

Targeted Remedial Education for Underperforming Teenagers: Costs and Benefits

Victor Lavy, *Hebrew University of Jerusalem, CEPR, and
National Bureau of Economic Research*

Analía Schlosser, *Hebrew University of Jerusalem*

This study evaluates the short-term effects of a remedial education program that provided additional instruction to underperforming high school students in Israel. The intervention prepared students for the matriculation exams. Using a comparison group of schools that enrolled in the program later and implementing a difference-in-differences estimation strategy, we found that the program raised the school mean matriculation rate by 3.3 percentage points. This gain reflects an effect on targeted participants and the absence of externalities on their untreated peers. The program was found to be less cost effective than two alternative interventions based on incentives for teachers and students.

I. Introduction

Recent increases in wage differentials between skilled and unskilled workers have led to renewed interest in ways and means of enhancing

We thank Daron Acemoglu, Joshua Angrist, Esther Duflo, Daniele Paserman, and seminar participants at Bar-Ilan University, Hebrew University, Tel Aviv University, and MIT for their helpful comments and suggestions. The Bagrut 2001 program was funded by the Israel Ministry of Education and the Branco Weiss Foundation. The statements in this article reflect the views of the authors and have not been endorsed by the program's sponsors. Contact the corresponding author, Analía Schlosser, at ani@mscc.huji.ac.il.

[*Journal of Labor Economics*, 2005, vol. 23, no. 4]
© 2005 by The University of Chicago. All rights reserved.
0734-306X/2005/2304-0006\$10.00

secondary and postsecondary education, especially among disadvantaged students. As a result, a rather large number of interventions targeting disadvantaged students have been implemented in recent years in developed countries. Evidence of their effectiveness is, however, still scanty.¹ Most studies are based on small samples and lack proper counterfactuals and cost-benefit analyses.² Particularly important is the lack of evidence on how effective high school interventions are, given the debate over the relative merit of early versus late childhood intervention.³

This study evaluates the effects of a multiyear program implemented in Israel since 1999 of remedial education for underperforming high school students. The objective of the program is to increase the percentage of students who earn matriculation certificates (hereinafter, the “matriculation rate”), mainly by means of a targeted increase in instruction time. The program was rapidly expanded and is now being conducted in about one-third of all high schools countrywide. Since it resembles remedial programs implemented in other developed countries, as noted, the lessons learned from it may be highly relevant.

Schools were chosen to participate in the program based on their (low) mean matriculation rate. The gradual phasing of schools into the program provides an opportunity to base the evaluation on a comparison of early and late enrollers. We estimate the effect of the program on average school performance even though only a fraction of the students in each school received treatment. We also attempt to distinguish the program effects on participating students from the spillover effects on their classmates.

By looking at how remedial education affects achievement in high school, this article contributes to the debate over whether investments in the later stages of a child’s development have positive payoffs. Recent evidence in support of the case for early childhood intervention is documented in Currie (2001), Currie and Thomas (2001), Krueger and Whit-

¹ For example, 21st Century Community Learning Centers provide after-school programs in about 7,500 rural and inner-city public schools in more than 1,400 U.S. communities (U.S. Department of Education 2003); Bell After-School Instructional Curriculum (BASICS) serves schools in Boston, New York City, and Washington, DC (<http://www.bellnational.org>); and Boys and Girls Clubs of America offer programs in several locations across the United States (Schinke, Cole, and Poulin 2000). A description of various tutoring/extra-instruction programs in the spirit of the one evaluated in this study can be found on the Harvard Family Research Project Web site at <http://www.gse.harvard.edu/hfrp/projects/afterschool/bibliography/tutoring.html>.

² Recent exceptions are Banerjee et al. (2004), about remedial education for third and fourth graders in India; Jacob and Lefgren (2004), on the effects of summer school and grade retention in Chicago public primary schools; and Machin, McNally, and Meghir (2004), on additional resources targeting secondary schools in disadvantaged urban areas of England.

³ For a discussion of this issue, see Carneiro and Heckman (2003).

more (2001), and Garces, Thomas, and Currie (2002). Carneiro and Heckman (2003) suggest that there is a consensus among practitioners in the field that early childhood interventions are preferred, mainly because it is much more difficult to improve an individual's cognitive and noncognitive abilities later in life. The evidence on high school interventions is mostly based on randomized trials in the United States, where an array of service-oriented dropout-prevention programs for American teens has failed to increase graduation rates (Dynarski and Gleason 1998). Our evidence, in contrast, shows positive returns for an intervention implemented at a relatively late stage of schooling. Our research is also related to the recent debate on the effects of school inputs and expenditures on student success and achievement.⁴ The evidence we provide in this article concerns a relatively neglected form of intervention, that of augmenting instruction time for targeted students rather than for an entire class.

After describing the main features of the Israeli high school system and the remedial education program in Section II, we explain the study design and identification strategy in Section III. Section IV presents evidence of the overall effect of the program, and Sections V and VI break it down into two components: effects on treated students and spillover effects on untreated students. Section VII presents a cost-benefit analysis of the program and a comparison with alternative interventions, and Section VIII offers some conclusions.

The empirical evidence shows a positive impact of the program on treated schools: approximately a 3.3 percentage point increase in the mean matriculation rate, implying an improvement of 6%. The evidence also suggests that the program affected the achievements of participating students only, increasing their probability of earning a matriculation certificate by 12 percentage points, on average implying an improvement of 22%. No evidence was found of program externalities on nonparticipating peers.

II. The High School System in Israel and the Bagrut 2001 Program

Israeli postprimary education consists of junior high school (grades 7–9) and high school (grades 10–12). High school students are enrolled either in an academic track leading to a matriculation certificate (*bagrut* in Hebrew) or in an alternative track leading only to a high school di-

⁴ Recent work has investigated the effects of class size on student achievements, using data from both randomized (Krueger 1999; Krueger and Whitmore 2001) and natural experiments (Angrist and Lavy 1999; Hoxby 2000); the effects of class size and teacher quality on the economic returns on schooling (Card and Krueger 1992; Betts 1995); the effects of teacher training and computers on pupils' learning (Angrist and Lavy 2001, 2002); and the effects of incentives for teachers on student achievements (Lavy 2002, 2004).

ploma.⁵ The bagrut is completed by passing a series of national exams in core and elective subjects beginning in tenth grade, with more tests taken in eleventh grade and most taken in twelfth grade. Students choose to be tested at various levels of proficiency, with each test awarding from one to five credit units per subject, depending on difficulty. Some subjects are mandatory, and for many the most basic level is three credits. A minimum of 20 credits is required to qualify for a matriculation certificate. About 52% of high school graduates and 46% of members of the relevant age cohort received matriculation certificates in 2002.⁶

In recent years, the Israel Ministry of Education has singled out as its top priority the need to raise the matriculation rate, especially among disadvantaged students and students in peripheral communities (Ministry of Education, Culture, and Sport 2001). The ministry instituted several programs toward this goal: achievement awards for high school students (Angrist and Lavy 2004), school incentives (Lavy 2002), teacher incentives (Lavy 2004), and the large-scale remedial education program investigated here. These initiatives provide an opportunity to gauge the effectiveness of interventions in high school. Moreover, since all the interventions were implemented during the same period and share similar goals, their effectiveness may be compared. Whereas the first three interventions cited above were aimed at improving the incentives for teachers and students, the intervention analyzed in this study is based on the concept of injecting added school resources in the form of extra instructional hours for underperforming students.

The Bagrut 2001 program targets low-achieving high schools. In its first phase, which began in October 1999, it included 10 schools. During that academic year, nine additional schools were added to the program. In the following year, there were 42 participating schools. The program continued to expand; in 2002, it included 130 schools. The gradual implementation was due mainly to budget constraints. Schools were enrolled into the program in no particular order, although the schools participating in the first year include a higher proportion of disadvantaged students and have lower tests scores and matriculation rates. Figure 1 presents the program timeline and the dates of the matriculation examinations that constitute the basis for the pre- and postprogram achievements used in the evaluation.

The intervention included individualized instruction in small study

⁵ The matriculation certificate is a prerequisite for university admission, and receiving it is one of the most economically important educational milestones. Many countries and some U.S. states have similar systems of exams and/or diplomas, e.g., the French Baccalaureate, the German Certificate of Maturity (*Reifezeugnis*), the Italian Diploma di Maturità, the New York State Regents exams, and the recently instituted Massachusetts Comprehensive Assessment System.

⁶ See the Israel Ministry of Education Web site at <http://www.education.gov.il>.

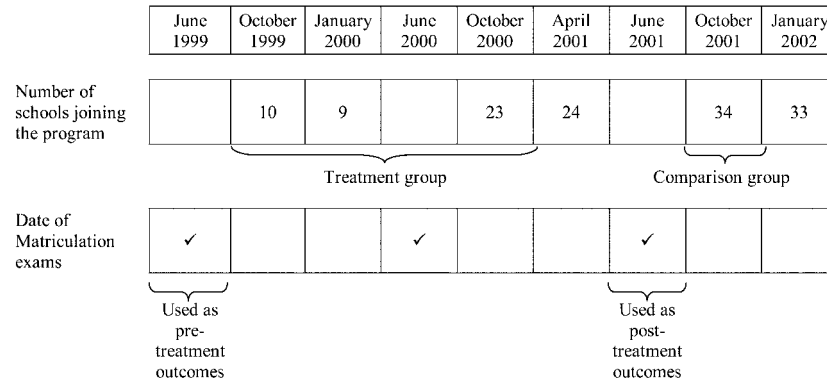


FIG. 1.—Program time line. Two treatment schools have been excluded from the analysis. One of these schools is missing from the 1999 scores dataset, and therefore there are no pretreatment outcomes for this school. The second school did not provide any data about the program implementation. Therefore, the proportion of students participating in the program cannot be measured for this school. The 24 schools that joined the program in April 2001 were not included in the analysis because their treatment lasted for only 2 months. See the text for more details.

groups of up to five students for tenth, eleventh, and twelfth graders. The aims of the intervention were (1) to design individualized instruction based on students' needs; (2) to increase the matriculation rate; (3) to enhance the scholastic and cognitive abilities, self-image, and leadership aptitudes of underperforming students. Participants were chosen by their teachers based on the likelihood of their passing the matriculation exams. Although no single quantifiable measure was used to apply this guideline, teachers were instructed to select students who had up to three failing subjects. The additional tutoring was focused on these subjects; the remedial classes were held during after-school hours and were taught by the classroom teachers.

III. Design of the Study and Identification Strategy

The first year of the study (1999–2000) was a pilot in which only twelfth-grade students from a few schools participated. Therefore, the evaluation focuses on the first year of full implementation, 2000–2001.⁷ Thus, 1998–99 is considered a pretreatment year.

Table 1 presents the treated population (4,100 students, one-fifth of all

⁷ We decided not to include in the analysis 24 schools that were added to the program in April 2001 because the program was implemented in these schools just during the last 2 months of the school year and its participants are not known. Furthermore, 19 of these schools were selected to the program as part of a package of several programs provided by the government to towns located near the country borders.

Table 1
Number and Proportion of Participants in the Bagrut 2001 Program

	Number of Students in Treatment Schools	Number of Program Participants	Program Participation Rate (%)		
			Mean	Minimum	Maximum
Grades 10–12	20,094	4,117	20	7	71
Grade 12	6,450	1,789	28	11	79

NOTE.—The data presented in the table are from the 40 schools that belong to the treatment group.

students in the treated schools). Our evaluation focuses on the effects of the program on the matriculation status of students, which is a comparable outcome for all twelfth graders.⁸ Of all twelfth graders, 28% participated in the program, but this rate varied across schools, from 11% to 79%. The large variability was the result of a rule that allowed each school to include up to 100 students in the program irrespective of school size. Although this rule was not strictly enforced, it led to a negative correlation between the proportion enrolled in the program and school size (see figs. 2 and 3).

Figures 4 and 5 show the relationship between the preprogram school matriculation rate and the program participation rate. Figure 4 shows the results for grades 10–12 and figure 5 for twelfth grade only. From both figures it is seen that there was no correlation between the program participation rate and the preprogram school matriculation rate.

Schools and students were not chosen randomly to participate in the program. Therefore, we based the identification strategy on the gradual implementation of the program, which yields a natural comparison group composed of schools that were enrolled in the program later (2001–2). The mean characteristics of these schools strongly resemble those of the schools that enrolled first. Since this similarity is found both for the pretreatment cohort of seniors and for the treated cohort, it provides some support for our claim that, during the first few years of the program, schools were enrolled in no particular order.

Table 2 presents descriptive statistics and balancing tests based on data for the 1999 graduating cohort among three groups of schools: schools that were treated in the first year of the program (col. 1), schools to be enrolled in the second year (col. 2), and all other high schools in Israel (col. 4). The table shows only a few minor differences between treated and comparison schools: the number of siblings and the proportion of students of Asian or African ethnic origin are marginally higher in the

⁸ We cannot carry the same analysis for eleventh graders because of the high variability in the tested subjects in this grade. We also lack the information that allows linking the subject in which the student received treatment to the subject tested.

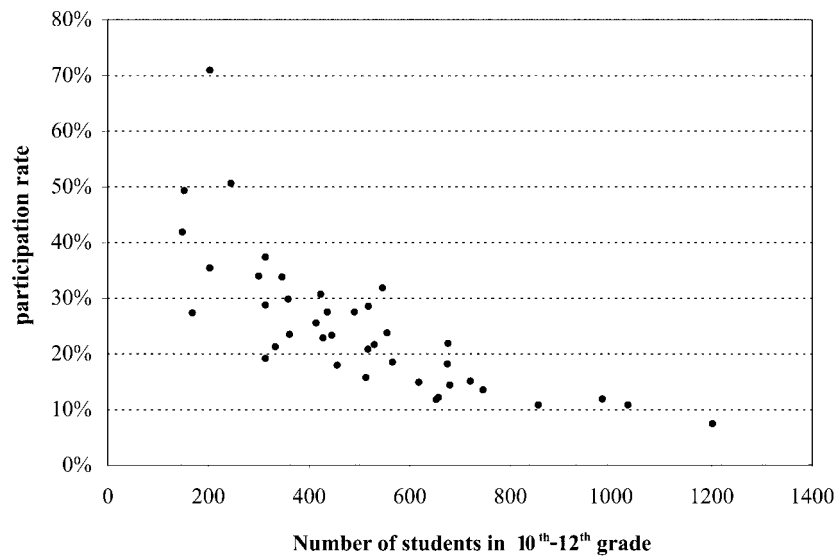


FIG. 2.—Program participation rate by school size

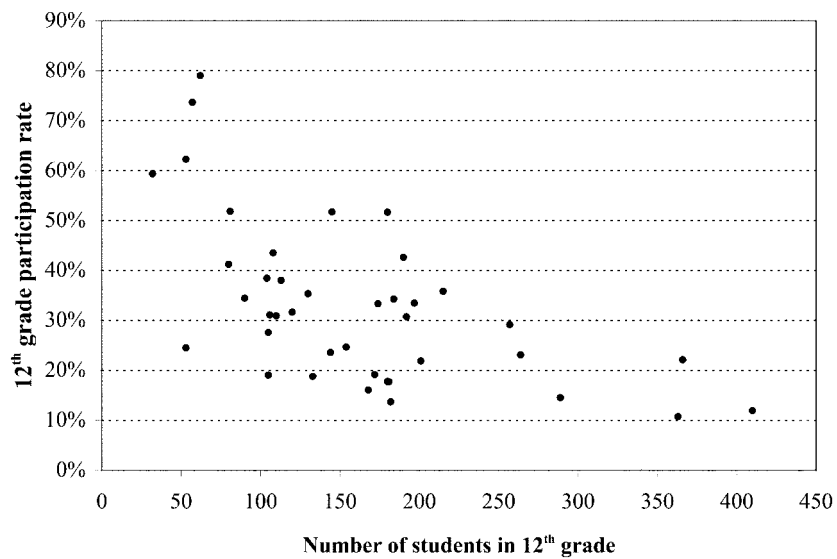


FIG. 3.—Twelfth-grade participation rate by number of twelfth-grade students at school

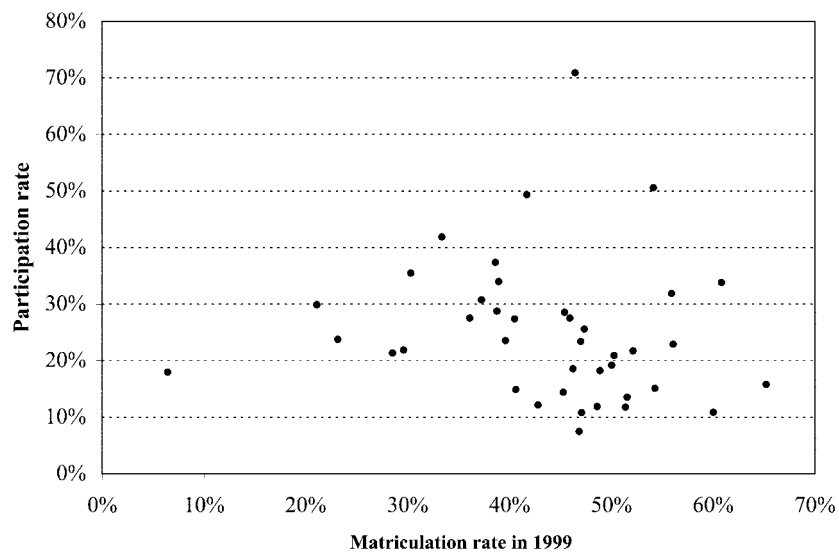


FIG. 4.—Program participation rate by pretreatment matriculation rate

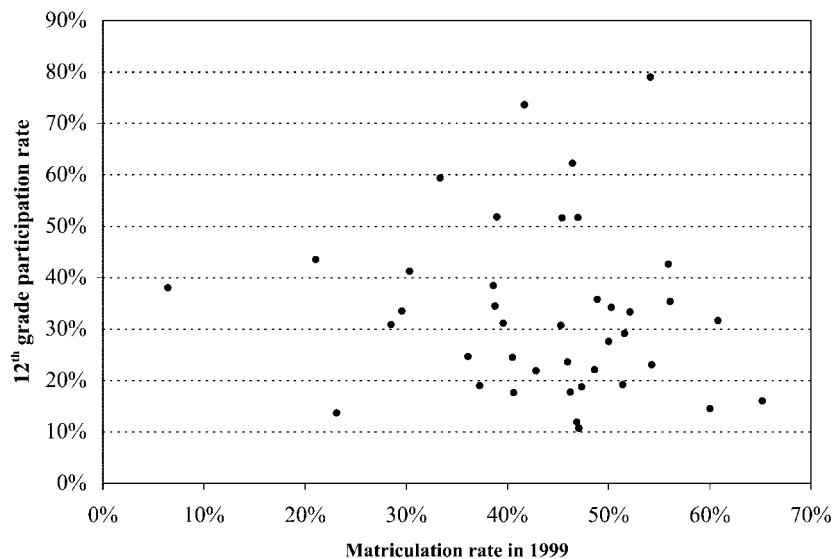


FIG. 5.—Twelfth-grade participation rate by pretreatment matriculation rate

Table 2
Comparison of Mean Covariates and Outcomes for the Pretreatment Year (1999), Treated, Comparison, and All Other Schools

	Treated Schools (1)	Comparison Schools (2)	Difference (1 – 2) (3)	All Other Schools (4)	Difference (1 – 4) (5)
Student's characteristics:					
Father's years of schooling	10.35	10.48	–.13 (.45)	12.14	–1.79 (.34)
Mother's years of schooling	10.73	10.95	–.21 (.41)	12.20	–1.46 (.28)
Number of siblings	2.58	2.25	.33 (.19)	2.28	.30 (.13)
Immigration status (= 1 if immigrated within the last 4 years)	.03	.06	–.03 (.01)	.04	–.01 (.01)
Male	.44	.46	–.02 (.020)	.48	–.04 (.02)
Parents' ethnic origin:					
Asia-Africa (excluding Ethiopia)	.37	.28	.09 (.03)	.20	.16 (.02)
Former Soviet Union	.19	.22	–.03 (.03)	.14	.05 (.02)
Ethiopia	.02	.05	–.03 (.02)	.02	–.00 (.01)
Europe-America (excluding former Soviet Union)	.07	.09	–.02 (.01)	.12	–.06 (.01)
Israel	.36	.37	–.01 (.03)	.52	–.16 (.02)
Student's achievements:					
Average lagged score in matriculation exams (prior to twelfth grade)	65.38	67.18	–1.80 (1.54)	73.44	–8.06 (.98)
Number of credit units received prior to twelfth grade	5.09	5.64	–.55 (.37)	6.50	–1.40 (.25)
School's characteristics:					
Matriculation rate	.44	.50	–.06 (.04)	.64	–.20 (.02)
Enrollment in twelfth grade	212.4	239.4	–27.1 (33.0)	222.9	–10.5 (23.6)
Enrollment in tenth grade through twelfth grade	652.8	720.6	–67.8 (101.1)	677.2	–24.4 (72.2)
Number of schools	40	34		367	
Number of students	6,725	5,508		48,710	

NOTE.—The table reports descriptive statistics from the pretreatment year (1999). The sample in col. 4 includes all remaining Jewish matriculation-track high schools in Israel. Standard errors for the differences corrected for clustering at the school level are reported in parentheses.

treated schools, suggesting that these schools have a higher proportion of students from disadvantaged families. The preprogram matriculation rate was also lower in treated schools (44% vs. 50%), although this difference is only marginally significant ($t = -1.74$).

Table 3 presents an equivalent analysis to table 2 but is based on the 2001 graduating cohort. The table shows similarity between treated and comparison schools, as is found in table 2. For the 2001 graduating cohort of table 3, it is also worth emphasizing the lack of significant differences in terms of lagged achievements (number of credits earned before twelfth grade and average score), which suggests that students from treated and comparison schools had similar achievements before the program started. On the other hand, the comparison of the treated schools with all high schools in the “other” category (i.e., those that are neither treated nor comparison schools) reveals a very different pattern. The two groups are significantly different in almost all student characteristics: students in treated schools have a lower socioeconomic status and poorer achievements, they come from larger families, their parents have less schooling, and a higher proportion are of Asian or African ethnic origin. Student achievements during the pretreatment period were also considerably higher in the “other” schools than in treated schools. For example, the matriculation rate in 1999 (the pretreatment year) was 20 percentage points higher, the 2001 seniors had an average lagged score seven points higher on matriculation exams taken before twelfth grade, and they had earned one more full credit unit before twelfth grade.

Some of the Bagrut 2001 program schools were also included in two other programs that started at the same year, 2001, as mentioned in the previous section. Ten of the Bagrut 2001-treated schools and nine of the comparison schools participated in the teacher bonus program. There are no observed differences between these two groups of schools in terms of characteristics and preprogram outcomes. The 10 Bagrut 2001-treated schools that participated in the teacher bonus program and the other Bagrut 2001-treated schools that did not are also very similar.⁹ Regardless of the close balance in terms of observables and also in terms of the number of comparison and treated schools that participated in this program, it is still of importance to assess empirically the sensitivity of the results while controlling for participation in the teacher bonus program. We should also mention that one Bagrut 2001-treated and two comparison schools participated in the student awards program. We will therefore examine as well the sensitivity of the results to controlling for this additional program.

⁹ This evidence is not shown in this article, but results are available on request from the authors.

Table 3
Comparison of Mean Covariates and Outcomes for the Treatment Year
(2001), Treated, Comparison, and All Other Schools

	Treated Schools (1)	Comparison Schools (2)	Difference (1 – 2) (3)	All Other Schools (4)	Difference (1 – 4) (5)
Student's characteristics:					
Father's years of schooling	10.62	10.69	–.07 (.41)	11.53	–.91 (.34)
Mother's years of schooling	10.75	10.99	–.25 (.48)	11.41	–.67 (.37)
Number of siblings	2.52	2.07	.45 (.18)	2.26	.26 (.15)
Immigration status (= 1 if immigrated within the last 4 years)	.02	.04	–.02 (.01)	.03	–.01 (.01)
Male	.45	.45	–.01 (.02)	.48	–.03 (.02)
Parents' ethnic origin:					
Asia-Africa (excluding Ethiopia)	.33	.23	.10 (.03)	.18	.15 (.02)
Former Soviet Union	.21	.24	–.03 (.04)	.14	.08 (.03)
Ethiopia	.02	.05	–.03 (.02)	.02	–.00 (.01)
Europe-America (excluding former Soviet Union)	.06	.08	–.03 (.01)	.12	–.06 (.01)
Israel	.38	.40	–.02 (.03)	.54	–.16 (.98)
Student's achievements:					
Average lagged score in matriculation exams (prior to twelfth grade)	68.92	70.66	–1.74 (1.70)	76.09	–7.17 (.89)
Number of credit units received prior to twelfth grade	5.86	6.16	–.31 (.36)	6.95	–1.1 (.28)
School's characteristics:					
Matriculation rate	.53	.58	–.05 (.04)	.68	–.15 (.02)
Enrollment in twelfth grade	206.8	230.8	–24.0 (31.9)	217.8	–11.0 (23.2)
Enrollment in tenth grade through twelfth grade	720.6	626.5	–100.3 (89.4)	673.7	–47.2 (63.5)
Number of schools	40	34		367	
Number of students	6,450	5,321		48,292	

NOTE.—The table reports descriptive statistics from the treatment year (2001). The sample in col. 4 includes all remaining Jewish matriculation-track high schools in Israel. Standard errors for the differences corrected for clustering at the school level are reported in parentheses.

IV. Effect of Treatment on Treated Schools

We first studied the effects of the program on the school mean matriculation rate, considering the school to be the unit of treatment even if only a small proportion of its students participated in the program. This approach allows us to estimate the overall effect of the program irrespective of the issue of selection of students for treatment. The risk, however, is that when only a small proportion of students receive treatment, accurate detection of an effect at the school level may be impossible. Furthermore, the effect on the treated students may be determined only if we assume that the program had no spillover effects or externalities on treated students' classmates.

For estimation purposes, we use the two cohorts of twelfth graders in 1999 and 2001. The repeated cross-section data are used to compare the change in the school mean matriculation rate in the treated and comparison schools. This approach amounts to a difference-in-differences estimation when school fixed effects are included in the model. The main underlying assumption of this strategy is that, in the absence of any intervention, the average change in matriculation rates would have been the same for treated and comparison schools.¹⁰ Although the validity of this assumption cannot be verified, table 4 shows that there were no significant differences in the matriculation rate trend between treated and comparison schools in the pretreatment years. In particular, there is no evidence of an "Ashenfelter dip" in outcomes in the years before treatment.

Table 4
Differences in Matriculation Rate Trends between
Treated and Comparison Schools for Pretreatment
Years, 1996–99

Regressor	Unweighted (1)	Weighted (2)
Treatment × 1997	−.037 (.058)	−.016 (.065)
Treatment × 1998	.010 (.056)	.016 (.067)
Treatment × 1999	−.023 (.050)	.008 (.059)
Number of schools	67	67

NOTE.—The table reports the interaction coefficients between treatment status and year dummies. The dependent variable is the school matriculation rate. The additional regressors are year dummies and a treatment status dummy. Data are aggregated at the school level. Estimates in col. 2 are weighted by cohort size. Robust standard errors are in parentheses. Seven schools missing in the test scores datasets for years 1996–98 are not included in the sample.

¹⁰ The model also assumes constant and additive treatment effects.

A. Empirical Models

Treatment effects are estimated using both micro data at the student level and aggregated data at the school level. In a model without additional covariates, the analyses using micro data or school averages weighted according to the number of twelfth graders provide identical results. However, when school and student characteristics are added to the model, the two estimates may differ. Adding these covariates improves the precision of the estimates, especially when student lagged achievements are added, because these achievements correlate strongly with students' matriculation status. Although the difference-in-differences models include school fixed effects, it is important to control for student and school characteristics, particularly when we are analyzing data from low-achieving schools, which often experience severe year-to-year volatility in student characteristics (Koretz 1996).

We first estimate the following model:

$$\bar{Y}_{st} = \beta' \bar{X}_{st} + \omega' S_{st} + \alpha T_{st} + \bar{\epsilon}_{st}, \quad (1)$$

where s indexes schools; $t = \{1999, 2001\}$ indexes years; $\bar{Y}_{st}$ is the proportion of students in school s who were awarded matriculation certificates in year t ; $\bar{X}_{st}$ is a vector of mean student characteristics that includes parents' schooling, gender, number of siblings, immigration status, and two measures of pretreatment achievements indicating the quartile of a student's average score on matriculation exams taken before twelfth grade and the quartile of the number of credit units earned before twelfth grade;¹¹ S_{st} is a vector of school characteristics that includes the school's religious status, the number of students in the school, its square, and the number of students in twelfth grade; and T_{st} is a dummy variable that indicates program participation status. To control for preexisting differences between treated and comparison schools and to improve the efficiency of the estimates, we stacked the data for pretreatment and treatment years and estimated a difference-in-differences model with school and year fixed effects as follows:

$$\bar{Y}_{st} = \beta' \bar{X}_{st} + \omega' S_{st} + \varphi_s + \delta_t + \alpha D_{st} + \bar{\epsilon}_{st}, \quad (2)$$

where φ_s is a school fixed effect, δ_t is a time effect, and D_{st} equals the interaction between the dummy for 2001 (the treatment year) and treatment status for school s . Similar equations are estimated based on micro

¹¹ Since in 2000 the program was implemented only among twelfth graders, the 2001 seniors that participated in the program were first exposed to treatment only in twelfth grade. Thus, their schooling achievements in the tenth and eleventh grades can safely be considered as preprogram outcomes.

data at the student level. For example, the difference-in-differences model in this case may be written as

$$Y_{ist} = \beta'X_{ist} + \omega'S_{st} + \varphi_s + \delta_t + \alpha D_{st} + \nu_{st} + \varepsilon_{ist}, \quad (3)$$

where Y_{ist} is the matriculation status of student i attending school s in year t , X_{ist} is a vector of student characteristics, and $\nu_{st} + \varepsilon_{ist}$ is the error term, which is composed of a school year-specific random element and an individual random element.

The wide variation in the proportions of participating students among schools may facilitate a more precise estimate of the treatment effect. We take advantage of the differences among schools in terms of the proportion of treated students to obtain an alternative measure of the effects of the program.¹² We use two alternative measures of treatment intensity: INT_{12s} expresses the proportion of twelfth-grade students receiving treatment at school s , and INT_s expresses the proportion of tenth-grade through twelfth-grade students receiving treatment at school s . Variations in treatment intensity reflect differences between treated and comparison schools, as well as differences within treated schools. Since the twelfth-grade treatment intensity might be correlated with the twelfth graders' perceived potential, we also use school-level treatment intensity as an alternative measure.

The main concern of using this strategy is the possible endogeneity of treatment intensity. To examine this issue, we evaluate whether preprogram achievements pertaining to the school or students predict treatment intensity while controlling for school size. As preprogram achievements, we use the mean matriculation rate of the previous cohort and, alternatively, the treated cohort's lagged outcomes.

Table 5 shows that treatment intensity is not correlated with lagged school achievements or lagged student achievements but is highly negatively correlated with school size.¹³ In addition, results presented in table A1 in the appendix suggest that there were not significant differences in the matriculation rate trend between schools according to the two alternative measures of treatment intensity. This evidence suggests that, conditional on school size, treatment intensity may be considered independent of potential achievements. The estimated difference-in-differences models include school fixed effects, which eliminate any remaining per-

¹² This strategy is similar to that applied by Card and Krueger (1994).

¹³ For example, in a model with treatment intensity in twelfth grade as the dependent variable, the coefficient of preprogram matriculation rate is not significantly different from zero (its value is 0.043 with a standard error of 0.169), and the coefficient of school size is -0.00031 with a standard error of 0.00005.

Table 5
The Relationship between Program Intensity and Preprogram Outcomes

Dependent Variable	1999 Matriculation Rate (1)	1999 School Size (2)	2000 Matriculation Rate (3)	2000 School Size (4)	Lagged Average Score (5)	School Size (6)	Lagged Credit Units Received (7)	School Size (8)
All treated schools ($N = 40$):								
Treatment intensity in twelfth grade	.043 (.169)	-.00031 (.00005)	...	...	.003 (.004)	-.00033 (.00005)	.003 (.013)	-.00033 (.00005)
Treatment intensity in tenth grade through twelfth grade	-.102 (.116)	-.00025 (.00004)	...	...	.000 (.002)	-.00030 (.00004)	.006 (.006)	-.00029 (.00004)
Schools enrolled in the program in 2001 ($N = 23$):								
Treatment intensity in twelfth grade	.110 (.237)	-.00032 (.00008)	.175 (.208)	-.00036 (.00008)	.004 (.005)	-.00032 (.00006)	.004 (.015)	-.00032 (.00006)
Treatment intensity in tenth grade through twelfth grade	-.008 (.134)	-.00025 (.00004)	-.043 (.104)	-.00028 (.00005)	.001 (.003)	-.00027 (.00004)	.004 (.008)	-.00026 (.00004)

NOTE.—The table presents results of four distinct regressions of treatment intensity on a preprogram outcome and school size. Two alternative measures of treatment intensity are used as the dependent variable: treatment intensity in the twelfth grade and treatment intensity in tenth through twelfth grade. The parameters reported in cols. 1 and 2 are from a regression of treatment intensity on school matriculation rate in 1999 and school size; those reported in cols. 3 and 4 are from a similar regression based on regressors from the year 2000. The parameters reported in cols. 5 and 6 are the estimates from regression of treatment intensity on average lagged score of the treated cohort and school size, and those reported in cols. 7 and 8 are based on a regression of treatment intensity on lagged number of credit units and school size. For results based on all treated schools, the school preprogram outcomes are taken from 1999 only. For results based on schools that were added to the program in 2001, the school preprogram outcomes are from 1999 and 2000. All models are based on aggregated data at the school level weighted by cohort size. Robust standard errors are reported in parentheses.

manent differences among schools that might be associated with variations in program intensity. The empirical models are stated as follows:

$$\bar{Y}_{st} = \beta' \bar{X}_{st} + \omega' S_{st} + \varphi_s + \delta_t + \alpha D_{st} \text{INT}_{12st} + \bar{\varepsilon}_{st}, \quad (4)$$

$$\bar{Y}_{st} = \beta' \bar{X}_{st} + \omega' S_{st} + \varphi_s + \delta_t + \alpha D_{st} \text{INT}_{st} + \bar{\varepsilon}_{st}, \quad (5)$$

B. Empirical Results

Table 6 presents the controlled differences in the matriculation rate between treated and comparison schools for both pretreatment (1999) and treatment (2001) years, as well as the difference-in-differences estimates. Columns 1–3 are based on specifications that use treatment status as a binary variable. Columns 4–6 are based on treatment intensity in twelfth grade, and columns 7–9 are based on average treatment intensity in grades 10–12.

In the preprogram period (1999), treated schools had a slightly lower matriculation rate than comparison schools, although the difference was not significantly different from zero for any of the specifications. The controlled differences for the postprogram period (2001) show that treated schools did better than comparison schools, suggesting a positive impact of the program; however, the estimates are significantly different from zero only for the specifications based on the treatment intensity indicators. Adjusting for preprogram differences and adding school fixed effects by estimating difference-in-differences increases the precision of the estimates. The difference-in-differences results suggest that the program boosted the matriculation rate of treated schools by 3 or 4 percentage points. The three different specifications based on alternative measures of treatment (treatment as a binary indicator, treatment intensity in twelfth grade, and treatment intensity in grades 10–12) provide similar results. For example, based on weighted aggregate school data (row 2), the estimated effect of treatment measured as a binary variable is 0.033 (SE = 0.014). The estimated effect based on treatment intensity in twelfth grade is 0.117 (SE = 0.050), which, when evaluated at the average intensity (0.279), provides an estimate of the average program effect equal to 0.033. The estimated effect based on average treatment intensity in grades 10–12 is 0.186 (SE = 0.078), implying at average intensity (0.203) an effect equal to 0.038. Estimating these models based on the micro data yields qualitatively similar results, though on average the point estimates suggest a somewhat lower program effect.

It is worth noting that the estimates based on treatment intensity not only are a result of rescaling the binary treatment variable but also reflect the fact that within treated schools, a higher intensity of treatment led to a greater improvement in the matriculation rate. Table A2 in the appendix

Table 6
The Estimated Effect of Additional Targeted Individualized Instruction on the Matriculation Rate, Treatment Status Defined at the School Level

	Treatment Regressor								
	Binary Indicator			Intensity in Twelfth Grade			Intensity in Tenth Grade through Twelfth Grade		
	1999	2001	Difference-in-	1999	2001	Difference-in-	1999	2001	Difference-in-
	(1)	(2)	Differences	(4)	(5)	Differences	(7)	(8)	Differences
Unweighted grouped data	-.018 (.022)	.032 (.026)	.040 (.021)	-.018 (.050)	.123 (.064)	.155 (.073)	-.042 (.070)	.149 (.085)	.237 (.104)
Weighted grouped data	-.028 (.023)	.018 (.025)	.033 (.014)	-.045 (.054)	.133 (.066)	.117 (.050)	-.070 (.068)	.161 (.093)	.186 (.078)
Micro data	-.005 (.021)	.020 (.024)	.026 (.010)	.018 (.046)	.123 (.059)	.083 (.036)	.015 (.060)	.156 (.085)	.127 (.054)
Number of schools	74	74	148	74	74	148	74	74	148
Number of students	12,235	11,773	24,008	12,235	11,773	24,008	12,235	11,773	24,008

NOTE.—The table reports estimates of the difference in the matriculation rate between treated and comparison schools in 1999 (pretreatment year), 2001 (posttreatment year), and difference-in-differences estimates that include year and school fixed effects. Three alternative measures of treatment status are used in the regressions: treatment as a binary variable (cols. 1–3) and two versions of treatment intensity. Columns 4–6 use the proportion of twelfth-grade students receiving treatment (average 0.277); cols. 7–9 use the proportion of tenth grade through twelfth grade students receiving treatment (average 0.203). The regressions include the following control variables: gender, immigration status, parental schooling, number of siblings, two measures of pretreatment outcomes indicating the quartile of a student's average score in matriculation exams taken before the twelfth grade and the quartile of the number of credit units received before the twelfth grade, school's religious status, school size, its square, and the number of students in the twelfth grade. Estimates based on unweighted and weighted grouped data at the school level are presented in rows 1 and 2, respectively. The weighted regressions are weighted by cohort size. Estimates presented in row 3 are based on micro data at the student level. Robust standard errors are reported in parentheses. Standard errors for the estimates based on micro data are corrected for clustering at the school-year level.

shows the difference-in-differences estimates, with the treatment intensity indicators (intensity in twelfth grade and in grades 10–12) broken down into three dummy variables denoting high, medium, and low treatment intensity.¹⁴ The estimates suggest that the higher the proportion of students participating in the program (higher treatment intensity), the greater the improvement in the school mean matriculation rate. For example, the estimates for low, medium, and high intensity in twelfth grade based on the weighted aggregate school data are 0.022, 0.031, and 0.064, respectively.

As noted earlier, 11 treated and 11 comparison schools participated in two additional programs implemented at the same time. In order to verify that the results from the school-level analysis reported above are not confounded by these interventions, we estimated the same models that are presented in table 6 while adding controls for school participation in either of these programs. Results, shown in table A3 in the appendix, are almost identical to those presented in table 6, which suggests that the increase in the matriculation rate among treated schools can be attributed to the Bagrut 2001 program.

C. Treatment Intensity: Instrumental Variables Strategy

The negative correlation between treatment intensity and school size gives us an opportunity to implement an alternative identification strategy. The larger a school is, the lower its level of treatment intensity. However, since school size itself may have an independent direct effect on school achievements, it cannot be used simply as an instrument of treatment intensity. This difficulty can be overcome by estimating the direct effect of school size on school achievements based only on data from the untreated cohorts (comparison schools in 1999 and 2001 and from treated schools in 1999). That is, the untreated cohorts enable us to partial out the direct effect of school size on school achievements.¹⁵ The instrument

¹⁴ The limits for low, medium, and high treatment intensity are defined by splitting the treated schools into three groups according to treatment intensity in twelfth grade or treatment intensity in grades 10–12. Low intensity in twelfth grade includes schools in the first quartile of the distribution (intensity lower than 20%); medium intensity in twelfth grade includes schools in the second and third quartiles (intensity between 20% and 40%); and high intensity in twelfth grade includes schools in the fourth quartile (intensity greater than 40%). The dummy variables for treatment intensity in grades 10–12 are defined similarly. Low intensity in grades 10–12 includes schools with intensity lower than 15%, medium intensity refers to the 15%–30% range, and high intensity includes schools with intensity greater than 30%.

¹⁵ This can be viewed as a two-step procedure. In the first step, we simply estimate the effect of school size on school achievements using only the pretreatment data (1999) for treatment and comparison schools and the 2001 data of the comparison schools. In the second step, we then use this estimate to partial out the effect of school size from the 2001 achievement variable in treated schools. In practice these two steps are estimated simultaneously.

of treatment intensity is then defined by the interaction of school size with the year 2001 dummy and treatment status. To allow for a more flexible relationship between treatment intensity and school size, we split the instrument into three variables based on the distribution of school size: the first corresponds to the first quartile, the second to the second and third quartiles, and the third to the fourth quartile. Each of these variables includes actual school size in its segment of the distribution and assigns zero values to the other two parts of the school size distribution. The following equations make up the first stage and the reduced-form structure of this instrumental variable strategy:

$$\begin{aligned} \text{INT}_{st} = & \beta' \bar{X}_{st} + \omega' S_{st} + \varphi_s + \delta_t + \gamma_1 \text{small}_{st} + \gamma_2 \text{med}_{st} + \gamma_1 \text{large}_{st} \\ & + \alpha_1 \text{inst1}_{st} + \alpha_2 \text{inst2}_{st} + \alpha_3 \text{inst3}_{st} + \bar{\epsilon}_{st}, \end{aligned} \quad (6)$$

$$\begin{aligned} \bar{Y}_{st} = & \beta' \bar{X}_{st} + \omega' S_{st} + \varphi_s + \delta_t + \gamma_1 \text{small}_{st} + \gamma_2 \text{med}_{st} + \gamma_1 \text{large}_{st} \\ & + \alpha_1 \text{inst1}_{st} + \alpha_2 \text{inst2}_{st} + \alpha_3 \text{inst3}_{st} + \bar{\epsilon}_{st}, \end{aligned} \quad (7)$$

where INT_{st} denotes the school's average treatment intensity, which is equal to zero for comparison schools in 1999 and 2001 and for treated schools in 1999 and has nonzero intensity values for treated schools in 2001; small_{st} , med_{st} , and large_{st} are the three school size variables described above; and the three instrumental variables are defined as follows: $\text{inst1}_{st} = \text{small}_{st} \times \text{treated}_s \times I(\text{year}=2001)$; $\text{inst2}_{st} = \text{med}_{st} \times \text{treated}_s \times I(\text{year}=2001)$; and $\text{inst3}_{st} = \text{large}_{st} \times \text{treated}_s \times I(\text{year}=2001)$, where $I(\text{year}=2001)$ denotes the treatment year. Similar equations are estimated on the basis of twelfth-grade treatment intensity.

The first-stage, reduced-form, and 2SLS results are presented in table 7. Columns 1 and 4 show the estimates of the relationship between school size and treatment intensity in twelfth grade on the basis of unweighted and weighted regressions, respectively, and columns 2 and 5 show the estimates for treatment intensity in grades 10–12. Treatment intensity (either in twelfth grade or in grades 10–12) seems to decrease monotonically as a function of school size. The same pattern is found in the relationship between school size and matriculation rate, although the estimates are less precise. In smaller schools (which, on average, had higher treatment intensity), the program appears to have caused a larger increase in the matriculation rate in the treatment year. The 2SLS estimates are slightly higher than the estimates obtained from the differences-in-differences strategies, but nevertheless they are consistent with the results presented above. For example, the estimated effect of treatment intensity in twelfth grade obtained from the weighted regression is 0.154 (SE = 0.054) and the corresponding estimate based on treatment intensity in grades 10–12 is 0.202 (SE = 0.067).

Table 7
First-Stage, Reduced-Form, and Two-Stage Least Squares Estimates of the Effect of Treatment Intensity

	Unweighted			Weighted		
	Treatment Intensity in Twelfth Grade (1)	Treatment Intensity in Tenth through Twelfth Grades (2)	Matriculation Rate (3)	Treatment Intensity in Twelfth Grade (4)	Treatment Intensity in Tenth through Twelfth Grades (5)	Matriculation Rate (6)
	First Stage	First Stage	Reduced Form	First Stage	First Stage	Reduced Form
Instruments:						
Small schools (size × year 2001 × treated)	.00229 (.00023)	.00186 (.00023)	.00056 (.00024)	.00220 (.00024)	.00178 (.00024)	.00051 (.00022)
Medium schools (size × year 2001 × treated)	.00053 (.00008)	.00040 (.00005)	.00003 (.00004)	.00050 (.00008)	.00038 (.00005)	.00004 (.00003)
Large schools (size × year 2001 × treated)	.00028 (.00006)	.00016 (.00004)	.00003 (.00003)	.00024 (.00006)	.00013 (.00004)	.00003 (.00002)
<i>F</i> (3, 54)	50.14	50.06	2.04	41.98	49.48	3.13
	2SLS	2SLS		2SLS	2SLS	
2SLS estimates	.174 (.074)	.228 (.088)		.154 (.054)	.202 (.067)	

NOTE.—Columns 1, 2, 4, and 5 of the top portion of the table present estimates from a first-stage regression for two alternative endogenous regressors: treatment intensity in the twelfth grade and treatment intensity in the tenth grade through twelfth grade. Columns 3 and 6 of the top portion of the table present estimates of the reduced form. The instruments used in the regressions are the interactions of school size with year 2001 dummy and treatment status at three different points of the school-size distribution: the first quartile of the school-size distribution (small schools), the second and third quartiles (medium-size schools,) and the fourth quartile (large schools). The reported *F*-statistics corresponds to a test of a joint significance of the three instruments. The lower portion of the table presents the 2SLS estimates. The regressions include school and year fixed effects and controls for student's and school's characteristics reported in table 6 and for three variables for school size at different points of the school size distribution. The models are based on aggregated data at the school-year level. Estimates presented in cols. 4–6 are weighted by cohort size. Robust standard errors are reported in parentheses.

V. Effects on Treated Students and Spillover Effects on Nonparticipants

Our next task is to identify the program effect on the treated students. The average school effect may be broken down into two parts: the effect on the participants and the effect on the participants' peers through externalities and spillover.¹⁶ Since there is no natural comparison group for the treated students, we resort to a model where selection is based on observables. The central identifying assumption in this case is that treatment status is independent of potential achievements conditioned on observable covariates and lagged achievements. Although this assumption is hardly satisfied in most nonexperimental studies, the availability of a wealth of data on student characteristics and student achievement on matriculation exams taken before the program was implemented make this a credible approach in our case.¹⁷ As Smith and Todd (2005) note, a critical condition for reducing selection bias is to have data containing a rich set of variables that affect both program participation and student achievement.

A comparison group was selected from among the students in the comparison schools. We avoided choosing students from treated schools because, first, they might have been exposed to treatment through potential externalities of the program and second, they might be different from program participants in observable and unobservable characteristics, as hinted by the fact that they were not chosen for the program.

As a benchmark we first estimated OLS regressions in which we controlled for the following student and school covariates: gender, ethnic origin, parents' education, number of siblings, immigration status, number of matriculation credits earned in grades 10 and 11, average weighted

¹⁶ Students in treated schools who were not in the program could have been affected in various ways. For example, the program may have freed some of the classroom teacher's time, which could then be directed to untreated students. Another possibility is that the improved learning of program participants reduced the frequency of disruptions during regular class sessions. On the other hand, negative externalities may have ensued if the close relationship between teachers and treated students that developed during treatment resulted in attention in regular classes being diverted from nonparticipants to participants.

¹⁷ The claim that the program itself may have changed the students' behavior and achievement on the examinations taken before participation in the program is not a concern in this case for two reasons: first, for most schools, it was the first year of implementation of the program, making it highly unlikely that the students were aware of the program before they were given the opportunity to participate. Second, the preprogram achievements are based on the matriculation examinations, which have a strong influence on opportunities for higher education and jobs. Thus, it does not stand to reason that students would risk future opportunities and underperform on these exams in order to be admitted to the program.

score on these credits, school size, its square, twelfth-grade enrollment, religious status of school, and school matriculation rate in 1999. Although this method controls for a large set of covariates and preprogram achievements, we might have been comparing dissimilar populations, since the achievements for treated students would be contrasted with the achievements for the whole sample of students from comparison schools.¹⁸ An alternative approach is to implement a nonparametric matching method, as Angrist (1998) did. This would assure that the subsets of students that form the treated and comparison groups have similar distributions of covariates. It also reduces the need to rely on model extrapolations and functional form. However, this procedure has a major problem in that as the number of covariates increases, the probability that both treated and comparison students will have the same set of covariates approaches zero. Rosenbaum and Rubin (1983) proposed an alternative approach that circumvents the curse of dimensionality. They stated that if treatment assignment can be ignored given x , then it can be ignored given any balancing score and, in particular, given a propensity score. We applied various methods based on the propensity score (nearest neighbor matching, kernel matching, stratification on the propensity score, weighted regression on the basis of the score, etc.), as well as nonparametric matching, in order to contrast the achievements for treated students with those for students from comparison schools and to estimate the effects of the program.¹⁹ The propensity score was estimated on the basis of all the aforementioned student covariates, preprogram achievements, and school characteristics.²⁰

A point to emphasize is that the matching of the propensity score is not meant to emulate the selection process used in the program; it is based

¹⁸ Because some of the covariates are nondiscrete, OLS estimates will be based on the whole sample even if for some values of the covariates there are only treated or comparison students.

¹⁹ The results reported in this article are based on nearest neighbor matching and kernel matching. The other methods provide very similar results.

²⁰ We tried various specifications in which we omitted one of the preprogram achievements from the propensity score equation and checked whether the matched sample was balanced for this covariate. The similar achievement of treated and comparison students in preprogram data omitted from the propensity score equation reinforces the credibility of the assumption that treated and comparison students would have done similarly afterward, too, if the treated students had not been treated. The standard errors of the estimates of the program effects were computed using two-stage bootstrap methods that preserve the clustered structure of the data. In the first stage, we sampled schools with replacement; in the second stage we sampled students within schools. For each generated sample, we estimated a propensity score function, and we then applied matching procedures to generate the estimates. The standard errors were derived from the results obtained in these replications. For the theoretical context, see Davison and Hinkley (1997); for an application of this procedure, see Subramanian and Deaton (1996).

mainly on the same central assumption of selection on observables that underlies the linear regression. The main purpose of propensity score matching is to restrict the nonexperimental comparison group to a sample that has the same distribution of covariates as the treated students. Table A4 in the appendix compares the characteristics of treated and comparison students matched by nearest neighbor matching.²¹ A total of 1,183 students from comparison schools were matched with 1,789 treated students. We were able to find a match for every treated student. The matched comparison students were selected from all comparison schools, reinforcing the argument that treated and comparison schools have similar characteristics. The comparison of treated and matched students shows no significant differences between the two groups in any of the student or school covariates.

To identify the program effects on nontreated students, we applied the same methodology described for treated students. Nontreated students are compared to students from comparison schools, using linear regressions and alternative matching methods.

A. Empirical Results

Table 8 presents the separate estimates of the program effects on treated students and their classmates. The estimates for program participants are reported in columns 1–3, and the estimates for nonparticipants are reported in columns 4–6. The alternative methods suggest that the program affected the achievements of treated students only, increasing their probability of earning a matriculation certificate by 13 percentage points on average. The effect of the program declines monotonically with student ability (measured on the basis of preprogram achievements). For example, the effect on the lowest ability quartile is twice that on the third ability quartile. However, this pattern of sharp monotonic decline in the effect of the program by student ability may be a result of inappropriate matching of students. An illustration may help clarify this point. Some students may have been selected to participate in the program despite being in the upper ability quartile (having high lagged achievements) because the school staff knew something about them that was not known to the econometrician (unobservable characteristics). These students will have low propensity scores and will be matched with comparison students who are truly high performers. Such circumstances may lead to a correlation between the propensity score estimates and pre- and postprogram achieve-

²¹ We first limited the sample to those treated and control students having propensity scores in the region of common support. Each treated student was matched with a student from comparison schools that had the nearest propensity score. Some comparison students were matched to more than one treated student. Nevertheless, most comparison students were used a maximum of two times.

Table 8
The Estimated Effect of Additional Targeted Individualized Instruction on
the Matriculation Rate: Student Level Analysis

	Program Participants			Nonparticipants		
	OLS (1)	Nearest Neighbor Matching (2)	Kernel Matching (3)	OLS (4)	Nearest Neighbor Matching (5)	Kernel Matching (6)
Treatment effect	.132 (.026)	.139 (.047)	.132 (.035)	.007 (.016)	.007 (.044)	.002 (.037)
Treatment effects by ability quartiles:						
Quartile 1 (lowest)	.217 (.050)	.215 (.065)	.218 (.060)	.001 (.015)	-.017 (.030)	-.009 (.022)
Quartile 2	.164 (.037)	.163 (.065)	.156 (.046)	.013 (.030)	.009 (.050)	.006 (.039)
Quartile 3	.103 (.030)	.101 (.065)	.091 (.046)	.016 (.027)	.010 (.059)	.007 (.044)
Quartile 4	.070 (.028)	.038 (.058)	.036 (.041)	.001 (.018)	.003 (.044)	.000 (.030)

NOTE.—The table reports estimated treatment effects on the student matriculation status for program participants (cols. 1–3) and for students that studied in treated schools but did not participate in the program (cols. 4–6). Row 1 reports the mean treatment effect and rows 2–5 report treatment effects by quartiles of the student's average score in matriculation exams taken before twelfth grade. Columns 1 and 4 present OLS estimates from models that control for student's and school's characteristics reported in table 6. Estimates for cols. 2, 3, 5, and 6 are based on nearest neighbor matching and kernel matching methods and were computed using Stata procedures developed by Becker and Ichino (2002). The following variables were used for the estimation of the propensity score: gender, immigration status, parental schooling, number of siblings, two measures of pretreatment outcomes indicating the quartile of a student's average score in matriculation exams taken before the twelfth grade and the quartile of the number of credit units awarded before the twelfth grade, school matriculation rate in 1999, school's religious status, school size, its square, and the number of students in the twelfth grade. Standard errors (in parentheses) are corrected for clustering at the school level in OLS regressions. Standard errors (in parentheses) for matching methods are computed using two-stage bootstrap methods that preserve the clustered structure of the data.

ments. In figures 6 and 7 we plotted the pre- and posttreatment achievements of treated students as a function of their propensity scores. The achievements are averaged according to the percentile of the propensity score. Figure 6 suggests that there is no clear relationship between student performance in grades 10 and 11 and the estimated propensity score. The same pattern appears in figure 7, where the matriculation rate is plotted against the propensity score percentiles. The matriculation rate of students with low estimated propensity scores is not lower than that of students with higher propensity scores. This evidence suggests that the propensity score estimates did not suffer from such bias and that the decline observed in the effect of the program by student ability is not erroneous.

The evidence in table 8 also suggests that the program did not affect the achievements of participants' classmates. This enhances the credibility of the results obtained in the analysis of the program effects on treated students.²² Assuming that the program did not have spillover effects, as

²² This result is also in line with the results reported in table A3, which suggest that the increase in the matriculation rate is due to the Bagrut 2001 program and not to the two other interventions, the teachers and students bonuses programs.

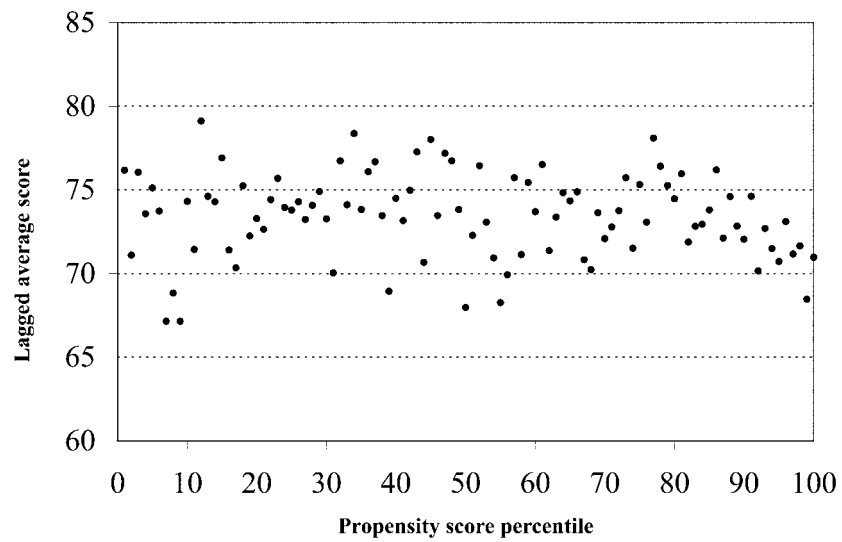


FIG. 6.—Lagged average score by propensity score percentiles

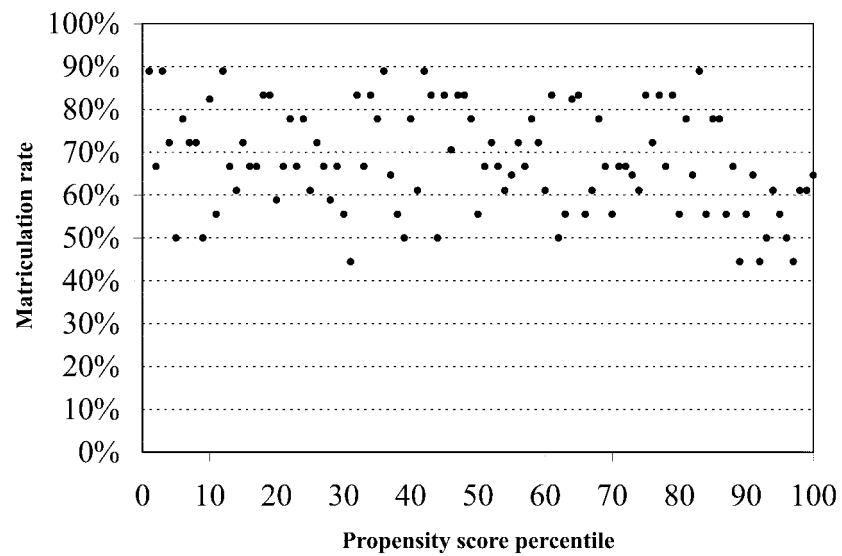


FIG. 7.—Matriculation rate by propensity score percentiles

the results indicate, we may return to the analysis of the program effects at the school level and obtain an estimate of the program effects on treated students.²³ Taking, for example, the estimate of the program effect obtained from the weighted regression based on aggregate school data (0.033) and dividing it by the proportion of treated students (0.277), we obtain a 0.119 estimate of the program effect on treated students. This estimate is very similar to the alternative estimates obtained from the analysis based on treated students.

VI. Estimates Based on Difference-in-Differences Matching Estimators

The availability of preprogram achievements allows us to apply an additional method to test for program effects. The use of difference-in-differences matching estimators lets us control for any time-invariant unobservable differences between treated and comparison students.²⁴ An example may be constant differences in school characteristics between treated and comparison schools. The assumption required for identification is that the difference between postprogram and preprogram achievements for comparison students provides the counterfactual for the treated students.

Two measures of achievements are available for the preprogram and postprogram periods. The first is the number of matriculation credits earned. Since we have seen that the program increases the matriculation rate and since a minimum of 20 credits is required to qualify for a matriculation certificate, we should expect to find a positive effect on mean credits earned by treated students. The second measure is the average score on matriculation exams. For this one, however, the effect may go either way.²⁵

The top portion of table 9 presents the comparison between the treated group and the matched comparison group. There are no differences between the two groups in mean preprogram achievements, such as the

²³ This is equivalent to a Wald estimator that can be stated as the sample analog of the following expression:

$$\alpha_{iv} = \frac{\{E[Y|T = 1, \text{year} = 2001, X] - E[Y|T = 1, \text{year} = 1999, X]\} - \{E[Y|T = 0, \text{year} = 2001, X] - E[Y|T = 0, \text{year} = 1999, X]\}}{\{E[S|T = 1, \text{year} = 2001, X] - E[S|T = 1, \text{year} = 1999, X]\} - \{E[S|T = 0, \text{year} = 2001, X] - E[S|T = 0, \text{year} = 1999, X]\}},$$

where $T = 1$ if the school was treated and $S = 1$ if the student was treated. The denominator can be reduced to $P[S = 1 | T = 1, \text{year} = 2001, X]$, which is equal to the proportion of treated students in treated schools, since the other elements are zero.

²⁴ For the theory and an application of this estimator, see Heckman, Hidehiko, and Todd (1997) and Smith and Todd (2005).

²⁵ On the one hand, treated students may have improved their scores; on the other hand, they may have been tested at higher levels of proficiency than comparison students.

Table 9
Difference-in-Differences Matching Estimates

	Preprogram Outcomes in Tenth and Eleventh Grades		Postprogram Outcomes in Twelfth Grade		Difference-in-Differences (5)
	Mean (1)	Difference (2)	Mean (3)	Difference (4)	
Treated versus matched comparison students (<i>N</i> = 3,578):					
Number of credit units received	6.791	.052 (.383)	13.357	2.003 (.623)	1.951 (.773)
Weighted average score	71.933	1.532 (.973)	68.628	2.926 (1.243)	1.255 (1.124)
Untreated versus matched comparison students (<i>N</i> = 9,318):					
Number of credit units received	5.447	-.001 (.449)	12.157	-.827 (.578)	-.830 (.559)
Weighted average score	65.737	1.389 (2.399)	64.024	-.016 (1.975)	-1.484 (1.394)

NOTE.—The table reports estimates of the difference in Bagrut outcomes between students from treated schools and matched students from comparison schools for the preprogram period (col. 2) and the postprogram period (col. 4) and difference-in-differences estimates (col. 5). Columns 1 and 3 present the outcomes' means. The top portion of the table reports estimates for treated students, and the bottom portion of the table reports estimates for students from treatment schools that did not participate in the program (untreated students). The matched comparison group for treated and untreated students was selected applying nearest neighbor matching. The variables used for the estimation of the propensity score are identical to those reported in table 8. Standard errors (in parentheses) are corrected for clustering at the school level.

number of credits in grades 10–11. However, we find a positive difference of two credit units ($SE = 0.623$) for treated students in the postprogram period. The differences-in-differences estimate is almost the same: 1.951 ($SE = 0.773$). No similar gain was found for the average score on the matriculation exams. This suggests that program participants gained on average two additional credits without lowering their average score. The bottom portion of table 9 presents a similar analysis based on nontreated students. The results suggest that the program did not have any effect (positive or negative) on the achievements of this group.

VII. Cost-Benefit Analysis and Alternative Interventions

The average cost of the program was \$1,100 per participant and \$300 per twelfth-grade student. We focus the cost-benefit analysis on the latter because it allows comparison with other school-based interventions and because the estimates of program effects at the school level lack the potential pitfalls of the estimates derived from the sample of program participants.

An ideal cost-benefit analysis should contrast the aforementioned cost with student achievements after graduation from high school. Unfortunately, we will have to wait many years to pursue this analysis. As an alternative, however, we may estimate the potential monetary returns on an increase in the matriculation rate in terms of the present value of the

Table 10
Comparison of Alternative Programs, Costs, and Benefits

Program	Cost per Student	Absolute Gain (in Percentage Points)	Pretreatment Bagrut Rate (%)
Matriculation awards (Angrist and Lavy 2004)	385	7	19
Teacher bonuses (Lavy 2004)	170	4	41
Bagrut 2001	300	3.3	44

NOTE.—The cost per student is in terms of 2001US\$.

increment in lifetime earnings.²⁶ Angrist and Lavy (2004) estimated the returns to the Bagrut certificate at 24%. This rate of return reflects partly the increased likelihood of higher education, but also the monetary returns to the Bagrut certificate itself.²⁷ Given that the program boosted the average school matriculation rate by 3.3 points and taking as a baseline the earnings of workers with 11–12 years of schooling in the year 2000, which were about \$16,100 (Central Bureau of Statistics 2002), the annual increment in earnings due to the program is estimated as follows: $\$16,100 \times 0.24 \times 0.033 = \127.5 . Assuming that this increment remains constant across lifetime earnings and that the positive payoffs begin 4 years after the costs are incurred and remain in effect for 43 years (until the individual reaches age 65), the internal rate of return is 20%. This exceeds any reasonable real rate of interest.

Although it seems that the program yields a positive rate of return, we should ask how it ranks in terms of efficiency relative to other potential interventions. To address this question, we compared the remedial education program with other programs that had the same objectives and were implemented in Israel in the same academic year. Table 10 presents a comparative analysis of Bagrut 2001, a student matriculation awards program that provided monetary bonuses to students who earned matriculation certificates (Angrist and Lavy 2004) and a teacher-bonus program that paid teachers bonuses on the basis of their students' performance on matriculation exams (Lavy 2004). The Bagrut 2001 program produced a gain similar to that of the teacher bonus program but at almost twice the per student cost. The matriculation awards program doubled the improvement achieved by the Bagrut 2001 program while its costs were almost equal. However, the matriculation awards program was implemented in much weaker schools, somewhat limiting the relevance of the

²⁶ We assume that the program affects only the wages of its participants and has no general equilibrium effects. This assumption seems plausible, since less than 1.5% of students in their age group received treatment. This point will have to be taken into account in the future as the program is expanded to a large proportion of Israeli schools.

²⁷ Angrist and Lavy (2004) report that even in a sample limited to those with exactly 12 years of schooling, Bagrut holders earn 13% more.

comparison. A more complete evaluation should also consider the long-term effects, including the effect on completion of higher education, employment, and even earnings.

VIII. Conclusions

In this article we have studied the effects of a relatively traditional and somewhat forgotten intervention—targeted additional instruction time—on the quality of schooling. This intervention can be viewed as a combination of two alternative policies targeted toward disadvantaged students: additional instruction time and smaller classes. The results suggest that this remedial education program led to a 3–4 percentage point increase in the school mean matriculation rate, which represents an 11–12 percentage point increase in the matriculation rate among participating students—an improvement of 22%. This estimate of the effect of treatment on treated students represents a relatively large gain in terms of empirical educational production functions. However, it comes at a relatively high cost because the program expenditure per participant was about 40% of the annual expenditure per high school student in Israel. The cost-benefit comparison with other relevant programs that are based on achievement incentives for teachers or students indicates that the remedial program is less effective. However, a comparison of the yield in terms of future earnings and program costs suggests a high rate of return.

Three policy implications of these results should be noted. First, in the context of early versus late childhood interventions, the results imply that for a not too outrageous cost some disadvantaged or underperforming teenagers can be helped to earn matriculation certificates, an important educational achievement with a high payoff for postsecondary schooling and success in the labor market. Remedial education targeted at adolescents while still in school appears to be a relatively successful intervention, especially given the low effectiveness of most high school dropout prevention programs in the United States and also the low rate of return to a GED certification (Heckman and Rubinstein 2001).

It is not possible to perform a rigorous comparison of the returns of the Bagrut 2001 program relative to early childhood interventions. However, a contrast of costs and benefits of preschool interventions with Bagrut 2001 can provide an indication for the relative magnitude of the returns.²⁸ Two of the most cited early childhood programs are the Perry

²⁸ First, as there is no study of the long-run effects of any early childhood intervention in Israel, we have to resort to comparisons with U.S. programs. Second, each study used different methodologies and assumptions for determining the program costs and evaluating the program impacts and used different discount rates. In addition, they have estimated benefits by monetizing not exactly the same social impacts.

Preschool Project and the Carolina Abecedarian Project. The evaluation of these randomized interventions included also estimates of their long-term effects. The cost-benefit analysis of the Perry Preschool Program reveals a \$7 return per dollar spent (Karoely et al. 1998). A more recent study based on a follow-up to age 40 suggests a return of \$17.07 per dollar invested (Schweinhart 2004). The study of the Carolina Abecedarian Project reports returns of about \$3.78 per dollar spent (Masse and Barnett 2002). Not having estimates of the long-term effects of the Bagrut 2001 program, we adopted a conservative approach, limiting the assessment of benefits to predicted increment in lifetime earnings. This approach led to an estimated return of \$7.40 per dollar spent; this ratio is within the range of the returns of the Perry Preschool and the Abecedarian projects.²⁹

The second policy implication of this article's results arises in the context of the current drive to reduce class size. The Bagrut 2001 remedial education program presents an effective alternative to class size reduction by providing individualized instruction to those who need it most.³⁰ Most studies that were able to identify a significant effect of class size (e.g., Angrist and Lavy 1999; Krueger 1999) show that class size reduction has an effect mainly on minorities and underachieving students. Therefore, it may be more effective to target additional resources directly to these groups instead of reducing class size for everyone. The relative unattractiveness of class size reduction is amplified by the increased demand for teachers that it implies and the consequences in terms of average teaching quality (Jepsen and Rivkin 2002), a constraint less relevant to targeted remedial education.

Finally, the results presented here on targeted remedial education are consistent with findings that point to summer schooling as a very effective intervention (Cooper et al. 2000; Jacob and Lefgren 2004). Summer schooling is generally targeted and has recently been gaining popularity in the United States and Europe as a form of remedial education.³¹ Both interventions—summer schooling and remedial education during the school year—may be viewed as ways of providing additional instruction

²⁹ We discounted the costs and benefits of the Bagrut 2001 program to age 0 at a discount rate of 4%. It is worth noting that unlike the two preschool programs we do not have data on the long-term effects of the program.

³⁰ Although most reductions in class size target preschool and early grades, some U.S. states have recently initiated reforms aimed at reducing class size in high schools. Unfortunately, we were not able to find any evidence about costs and benefits of class size reduction in high school.

³¹ Many states, including Arkansas, Delaware, Maryland, and West Virginia, have initiated summer schooling programs for disadvantaged groups. For a survey and more details, see http://www.sreb.org/programs/srr/pubs/Summer_School.pdf.

time to disadvantaged or underperforming students. The former attempts primarily to remedy failure; the latter is designed to prevent it.

Appendix

Table A1
Differences in Matriculation Rate Trends between Schools by Treatment Intensity for Pretreatment Years, 1996–99

Regressors	Unweighted		Weighted	
	Treatment Intensity in Twelfth Grade (1)	Treatment Intensity in Tenth through Twelfth Grades (2)	Treatment Intensity in Twelfth Grade (3)	Treatment Intensity in Tenth through Twelfth Grade (4)
Treatment intensity × 1997	-.202 (.138)	-.269 (.162)	-.147 (.161)	-.235 (.205)
Treatment intensity × 1998	.045 (.131)	.031 (.162)	.072 (.164)	.016 (.215)
Treatment intensity × 1999	-.052 (.121)	-.073 (.141)	.041 (.152)	.027 (.192)
Number of schools	67	67	67	67

NOTE.—The table reports the interaction coefficients between treatment intensity and year dummies. The dependent variable is the school matriculation rate. The additional regressors are year dummies and a treatment intensity indicator. Two alternative measures of treatment intensity are used: treatment intensity in twelfth grade and treatment intensity in tenth grade through twelfth grade. Data are aggregated at the school level. Estimates in col. 2 are weighted by cohort size. Robust standard errors are reported in parentheses. Seven schools missing in the test scores datasets for years 1996–98 are not included in the sample.

Table A2
Treatment Effects on the Matriculation Rate by Ranges of Treatment Intensity

Range of Treatment Intensity	Unweighted (1)	Weighted (2)	Micro Data (3)
Twelfth grade:			
Low (<20%)	.005 (.030)	.022 (.020)	.026 (.012)
Medium (20%–40%)	.030 (.020)	.031 (.016)	.026 (.012)
High (>40%)	.097 (.051)	.064 (.039)	.028 (.026)
Tenth grade through twelfth grade:			
Low (<15%)	.009 (.021)	.014 (.016)	.025 (.013)
Medium (15%–30%)	.032 (.021)	.040 (.017)	.022 (.013)
High (>30%)	.087 (.051)	.061 (.040)	.040 (.024)

NOTE.—The table reports controlled difference-in-differences treatment estimates by ranges of treatment intensity. The treatment intensity indicators are divided into three dummy variables: low, medium, and high. The school matriculation rate is the dependent variable in the regressions based on grouped data (cols. 1 and 2). The student matriculation status is the dependent variable for the regressions based on micro data (col. 3). Estimates in col. 2 are weighted by cohort size. The controls for student's and school's characteristics are the same as those used in table 6. Robust standard errors are reported in parentheses. Standard errors for the estimates based on micro data are corrected for clustering at the school-year level.

Table A3
Difference-in-Differences Estimates Conditioned on Control for School
Participation in Other Interventions

	Treatment as a Binary Indicator (1)	Treatment Intensity in Twelfth Grade (2)	Treatment Intensity in Tenth Grade through Twelfth Grade (3)
Unweighted grouped data	.041 (.021)	.163 (.070)	.244 (.098)
Weighted grouped data	.032 (.014)	.120 (.050)	.192 (.075)
Micro data	.026 (.010)	.084 (.036)	.127 (.054)
Number of schools	74	74	74
Number of students	24,008	24,008	24,008

NOTE.—The table reports difference-in-differences estimates using models similar to those presented in table 6 but with the addition of dummy indicators for the school participation during 2001 in the teacher bonus program (10 treatment and nine comparison schools) or in the student awards program (one treatment and two comparison schools). Weighted regressions are weighted by cohort size. Robust standard errors are reported in parentheses. Standard errors for the estimates based on micro data are corrected for clustering at the school-year level.

Table A4
Comparison of Mean Covariates and Outcomes for Treated and
Matched Students

	Treated Students (1)	Matched Students (2)	Difference (1 – 2) (3)
Student's characteristics:			
Father's years of schooling	10.37	10.57	–.20 (.43)
Mother's years of schooling	10.63	10.71	–.08 (.51)
Number of siblings	2.51	2.45	.07 (.24)
Immigration status (= 1 if im- migrated within the last 4 years)	.02	.01	.00 (.01)
Male	.42	.43	–.01 (.03)
Parents' ethnic origin:			
Asia-Africa (excluding Ethiopia)	.32	.32	.00 (.04)
Former Soviet Union	.20	.20	.01 (.05)
Ethiopia	.02	.02	–.00 (.01)
Europe-America (excluding for- mer Soviet Union)	.06	.06	–.00 (.01)
Israel	.40	.41	–.00 (.04)
Student's achievements:			
Average lagged score in matric- ulation exams (prior to twelfth grade)	73.43	71.91	1.53 (.97)

Table A4 (Continued)

	Treated Students (1)	Matched Students (2)	Difference (1 – 2) (3)
Number of credit units received prior to twelfth grade	6.84	6.79	.05 (.38)
Matriculation rate	.68	.55	.13 (.04)
School's characteristics:			
Enrollment in twelfth grade	178	174	4.02 (21.32)
Enrollment in tenth grade through twelfth grade	544	567	–23.1 (61.9)
Number of schools	40	34	
Number of students	1,789	1,183	

NOTE.—The table reports descriptive statistics for treated and matched comparison students. Difference of means between treated and matched students is reported in col. 3. Standard errors for the differences corrected for clustering at school level are reported in parentheses. Each treated student was matched with a student from comparison schools that had the nearest propensity score. Some comparison students were matched to more than one treated student.

References

- Angrist, Joshua D. 1998. Estimating the labor market impact of voluntary military service using social security data on military applicants. *Econometrica* 6, no. 2:249–88.
- Angrist, Joshua D., and Victor Lavy. 1999. Using Maimonides' rule to estimate the effect of class size on scholastic achievement. *Quarterly Journal of Economics* 114, no. 2:533–75.
- . 2001. Does teacher training affect pupil learning? Evidence from matched comparisons in Jerusalem public schools. *Journal of Labor Economics* 19, no. 2:343–69.
- . 2002. New evidence on classroom computers and pupil learning. *Economic Journal* 112, no. 482:735–65.
- . 2004. The effect of high stakes high school achievement awards: Evidence from a group randomized trial. Unpublished manuscript, Department of Economics, Massachusetts Institute of Technology and Hebrew University.
- Banerjee, Abhijit, Shawn Cole, Esther Duflo, and Leigh Linden. 2004. Remedying education: Evidence from two randomized experiments in India. Photocopy, Department of Economics, Massachusetts Institute of Technology.
- Becker, Sascha O., and Andrea Ichino. 2002. Estimation of average treatment effects based on propensity scores. *Stata Journal* 2, no. 4:338–77.
- Betts, Julian. 1995. Does school quality matter? Evidence from the National Longitudinal Survey of Youth. *Review of Economics and Statistics* 77, no. 2:231–50.
- Card, David, and Alan Krueger. 1992. Does school quality matter? Return

- to education and the characteristics of public schools in the United States. *Journal of Political Economy* 100, no. 1:1–40.
- . 1994. Minimum wages and employment: A case study of the fast food industry in New Jersey and Pennsylvania. *American Economic Review* 84, no. 4:772–93.
- Carneiro, Pedro, and James Heckman. 2003. Human capital policy. Working Paper no. 9495, National Bureau of Economic Research, Cambridge, MA.
- Central Bureau of Statistics. 2002. *Statistical abstract of Israel*, vol. 53. Jerusalem: Central Bureau of Statistics, Israel.
- Cooper, Harris, Kelly Charlton, Jeffrey C. Valentine, and Laura Muhl-enbruck. 2000. Making the most of summer school: A meta-analytic and narrative review. *Monographs of the Society for Research in Child Development* 65, no. 1:1–118.
- Currie, Janet. 2001. Early childhood education programs. *Journal of Economic Perspectives* 15, no. 2:213–38.
- Currie, Janet, and Duncan Thomas. 2001. Early test scores, school quality, and SES: Long-run effects on wage and employment outcomes. In *Worker wellbeing in a changing labor market*, Research in Labor Economics, vol. 20, ed. Solomon W. Polachek, 20:103–32. New York: Elsevier Science.
- Davison, Antony C., and David Hinkley. *Bootstrap methods and their application*. Cambridge: Cambridge University Press, 1997.
- Dynarski, Mark, and Philip Gleason. 1998. How can we help? What have we learned from evaluations of federal dropout-prevention programs? Policy Research Report no. 8014–140, Mathematica, Princeton, NJ.
- Garces, Eliana, Duncan Thomas, and Janet Currie. 2002. Longer term effects of Head Start. *American Economic Review* 92, no. 4:999–1012.
- Heckman, James J., Ichimura Hidehiko, and Petra Todd. 1997. Matching as an econometric evaluation estimator. *Review of Economic Studies* 64, no. 4:605–54.
- Heckman, James, and Yona Rubinstein. 2001. The importance of non-cognitive skills: Lessons from the GED Testing Program. *American Economic Review* 91, no. 2:145–49.
- Hoxby, Caroline M. 2000. The effects of class size on student achievement: New evidence from population variation. *Quarterly Journal of Economics* 115, no. 3:1239–85.
- Jacob, Brian A., and Lars Lefgren. 2004. Remedial education and student achievement: A regression discontinuity analysis. *Review of Economics and Statistics* 86, no. 1:226–44.
- Jepsen, Christopher, and Steven Rivkin. 2002. What is the tradeoff between smaller classes and teacher quality? Working Paper no. 9205, National Bureau of Economic Research, Cambridge, MA.
- Karoly, Lynn A., Peter W. Greenwood, Susan S. Everingham, Jill Houbé,

- M. Rebecca Kilburn, C. Peter Rydell, Matthew Sanders, and James Chiesa. 1998. *Investing in our children: What we know and don't know about the costs and benefits of early childhood interventions*. Santa Monica, CA: Rand.
- Koretz, Daniel. 1996. Using student assessments for educational accountability. In *Improving American schools*, ed. Eric A. Hanushek and Dale W. Jorgenson, 171–95. Washington, DC: National Academy Press.
- Krueger, Alan B. 1999. Experimental estimates of education production functions. *Quarterly Journal of Economics* 114, no. 2:497–532.
- Krueger, Alan B., and Diane M. Whitmore. 2001. The effect of attending a small class in the early grades on college-test taking and middle school test results: Evidence from Project STAR. *Economic Journal* 111, no. 468:1–28.
- Lavy, Victor. 2002. Evaluating the effect of teachers' group performance incentives on pupil achievement. *Journal of Political Economy* 110, no. 6:1286–1317.
- . 2004. Performance pay and teacher's effort, productivity, and grading ethics. Working Paper no. 10622, National Bureau of Economic Research, Cambridge, MA.
- Machin, Stephen, Sandra McNally, and Costas Meghir. 2004. Improving pupil performance in English secondary schools: Excellence in cities. *Journal of the European Economic Association* 2, no. 2:396–405.
- Masse, Leonard, and W. Steven Barnett. 2002. A benefit-cost analysis of the Abecedarian Early Childhood Intervention. National Institute for Early Education Research, New Brunswick, NJ.
- Ministry of Education, Culture, and Sport. 2001. *Facts and figures about education in Israel*. Jerusalem: Publications Department, Ministry of Education, Culture, and Sport.
- Rosenbaum, Paul R., and Donald B. Rubin. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika* 70, no. 1:41–55.
- Schinke, S. P., Kristin Cole, and Stephen R. Poulin. 2000. Enhancing the educational achievement of at-risk youth. *Prevention Science* 1, no. 1: 51–60.
- Schweinhart, Lawrence J. 2004. The High/Scope Perry Preschool Study through age 40: Summary, conclusions, and frequently asked questions. High Scope Educational Research Foundation, Ypsilanti, MI.
- Smith, Jeffrey, and Petra Todd. 2005. Does matching overcome Lalonde's critique of nonexperimental estimators? *Journal of Econometrics* 125, no. 1–2:305–53.
- Subramanian, Shankar, and Angus Deaton. 1996. The demand for food and calories. *Journal of Political Economy* 104, no. 1:133–62.
- U.S. Department of Education, Office of the Under Secretary. 2003. *When*

schools stay open late: The national evaluation of the Twenty-First Century Community Learning Centers Program, first year findings. U.S. Department of Education, Office of the Under Secretary, Washington, DC.