-pencil tests of verbal and math ability.*

**Resumo:** Um estudo piloto foi conduzido com o objetivo de determinar se um currículo que ensina princípios estratégicos melhora o desempenho em testes padronizados. Duas salas de aula de uma escola israelita de baixa renda (Low SES – Alunos com Desvantagens Socioeconômicas) participaram do experimento. A sala de aula mais avançada academicamente foi exposta a jogos de estratégia. A sala de aula menos avançada foi exposta aos mesmos jogos e a uma série de princípios destinados a orientar seus pensamentos em uma situação complexa de escolhas. Estes princípios foram ensinados por meio dos jogos de estratégia e reforçados no andamento dos trabalhos escolares comuns. Os resultados indicam que a classe menos avançada superou o desempenho da sala mais avançada em testes computadorizados e nos testes escritos de habilidades matemáticas e verbais.

## Introduction

*The present study examines the effectiveness of a school-based curriculum called The Mind Lab that attempts 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 enhance strategic reasoning by drawing 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.*

*Previous research (Green and Gendelman 2003) has shown that the Mind Lab curriculum improves performance in game-playing ability, as gauged by computerized tests. The Green and Gendelman experiments randomly assigned children to treatment and control conditions. Both experimental groups were presented with an introduction to the rules of various games. Children in the treatment group were taught strategic principles useful to the solution of certain types of game-related puzzles, while children in the control group spent the equivalent time period practicing the games. Green and Gendelman find statistically significant increases in performance among those students assigned to receive instruction in strategic principles.*

*This evaluation extends this line of investigation. In addition to examining whether the analogy-based approach used by the Mind Lab improves performance in strategy games, we consider a more general treatment that integrates strategic reasoning into the day-to-day curriculum of a classroom. The outcome variable in question is not simply game performance but performance on standardized academic tests.*

## Introdução

O presente estudo examina a efetividade da Metodologia desenvolvida pela *Mind Lab Group*, que procura melhorar as habilidades estratégicas de raciocínio. É um programa que fornece instruções e materiais de ensino às escolas do Ensino Fundamental, utilizando jogos, com o objetivo de melhorar o raciocínio por meio de analogias com situações reais da vida. Por exemplo, ao ensinar crianças a raciocinar em jogos que envolvem problemas complexos, é provocada uma analogia com problemas que podem parecer maiores, a não ser que sejam divididos em uma série de passos mais simples e manejáveis. A ideia por trás da analogia é estabelecer metáforas fáceis de se lembrar, que tenham significado tanto em jogos quanto na vida real.

Em pesquisa anterior, Green e Gendelman (2003) demonstraram que esta Metodologia melhora o desempenho em jogos computadorizados. O experimento de Green e Gendelman selecionou crianças ao acaso para compor os grupos experimental e de controle. Ambos foram apresentados às regras de alguns jogos. Crianças do grupo experimental foram capacitadas com princípios estratégicos úteis na solução de certos tipos de jogos “*puzzle*”<sup>1</sup>, enquanto as crianças do grupo de controle praticaram os mesmos jogos pelo mesmo período de tempo. Green e Gendelman encontraram melhorias estatisticamente significativas no desempenho dos alunos que receberam instruções de princípios estratégicos.

A atual pesquisa amplia esta linha de investigação. Para além de se examinar se a aproximação metafórica, utilizada pela *Mind Lab Group*, melhora o desempenho em jogos de estratégia, foi considerada a integração do raciocínio estratégico no dia-a-dia do currículo da sala de aula. A variável de resultado pesquisada não é o desempenho no jogo, mas sim o desempenho em testes acadêmicos padronizados.

---

<sup>1</sup> Jogos não competitivos, envolvendo problemas lógicos. Ex.: Sudoku (nota do tradutor)

*This hypothesis is tested by means of a panel study in which two classrooms of third grade students in a low SES Israeli school were tracked over time. The classroom containing more advanced third graders, as indexed by their standardized scores at the start of the year, was assigned to a control group that played a series of abstract strategy games over a period of several weeks. Strategic principles were not part of their curriculum. The classroom containing less advanced third graders was assigned to the treatment group. They received analogy-based instruction designed to organize their strategic approach to each game and to convey the relevance of strategic principles for thinking in forms of schoolwork.*

*This essay is organized as follows. We begin by providing an overview of the research 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 improved performance in game-playing performance. Prior to instruction, students in the control group slightly outperform students in the treatment group; after instruction, the treatment group significantly outperforms the control group. We then examine the effects of the experimental intervention on math and verbal scores. Although the control group slightly outperformed the treatment group on math and verbal pretests, the treatment group pulled ahead in both post-tests. The gains in verbal performance were statistically significant, as were the gains in both verbal and math scores summed together.*

Esta hipótese é testada em um estudo realizado em duas classes de aula da terceira série de uma escola israelita de baixa renda (Low SES – Alunos com Desvantagens Socioeconômicas). A sala com alunos mais avançados, conforme avaliação padrão realizada no início do ano, compôs o grupo de controle, que jogou uma série de jogos estratégicos durante algumas semanas. Princípios estratégicos não faziam parte do seu currículo. A sala com alunos menos avançados formaram o grupo experimental. Estes alunos receberam instruções, baseadas em metáforas, para organizar suas estratégias de jogo e foram estimulados a refletir sobre a relevância destes princípios estratégicos para pensar em seus trabalhos escolares.

Este artigo está organizado da seguinte forma. Inicialmente, apresentamos um resumo do plano da pesquisa. Depois de descrever os sujeitos e a intervenção experimental, apresentamos o modelo estatístico utilizado para estimar os efeitos da intervenção. Seguem os resultados obtidos, mostrando que a aproximação pedagógica, utilizada pela *Mind Lab*, melhorou significativamente o desempenho no resultado dos jogos. Antes da instrução, alunos do grupo de controle superaram por muito pouco o desempenho dos alunos do grupo experimental. Depois das instruções, o grupo experimental superou significativamente o grupo de controle. Em seguida, examinamos os efeitos da intervenção nas notas de matemática e desempenho verbal. Apesar do grupo de controle ter superado por muito pouco o grupo experimental em pré-testes matemáticos e verbais, o grupo experimental teve desempenho superior em ambos pós-testes. O ganho em desempenho verbal foi estatisticamente significativo, assim como os ganhos verbais e matemáticos somados.

## Research Design and Analysis

### **Subjects**

*This study took place in an elementary school in the town of Migdal Haemeq. This town is noteworthy for its high proportion of Ethiopian and Russian immigrants, accounting for nearly half of the town's population. This town's population contains a large fraction of poor families, and the average family income is approximately \$12,000. In Migdal Haemeq, just 35% of high school age students graduate with a full diploma. The Shalom school was characterized as a "troubled school," where teachers faced the daunting task of instructing children whose average achievement level was well below average. The teacher in the experimental classroom was eager to try the proposed intervention in the hopes of improving the morale and performance of her students. She was assisted by an instructor from the Mind Lab program, who, from January through May of 2004, helped introduce children in both classes to the computerized strategy games studied here.*

### **Outcome measures**

*During each experimental session, children played strategy games using a computer interface. The interface provided an overview of the rules, offer a series of examples, and then present a series of puzzles to be solved by the student. 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 puzzles solved.*

## **Desenho da pesquisa e análise**

### **Sujeitos**

Este estudo aconteceu em uma escola de Ensino Fundamental na cidade de Migdal Haemeq. Esta cidade é notável pela alta proporção de imigrantes etíopios e russos, que compõem quase a metade da população da cidade. A população é formada por uma grande fração de famílias pobres, sendo a média da renda familiar de aproximadamente US\$ 12.000,00. Em Migdal Haemeq, somente 35% dos alunos se formam no Ensino Médio. A Escola Shalom foi caracterizada com uma "escola problemática", onde os professores se deparam com a tarefa de instruir crianças com nível de aprendizado bem abaixo da média. A professora na sala de aula do grupo experimental estava empolgada em tentar a intervenção proposta, com a expectativa de melhorar a moral e o desempenho dos seus alunos. Ela foi orientada por um instrutor da *Mind Lab Group* que, de janeiro a maio de 2004, ajudou a introduzir as crianças nas estratégias de jogos estudados aqui.

### **Procedimentos de mensuração**

Durante cada sessão experimental, crianças brincaram com jogos de estratégia utilizando a interface de um computador. A interface forneceu uma prévia sobre as regras, ofereceu uma série de exemplos e depois apresentou uma série de "puzzles" para serem resolvidos pelos alunos. Pelo fato de serem jogos de computador, dados referentes à qualidade e quantidade foram facilmente coletados para cada aluno, em três testes. A qualidade de cada jogada foi medida pelo número de "puzzles" resolvidos.

*In addition, students were given standardized tests in math and verbal ability. The math test was developed by the school district. Since the school district did not have a standard verbal test, the teaching aide from the Mind Lab developed one based on other tests of verbal ability for third graders. The treatment and control groups were tested, under identical testing conditions, before the study began (September 2003) and after it concluded (May 2004).*

### **Overview of Study Design**

*The Mind Lab curriculum was introduced over several weeks. The first session consisted of a pre-test in which both the treatment and control groups were taught a pair of games, Rush Hour and Four-in-a-Row. The results of this session provide a benchmark for assessment of subsequent treatment effects. The second session marked the beginning of the experimental intervention; the treatment and control groups were segregated, and the treatment group was taught strategic principles relevant to the game Rush Hour. This lesson stressed the importance of breaking down a complex problem into its component parts, and working in a methodical fashion to solve a formidable challenge. The treatment and control groups tested their performance in both games. The second session also saw the introduction of a new game, Touchdown. Treatment and control groups were taught the rules and, with similar background preparation, tested their ability against a series of computer puzzles. The third lesson introduced students in the treatment group to strategic principles relevant to Touchdown. This lesson offered an accessible introduction to the idea of backward induction, solving a puzzle by working backward from the endpoint to the current position. During this session, students were tested again on Rush Hour and Touchdown. The fourth session saw the introduction of a new game, Lunar Lockout, but no special instruction on principles relevant to it. Students were tested on Touchdown and Lunar Lockout. Thus, we have pre-tests and post-tests for the games Rush Hour, Four-in-a-Row, and Touchdown. In fact, for Rush Hour and Touchdown, we have two successive post-tests.*

Além disso, os alunos realizaram testes padronizados de matemática e de habilidades verbais. O teste de matemática foi desenvolvido por um distrito escolar. Como o distrito escolar não dispunha de um teste verbal padrão, a *Mind Lab Group* desenvolveu um teste baseado em outro já existente para alunos de terceira série. Os grupos experimental e de controle foram testados sob condições idênticas antes que o estudo começasse (setembro de 2003) e depois de concluído (maio de 2004).

### **Desenvolvimento do Estudo**

O currículo Mind Lab foi ministrado durante algumas semanas. A 1ª sessão consistiu de um pré-teste, no qual foram ensinados os jogos “Hora do Rush” e “Lig-4”. Os resultados desta sessão possibilitaram colher dados iniciais. A partir da 2ª sessão, o grupo experimental e de controle foram separados. Ao grupo experimental, foram ensinados princípios estratégicos relevantes para o jogo Hora do Rush, mostrando-se a importância de se quebrar um problema complexo em várias partes e trabalhar de forma metódica para resolver o desafio. Os dois grupos foram testados nos dois jogos. Foi introduzido mais um jogo: “Touchdown”. Na 3ª sessão, foi ensinada a estratégia “Indução Reversiva”: resolver o desafio “de trás para frente”, do fim para o começo. Os alunos foram testados em “Hora do Rush” e “Touchdown”. Na 4ª sessão, o jogo “Missão Lunar” foi introduzido, sem instruções de novos princípios. Os alunos foram testados em “Touchdown” e “Missão Lunar”. Assim, obtivemos pré-teste e pós-teste para os jogos “Hora do Rush”, “Lig-4” e “Touchdown”, sendo que nos jogos Hora do Rush e Touchdown temos dois pós-testes sucessivos.

*It should be stressed that throughout the experimental period, the teacher sought to integrate the Mind Lab lessons into the curriculum. Consider, for example, the way that two strategic ideas, the “Detective Method” and the “Ladder Method” were integrated into the curriculum of the treatment classroom. The Detective Method is premised on the idea that asking questions can be a tool to clarify one’s goals and identify the obstacles that prevent one from attaining them. Before every written exercise, students were encouraged to first think about their “detective questions” and say why the answer to the question will help them realize the objective of the exercise. The Ladder Method emphasizes the importance of breaking down a complex problem into its component parts and identifying secondary objectives that will serve one’s primary objective. Students were urged to use the Ladder Method when engaging math problems. For example, one child remarked that “In math lessons I know that I have to solve every stage, if I only miss one part then everything will go down.” When confronted with a problem like 37 plus 15, a student said “I should do it with the Ladder Method and not try to guess the answer in one step. First, I fill in a nice numbers like  $37 + 13 = 50$ , and now I have another 2 so the answer is 52.” This method also applied to verbal reasoning. As one student noted, “When I have to read a text I should ask myself, ‘What is the meaning of the first part of the text and then the second part?’”*

Durante o período experimental, a professora procurou integrar a Metodologia *Mind Lab* ao currículo. Vejamos, por exemplo, a maneira como duas estratégias metafóricas, o “Método do Detetive” e o “Método da Escada”, foram integradas ao currículo no grupo experimental. O Método do Detetive é baseado na ideia de que fazer perguntas pode ser uma ferramenta para clarear os objetivos e identificar os obstáculos que impedem que sejam alcançados. Antes de qualquer exercício escrito, os alunos eram encorajados a primeiro elaborar perguntas investigativas e a dizer por que a resposta àquela pergunta iria ajudá-los a entender o objetivo do exercício. O Método da Escada enfatiza a importância de se quebrar um problema complexo em diversas partes/componentes e identificar objetivos secundários que servirão como degraus para atingir os objetivo primário. Os alunos foram estimulados a utilizar o Método da Escada quando envolvidos em problemas matemáticos. Por exemplo, uma criança enfatizou que “*em lições matemáticas eu sei que tenho que resolver todos os estágios, se eu perder uma parte deles tudo pode dar errado.*” Quando confrontado com um problema tipo 37 mais 15, um aluno disse: “*antes de mais nada, eu deveria resolver utilizando o Método da Escada e não tentar adivinhar a resposta. Primeiro, eu faço um cálculo fácil, como  $37 + 13 = 50$ , e agora eu reponho o 2 e a resposta é 52*”. Este método também foi aplicado em raciocínio verbal. Um aluno salientou assim: “*quando eu tenho que ler um texto, eu deveria perguntar a mim mesmo: qual é o significado desta primeira parte do texto e depois da segunda parte?*”

### **Session-by-Session results**

*A total of 19 students (from the less advanced class) were assigned to the treatment group, and 16 students (from the more advanced class) were assigned to the control group. Due to the vagaries of student attendance, some of the students tested in the pre-test phase of the experiment were absent when subsequent tests were administered. The number of observations varies from 15 to 19 in the treatment group and from 13 to 16 in the control group.*

*The assignment of classrooms to experimental groups produced a treatment group that lagged behind the control group in terms of pre-intervention test scores. For example, the control outperformed the treatment group on a standardized math test used in the local school district (treatment mean = 67.8, control group mean = 70.9), on a standardized verbal test devised by the Mind Lab instructors (treatment group mean = 60.2, control group mean = 70.9), and in the Rush Hour pretest (average number of puzzles solved by the treatment group = 4.4, average numbers solved by the control group = 5.7).*

*As shown in Table 1, the second meeting saw the treatment group surpass the control group in terms of average performance on both of the games on which they had received instruction. The treatment group, whose scores had been poorer than the control group in the pre-test, now outperformed the control group in Rush Hour (treatment group average = 6.6, control group average = 5.1). The computer continued to dominate students in 4-in-a-Row, although it should be noted that both students who won any games at all against the computer were in the treatment group, which had earlier received instruction on relevant strategic principles. There was no difference in performance on the game Touchdown, which the students played initially without any instruction in strategic principles (treatment group average = 5.3, control group average = 5.4).*

### **Resultados Sessão por Sessão**

Um total de 19 alunos (da sala menos avançada) foi designado para o grupo experimental, e 16 alunos (da sala mais avançada) para o grupo de controle. Devido à variação da frequência dos alunos, alguns alunos que efetuaram os pré-testes estavam ausentes quando testes subsequentes foram administrados. O número de observações varia de 15 a 19 no grupo experimental e de 13 a 16 no grupo de controle.

Nas avaliações realizadas antes da intervenção, o grupo de controle superou o grupo experimental em testes padronizados de matemática (grupo experimental = 67,8; grupo de controle = 70,9), em testes verbais padronizados (grupo experimental = 60,2; grupo de controle = 70,9) e no pré-teste de “Hora do Rush” (média de “puzzles” solucionados pelo grupo experimental = 4,4; média de “puzzles” solucionados pelo grupo de controle = 5,7).

Como mostra a Tabela 1, no segundo encontro o grupo experimental superou o grupo de controle em termos de média de desempenho em ambos os jogos em que receberam instrução. O grupo experimental, cujos resultados eram inferiores aos do grupo de controle no pré-teste, agora superou o grupo de controle no jogo “Hora do Rush” (média do grupo experimental = 6,6; média do grupo de controle = 5,1). O computador continuou vencendo os alunos no “Lig-4”, apesar de que deve ser salientado que os alunos que no geral ganharam qualquer jogo contra o computador estavam no grupo experimental, que tinha recebido instruções relevantes de princípios estratégicos. Não houve diferença significativa de desempenho no jogo “Touchdown”, que os alunos jogaram inicialmente sem qualquer instrução de princípios estratégicos (média do grupo experimental = 5,3; média do grupo de controle = 5,4).



**Table 1:** Performance of Treatment and Control Groups, by Game and Experimental Session  
(Table entries are means with Ns in parentheses)

|                               | Session | Control Group | Treatment Group | Significance (two-tailed) |
|-------------------------------|---------|---------------|-----------------|---------------------------|
| <i>Pretest Four-in-a-Row</i>  | 1       | 0 (16)        | 0 (19)          | N/A                       |
| <i>Posttest Four-in-a-Row</i> | 2       | 0 (16)        | .13 (15)        | .14                       |
| <i>Pretest Rush Hour</i>      | 1       | 5.7 (16)      | 4.4 (19)        | .09                       |
| <i>Posttest Rush Hour</i>     | 2       | 5.1 (16)      | 6.6 (15)        | .06                       |
| <i>Posttest Rush Hour</i>     | 3       | 8.0 (14)      | 9.0 (16)        | .17                       |
| <i>Pretest Touchdown</i>      | 2       | 5.4 (16)      | 5.3 (14)        | .91                       |
| <i>Posttest Touchdown</i>     | 3       | 8.1 (15)      | 9.1 (16)        | .33                       |
| <i>Posttest Touchdown</i>     | 4       | 10.3 (15)     | 11.7 (15)       | .15                       |
| <i>Pretest Lunar Lockout</i>  | 4       | 4.1 (16)      | 4.2 (17)        | .81                       |

**Tabela 1:** Desempenho dos grupos experimental e de controle, por jogo e sessão  
(entradas na tabela são médias, com número de alunos entre parênteses)

|                               | Sessão | Média Grupo de Controle | Média Grupo experimental | Relevância |
|-------------------------------|--------|-------------------------|--------------------------|------------|
| <i>Pré-teste Lig-4</i>        | 1      | 0 (16)                  | 0 (19)                   | N/A        |
| <i>Pós-teste Lig-4</i>        | 2      | 0 (16)                  | ,13 (15)                 | ,14        |
| <i>Pré-teste Hora do Rush</i> | 1      | 5,7 (16)                | 4,4 (19)                 | ,09        |
| <i>Pós-teste Hora do Rush</i> | 2      | 5,1 (16)                | 6,6 (15)                 | ,06        |
| <i>Pós-teste Hora do Rush</i> | 3      | 8,0 (14)                | 9,0 (16)                 | ,17        |
| <i>Pré-teste Touchdown</i>    | 2      | 5,4 (16)                | 5,3 (14)                 | ,91        |
| <i>Pós-teste Touchdown</i>    | 3      | 8,1 (15)                | 9,1 (16)                 | ,33        |
| <i>Pós-teste Touchdown</i>    | 4      | 10,3 (15)               | 11,7 (15)                | ,15        |
| <i>Pré-teste Missão Lunar</i> | 4      | 4,1 (16)                | 4,2 (17)                 | ,81        |

The third session saw the students hone their skills on Rush Hour and Touchdown, and no new games were presented. The treatment group outperformed the control group in both games. In both cases, the mean in the treatment group was one point higher, indicating that on average students in the treatment group solved one additional puzzle. The fourth session gave both groups one more opportunity to test their skills in Touchdown, and the treatment group continued to solve more puzzles than the control group. When both groups were presented with a new game with no special instruction, however, the two groups performed almost identically. The near-equivalence of the two groups in the pretest for both Touchdown and Lunar Lockout may be interpreted in two ways. One interpretation relates these scores to the baseline and concludes that instruction in strategic principles helped buoy the scores of the treatment group, which prior to any of the interventions lagged behind the control group. Alternatively, one could ignore the baseline scores of the two groups and, seeing no difference in average scores, conclude that the instruction did nothing to distinguish the pretest scores of the treatment group.

### **Estimation**

In this section, we attempt to gauge the effects of the experimental treatment more precisely. Two statistical models are considered. The first is simply a linear regression model of the outcome measure (posttest scores) on the independent variable, experimental group. 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  $U$  represent a vector of disturbances. The regression model is

$Y = a + Xb + u$ , which turns out to be equivalent to a comparison of average outcomes in the treatment and control groups.

Na terceira sessão, os alunos praticaram suas habilidades em “Hora do Rush” e “Touchdown” e nenhum jogo novo foi apresentado. O grupo experimental superou o grupo de controle nos dois jogos. Em ambos os casos, a média do resultado no grupo experimental foi um ponto superior, indicando que a média dos alunos no grupo experimental resolveu um “puzzle” a mais. A quarta sessão deu aos dois grupos mais uma oportunidade para testar suas habilidades em “Touchdown”, sendo que o grupo experimental continuou a resolver mais “puzzles” do que o grupo de controle. Quando um novo jogo foi introduzido, sem instruções especiais, aos dois grupos, ambos tiveram um desempenho quase idêntico. A equivalência dos dois grupos no pré-teste para “Touchdown” e “Missão Lunar” pode ser interpretada de duas formas. Uma interpretação refere-se ao desempenho inicial dos grupos e conclui que as instruções em princípios estratégicos ajudou a manter o nível de performance do grupo experimental, que anteriormente era inferior ao do grupo de controle. Por outro lado, se ignorarmos os pontos de partida dos dois grupos, podemos concluir que a instrução não colaborou para diferenciar a pontuação dos grupos no pré-teste de um novo jogo.

### **Resultados estatísticos**

Neste item, buscamos determinar os efeitos do tratamento experimental de forma mais precisa. Dois modelos estatísticos foram considerados. O primeiro é simplesmente um modelo regressivo linear do resultado da medida (pontos do pós-teste) em uma variável independente: grupo experimental.  $Y$  representa o vetor dos pontos do pós-teste;  $X$  representa uma variável *dummy* de valor 1 se o aluno foi designado para o grupo experimental;  $U$  representa o vetor de interferências. O modelo de regressão é:  $Y = a + Xb + u$ , que assegura a comparação das médias dos resultados dos grupos experimental e de controle.

Given the small size of this panel study, pre-test scores may play a potentially useful role in estimation. By reducing the amount of unexplained variation in the dependent variable, these covariates improve the precision with which the experimental effects are estimated. Let  $P$  represent a matrix of pre-test game scores, and let  $S$  denote pre-intervention standardized test scores. Thus,  $P$  and  $S$  comprise the pre-test verbal score, the pre-test math score, and the pre-test score of the relevant strategy game. The regression model is:  $Y = a + Xb + Sc + Pd + u$ .

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.

### **Results: Curriculum Affects Game Performance**

Table 2 reaffirms the findings of Green and Gendelman (2003). The Mind Lab curriculum improves game-playing performance. This pattern holds for both Rush Hour and Touchdown, although the effect is most apparent statistically when the two games are analyzed jointly. Without controlling for covariates, we find that the treatment group completed an average of 6.5 more puzzles for both games. Controlling for covariates, this number jumps to 7.7. Both numbers are statistically robust ( $p < .01$ ).

Tendo em vista o pequeno recorte deste estudo, pontos do pré-teste podem desempenhar uma função útil na estimativa estatística. Através da redução da quantidade de variações sem explicação nas variáveis dependentes, estas covariáveis melhoram a precisão com a qual os efeitos experimentais são estimados.  $P$  representa uma matriz de pontuação de jogos pré-teste e  $S$  refere-se aos resultados de testes padronizados pré-intervenção. Portanto,  $P$  e  $S$  incluem os pontos do pré-teste verbal, pontos do pré-teste de matemática e os pontos dos pré-testes dos jogos de estratégia. O modelo de regressão é:  $Y = a + Xb + Sc + Pd + u$ .

As hipóteses centrais deste estudo retratam o parâmetro  $b$ : se a intervenção melhora o desempenho em testes,  $b$  é positivo. Portanto, um teste unilateral é utilizado para determinar a significância do resultado contra a hipótese nula, de que a intervenção não interferiu na melhora dos resultados.

### **Resultados: o currículo *Mind Lab* afeta o desempenho em jogos**

A tabela 2 reafirma os resultados da pesquisa de Green e Gendelman (2003). O currículo *Mind Lab* melhorou o desempenho nos jogos “Hora do Rush” e “Touchdown”, embora o resultado seja mais aparente quando os dois jogos são analisados juntos. Sem controlar esta covariável, percebemos que o grupo experimental completou uma média de 6,5 “puzzles” em ambos os jogos. Controlando as covariáveis, este valor salta para 7,7. Ambos os resultados são estatisticamente significativos ( $p < ,01$ ).

**Table 2: Effects of Treatment on Posttest Game Scores, with and without Covariates**

|                         | Rush Hour    | Rush Hour     | Touchdown    | Touchdown     | Both Games   | Both Games    |
|-------------------------|--------------|---------------|--------------|---------------|--------------|---------------|
| Treatment               | 3.0<br>(1.4) | 3.6<br>(1.5)  | 2.6<br>(2.0) | 3.9<br>(1.4)  | 6.5<br>(2.9) | 7.7<br>(2.1)  |
| Pretest Rush Hour Score |              | .4<br>(.3)    |              |               |              | .3<br>(.6)    |
| Pretest Touchdown Score |              |               |              | 1.7<br>(.3)   |              | 2.2<br>(.6)   |
| Pretest Verbal Score    |              | -.04<br>(.05) |              | -.04<br>(.04) |              | -.08<br>(.06) |
| Pretest Math Score      |              | .04<br>(.05)  |              | .07<br>(.04)  |              | .07<br>(.07)  |
| N                       | 27           | 26            | 30           | 25            | 26           | 24            |

**Tabela 2: Efeitos da intervenção na Pontuação dos Jogos Pós-Teste, com e sem covariáveis**

|  | <i>Hora do Rush</i> | <i>Hora do Rush</i> | <i>Touchdown</i> | <i>Touchdown</i> | <i>Ambos Jogos</i> | <i>Ambos Jogos</i> |
|--|---------------------|---------------------|------------------|------------------|--------------------|--------------------|
| <i>Tratamento</i>                          | 3,0<br>(1,4)        | 3,6<br>(1,5)        | 2,6<br>(2,0)     | 3,9<br>(1,4)     | 6,5<br>(2,9)       | 7,7<br>(2,1)       |
| <i>Pontuação do Pré-teste Hora do Rush</i> |                     | ,4<br>(,3)          |                  |                  |                    | ,3<br>(,6)         |
| <i>Pontuação pré-teste Touchdown</i>       |                     |                     |                  | 1,7<br>(,3)      |                    | 2,2<br>(,6)        |
| <i>Pontuação pré-teste Verbal</i>          |                     | -,04<br>(,05)       |                  | -,04<br>(,04)    |                    | -,08<br>(,06)      |
| <i>Pontuação pré-teste Matemático</i>      |                     | ,04<br>(,05)        |                  | ,07<br>(,04)     |                    | ,07<br>(,07)       |
| <i>N</i>                                   | 27                  | 26                  | 30               | 25               | 26                 | 24                 |

## **Results: Curriculum Affects Standardized Test Performance**

*That instruction improves performance in strategic games is a conclusion that is neither new nor altogether surprising. What sets this evaluation apart from previous research is that it estimates the effects of the Mind Lab curriculum on standardized test performance. Two tests were administered, one math test that was commonly used among nearby school districts and a standardized verbal test that was developed in collaboration with the classroom teacher.*

*Our regression model predicts post-test scores on each test with a dummy variable marking the experimental treatment and pre-test scores in math, verbal, and Rush Hour. Note that all of the pre-test covariates were measured before any Mind Lab instruction occurred. Although intuition suggests that strategic thinking might be more readily generalized to math as opposed to verbal reasoning, our results suggest that the effects are at least as strong for verbal scores. Table 3 reports that the treatment group scored on average 11.9 points higher than the control group ( $SE=5.1$ ,  $p = .014$ ), which is slightly more than half a standard deviation improvement in the verbal post-test. The results were also positive for math, with the treatment showing a 4.8 point gain. However, given the standard error of 3.1, this effect falls just shy of statistical significance ( $p = .065$ ). Summing the two post-tests together to form a composite score and repeating the regression estimation shows a statistically significant 17.5 point gain ( $SE=7.5$ ,  $p=.015$ ).*

## **Resultados: o currículo afeta o desempenho de testes padronizados**

Que a instrução da Metodologia Mind Lab melhora o desempenho em jogos de estratégia é uma conclusão que não é nem nova e nem traz espanto. O que diferencia este estudo da pesquisa anterior é que esta estima os efeitos do currículo Mind Lab sobre o desempenho em testes padronizados. Dois testes foram administrados, um de matemática, que era de uso comum nos distritos escolares dos arredores, e um verbal, que foi desenvolvido em colaboração com o professor da sala de aula.

Este modelo de regressão prognostica a pontuação de pós-testes em cada teste com uma variável dummy, indicando o tratamento experimental e a pontuação do pré-teste em matemática, desempenho verbal e “Hora do Rush”. Todas as covariáveis do pré-teste foram medidas antes que qualquer instrução da Mind Lab ocorresse. Apesar da intuição sugerir que estratégias de pensamento poderiam ser mais prontamente generalizados para a matemática (em oposição ao raciocínio verbal), os resultados obtidos sugerem que os efeitos são significativos para a pontuação verbal. A tabela 3 mostra que o grupo experimental pontuou cerca de 11,9 pontos a mais do que o grupo de controle ( $SE = 5,1$ ,  $p = ,014$ ), que é uma melhora de mais da metade de um desvio padrão no pós-teste verbal. Os resultados foram positivos também para matemática, com o grupo experimental mostrando um ganho de 4,8. Entretanto, tomando um erro padrão de 3,1, este efeito cai significância estatística ( $p= ,065$ ). Somando-se os dois pós-testes, de modo a formar uma pontuação composta, e repetindo-se a estimativa regressiva, evidencia-se uma significância estatística de um ganho de 11,5 pontos ( $SE = 7,5$ ,  $p = 0,015$ ).

The estimates in Table 3 vary markedly, depending on whether we control for baseline covariates. Without controls for baseline performance, the effects are weakly positive. When controls are included, the treatment effects become much stronger. This pattern reflects the contrasting baseline scores of the two groups. After the intervention, the treatment slightly outperformed the control group. However, when one takes into account the substantial gap between these two classrooms prior to the study, the effects of the *Mind Lab* curriculum become much larger. In other words, the *Mind Lab* curriculum appears to have closed the substantial gap between the two classrooms.

**Table 3: Effects of Treatment on Posttest Game Scores, with and without Covariates**

|                         | Verbal    | Verbal     | Math      | Math      | Verbal + Math | Verbal + Math |
|-------------------------|-----------|------------|-----------|-----------|---------------|---------------|
| Treatment               | 1.3 (6.1) | 11.9 (5.1) | 1.4 (4.7) | 4.8 (3.1) | 1.3 (9.6)     | 17.5 (7.5)    |
| Pretest Verbal Score    |           | .66 (.13)  |           | .18 (.08) |               | .85 (.20)     |
| Pretest Math Score      |           | -.03 (.12) |           | .47 (.08) |               | .42 (.18)     |
| Pretest Rush Hour Score |           | .91 (1.06) |           | .82 (.70) |               | 1.81 (1.57)   |
| <b>N</b>                | 30        | 30         | 34        | 33        | 30            | 30            |

As estimativas na tabela 3 variam notadamente, dependendo se controlamos as covariações da linha de partida. Sem o controle da variável “desempenho inicial”, os efeitos são pouco positivos. Quando é incluído, os efeitos da intervenção tornam-se bastante positivos. Este modelo reflete o contraste dos pontos de partida dos dois grupos. Após a intervenção, o grupo experimental superou levemente o grupo de controle. Assim, quando avaliamos a diferença substancial entre as duas salas de aula antes do estudo, os efeitos do currículo *Mind Lab* tornam-se muito maiores. Em outras palavras, o currículo da *Mind Lab* parece ter colaborado para suprimir a disparidade entre as duas salas de aula.

**Tabela 3: Efeitos da intervenção na pontuação pós-teste dos jogos, com e sem as covariáveis**

|                         | Verbal    | Verbal     | Matemática | Matemática | Verbal + Matemática | Verbal + Matemática |
|-------------------------|-----------|------------|------------|------------|---------------------|---------------------|
| Tratamento              | 1,3 (6,1) | 11,9 (5,1) | 1,4 (4,7)  | 4,8 (3,1)  | 1,3 (9,6)           | 17,5 (7,5)          |
| Pré-teste verbal        |           | ,66 (,13)  |            | ,18 (,08)  |                     | ,85 (,20)           |
| Pré-teste de Matemática |           | -,03 (,12) |            | ,47 (,08)  |                     | ,42 (,18)           |
| Pré-teste Hora do Rush  |           | ,91 (1,06) |            | ,82 (,70)  |                     | 1,81 (1,57)         |
| <b>N</b>                | 30        | 30         | 34         | 33         | 30                  | 30                  |

## Discussion

*The data presented above indicate that the Mind Lab curriculum had two statistically significant effects. First, as expected, it improved game-playing performance, as gauged by the number of puzzles each child solved during each session. Prior to the experimental intervention, the control group enjoyed a slight edge in game-playing performance; after the intervention, the treatment group significantly outplayed the control group.*

*Second, and more importantly, the treatment group outperformed the control group in standardized testing. The results were more decisive for verbal than for math scores; nevertheless, the findings lend surprising support for the idea that instruction in strategic reasoning improves academic scores.*

*Additional research is needed to understand why the Mind Lab curriculum improved verbal and math scores. One possibility is that the curriculum helped students in the treatment group negotiate standardized tests – they might have become more sensitive to the strategy of picking the best option from a field of choices. Another possibility is that games made ordinary schoolwork more fun for the children in the treatment group, enhancing their attentiveness to their lessons. A small post-intervention interview of the children lends some support to both of these hypotheses, as children expressed widespread enthusiasm for the Mind Lab curriculum, claimed that it gave them confidence, and often said that it helped them in their all around academic performance. It remains to be seen whether the promising results in this pilot study can be replicated in other grades and school environments.*

## Análise

Os dados apresentados indicam que o currículo *Mind Lab* teve, estatisticamente, dois efeitos significativos. Primeiro, como esperado, melhorou o desempenho nos jogos na medida em que cada “puzzle” foi sendo resolvido durante as sessões. Antes da intervenção experimental, o grupo de controle demonstrava um desempenho ligeiramente superior. Depois da intervenção, o grupo experimental superou significativamente o grupo de controle.

Segundo, e mais importante, o grupo experimental superou o grupo de controle nos testes padronizados. Os resultados foram mais evidentes nos testes verbais do que nos matemáticos. Assim, os resultados servem de apoio para a ideia de que a introdução de estratégias de raciocínio melhora os resultados acadêmicos.

Novas pesquisas são necessárias para entender por que o currículo *Mind Lab* melhora o desempenho em testes verbais e matemáticos. Uma possibilidade é que o currículo tenha ajudado os alunos, no grupo experimental, a resolver testes padronizados por ter tornados os alunos mais sensíveis à estratégia de selecionar a melhor opção a partir de um rol de escolhas. Outra possibilidade é que os jogos tornaram os trabalhos habituais de sala de aula mais divertidos para as crianças do grupo experimental, intensificando sua atenção nas lições. Uma pequena entrevista realizada com as crianças após a intervenção serviu de suporte a ambas as hipóteses. As crianças expressaram bastante entusiasmo em relação ao currículo *Mind Lab*; alegaram que deu a elas segurança, e com frequência disseram que os ajudou de várias formas no desempenho acadêmico. Cabe pesquisar se os resultados promissores neste estudo piloto podem ser replicados em outras séries e em outros ambientes escolares.

## References

Green, Donald P., and Dan Gendelman. 2003. **Teaching Children to Think Strategically: Results from a Randomized Experiment.** Unpublished manuscript, Institution for Social and Policy Studies at Yale University.

## Bibliografia

Green, Donald & Gendelman, Dan. **Teaching Children to Think Strategically: Results from a Randomized Experiment.** Unpublished manuscript: Institution for Social and Policy Studies at Yale University, 2003.