

Targeted Remedial Education for Under-Performing Teenagers:

Costs and Benefits*

Victor Lavy

The Hebrew University of Jerusalem, CEPR and NBER

and

Analia Schlosser

The Hebrew University of Jerusalem

May 2004

* We would like to thank Daron Acemoglu, Joshua Angrist, Esther Duflo, Daniele Paserman and seminar participants at the Hebrew University and MIT for their helpful comments and suggestions. All errors are our own. The Bagrut 2001 program was funded by the Israel Ministry of Education and the Branco Weiss Foundation. The statements in this paper reflect the views of the authors and have not been endorsed by the program's sponsors.

Email addresses of authors: msvictor@mscc.huji.ac.il, ani@mscc.huji.ac.il.

Targeted Remedial Education for Under-Performing Teenagers:

Costs and Benefits

Abstract

There is renewed interest in ways to enhance secondary education, especially among disadvantaged students. This study evaluates the short-term effects of a remedial-education program that provided additional instruction to under-performing high-school students in Israel. The program targeted 10th–twelfth graders who needed additional help to pass the matriculation exams. Using a comparison group of schools that enrolled in the program later and implementing a differences-in-differences estimation strategy, we found that the program raised the school mean matriculation rate by 3.3 percentage points. This gain reflects mainly 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.

Key Words: Remedial education, natural experiment

JEL CODES: I20, J24

1. Introduction

Recent increases in wage differentials between skilled and unskilled workers have led to renewed interest in ways and means of enhancing secondary and post-secondary 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 still scanty, however.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 multi-year program, implemented in Israel since 1999, of remedial education for under-performing 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.

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 US communities (US Department of Education, 2003); Bell After-School Instructional Curriculum (BASICS) serves schools in Boston, New York City, and Washington, D.C. (www.bellnational.org); and Boys and Girls Clubs of America offer programs in several locations across the US (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 website: www.gse.harvard.edu/hfrp/projects/afterschool/bibliography/tutoring.html.

² Recent exceptions are Banerjee, Cole, Duflo and Linden, 2003, about remedial education for third- and fourth-graders in India; Jacob and Lefgren, 2002, on the effects of summer school and grade retention in Chicago public primary schools; and Machin, McNally and Meghir, 2003, on additional resources targeting secondary schools in disadvantaged urban areas of England.

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

By looking at how remedial education affects achievement in high school, this paper 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 Krueger and Whitmore (2001), Currie (2001), Currie and Thomas (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 non-cognitive abilities later in life. The evidence on high-school interventions is mostly based on randomized trials in the US, 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 paper 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 2, we explain the study design and identification strategy in Section 3. Section 4 presents evidence of the overall effect of the program, and Sections 5 and 6 break it down into two components: effects on treated students and spillover effects on untreated students. Section 7 presents a cost-benefit analysis of the program and a comparison with alternative interventions, and Section 8 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 percent. 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 percent. No evidence was found of program externalities on non-participating peers.

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 and 2002); and the effects of incentives for teachers on student achievements (Lavy, 2002 and 2003).

2. The High-School System in Israel

Israeli post-primary 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)⁵ or in an alternative track leading only to a high-school diploma. 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 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 twenty credits is required to qualify for a matriculation certificate. About 52 percent of high-school graduates and 46 percent 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.⁷ The Ministry instituted several programs toward this goal: achievement awards for high-school students (Angrist and Lavy, 2002), school incentives (Lavy, 2002), teacher incentives (Lavy, 2003), 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 under-performing students.

The Bagrut 2001 Program

The Bagrut 2001 program targets low-achieving high schools. In its first phase, which began in October 1999, it included ten schools. During that academic year, nine additional schools were

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 US 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 website, <www.education.gov.il>.

⁷ Ministry of Education, *Facts and Figures* (2001). An online version can be found at: <www.education.gov.il/minhal_calcala/netunim.htm#A>.

<http://www.education.gov.il/minhal_calcala/netunim.htm#A>.

added to the program. In the following year there were forty-two participating schools. The program continued to expand; in 2002, it included 130 schools. The gradual implementation was due mainly to budget constraints. Schools were not phased into the program in any particular order, although the first-year participants seem to have been slightly weaker than the rest. Table 1 presents the program timeline and the dates of the matriculation examinations that constitute the basis for the pre- and post-program achievements used in the evaluation.

The intervention included individualized instruction in small study 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 after school hours and were taught by the classroom teachers.

3. Design of the Study

The first year (1999/2000) was a pilot in which only a few schools participated. Therefore, the evaluation focuses on the first year of full implementation, 2000/2001. Thus, 1998/1999 is considered a pre-treatment year.

Table 2 presents the population (4,100 students, one-fifth of all students in the treated schools) affected by the intervention. Our evaluation focuses on the effects of the program on the matriculation status: a comparable outcome for all twelfth graders. Of all twelfth-graders, 28 percent participated in the program, but this rate varied across schools, from 11 to 79 percent. 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. Though this rule was not strictly enforced, it led to a negative correlation between the proportion enrolled in the program and school size (Figures 1 and 2).

Figures 3 and 4 show the relationship between the pre-program school matriculation rate and the program participation rate. Figure 3 shows the results for grades 10–12 and Figure 4 for twelfth grade only. Both figures suggest that there was no correlation between the program participation rate and the pre-program school matriculation rate.

Identification Strategy

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 (2001–2002). The mean characteristics of these schools strongly resemble those of the schools that enrolled first. Since this similarity is found both for the pre-treatment 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.

Tables 3 and 4 present descriptive statistics on three groups of schools: first-year program schools, schools enrolled in the second year, and all other high schools in Israel. Table 3 presents the evidence based on the data on 1999 seniors (pre-treatment graduating cohort) and Table 4 on 2001 seniors (treatment cohort). Table 3 shows only a few slight 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 treated schools, suggesting that these schools are slightly weaker. The pre-program matriculation rate was also lower in treated schools (44 versus 50 percent), although this difference is only marginally significant ($t=-1.74$). The similarity between treatment and comparison schools is also found among the treatment cohort (Table 4). For this cohort it is worth emphasizing the lack of significant differences in terms of lagged achievements (number of credit units 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 pre-treatment period were also considerably higher in the “other” schools than in treated schools. For example, the matriculation rate in 1999 (the pre-treatment 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.

4. 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 receives 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 in 2001. These cross-section data are used to compare the change in the school mean matriculation rate in treated and comparison schools. This approach amounts to a differences-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.⁸ Although the validity of this assumption cannot be verified, Table 5 shows that there were no significant differences in the matriculation-rate trend between treated and comparison schools in the pre-treatment years. In particular, there is no evidence of an 'Ashenfelter dip' in outcomes in the years before treatment.

A. *Empirical Models*

Treatment effects are estimated using both microdata at the student level and aggregated data at the school level. In a model without additional covariates, the analyses using microdata 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 differences-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:

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

where s indexes schools and $t = \{1999, 2001\}$ 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

⁸ The model also assumes constant and additive treatment effects.

two measures of pre-treatment 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; 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 pre-treatment and treatment years and estimated a differences-in-differences model with school and year fixed effects as follows:

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

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 microdata at the student level. For example, the differences-in-differences model may be written as

$$(3) \quad Y_{ist} = \beta' X_{ist} + \omega' S_{st} + \phi_s + \delta_t + \alpha D_{st} + v_{st} + \epsilon_{ist}$$

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 $v_{st} + \epsilon_{ist}$ 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 between schools in terms of the proportion of treated students to obtain an alternative measure of the effects of the program.⁹ We use two alternative versions 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- to 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. Although treatment intensity in twelfth grade might provide a more accurate estimate of the program effects among twelfth graders, it might be correlated with the twelfth graders' perceived potential. For example, a school might choose mostly twelfth graders for the program if this grade is a particularly weak cohort. The use of an alternative measure (treatment intensity at the school level) provides an opportunity to test the robustness of the results.

The main concern of this strategy is the possible endogeneity of treatment intensity. The correlation between school treatment intensity and potential achievements may be either positive or

⁹ This strategy is similar to that applied by Card and Krueger (1994).

negative. For example, successful school principals (those whose schools have high matriculation rates) may get more program resources. On the other hand, program administrators may target more resources to weaker schools. To assess the likelihood of these possibilities, we estimate alternative models that evaluate whether pre-program achievements pertaining to the school or students predict school treatment intensity. The models control for school size in order to avoid any spurious correlation that could result from the existing correlation between school size and school achievement. The models are based on alternative measures of pre-program achievements: (1) The school pre-program achievements are the school matriculation rates in the years before treatment. (2) The student pre-program achievements are the achievement indices of the twelfth-grade cohort before treatment, such as the average score on matriculation exams taken before twelfth grade and the number of credits earned through those exams.

Table 6 shows that treatment intensity is not correlated with lagged school achievements or lagged student achievements but is highly negatively correlated with school size.¹⁰ This evidence suggests that, conditional on school size, treatment intensity may be considered independent of potential achievements. The estimated differences-in-differences models include school fixed effects, which eliminate any remaining permanent difference between schools that might be associated with variations in program intensity. The empirical models are stated as follows:

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

$$(5) \quad \bar{Y}_{st} = \beta' \bar{X}_{st} + \omega' S_{st} + \varphi_s + \delta_t + \alpha D_{st} INT_{st} + \bar{\epsilon}_{st}$$

B. Empirical Results

Table 7 presents the controlled differences in the matriculation rate between treated and comparison schools for both pre-treatment (1999) and treatment (2001) years, as well as the differences-in-differences estimates. Columns 1, 2, and 3 are based on specifications that use treatment status as a binary variable. Columns 4, 5, and 6 are based on treatment intensity in twelfth grade, and columns 7, 8, and 9 are based on average treatment intensity in grades 10–12.

In the pre-program 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 post-program period (2001) show that treated schools did better than comparison schools, suggesting a positive impact of the program; however,

¹⁰ For example, in a model with treatment intensity in twelfth grade as the dependent variable, the coefficient of pre-program matriculation rate is not significant (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.

the estimates are significantly different from zero only for the specifications based on the treatment-intensity indicators. Adjusting for pre-program differences and adding school fixed effects by estimating differences-in-differences increases the precision of the estimates. The differences-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, if we look at the results based on weighted aggregate school data (row 2), the estimate based on treatment as a binary variable is 0.033 (s.e.=0.014). The estimate based on treatment intensity in twelfth grade is 0.117 (s.e.=0.050), which, when evaluated at the average intensity (0.279), provides an estimate of the average program effect equal to 0.033. The estimate based on average treatment intensity in grades 10–12 is 0.186 (s.e.=0.078), implying at average intensity (0.203) an effect equal to 0.038. Estimating these models based on the microdata 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 A1 in the Appendix shows the differences-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.¹¹ 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.

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

¹¹ 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 the 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%); 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%.

treatment intensity. However, since school size may have a direct effect on school achievements, it cannot be used simply as an instrument of treatment intensity. However, the direct effect of school size on school achievement may be identified from the relationship in the pre-program baseline cohort, including schools in the comparison group. Therefore, school size in the treatment year may be used as an instrument of treatment intensity. More precisely, the interaction of school size with the year 2001 and treatment status serves as an instrument of treatment intensity. To allow for a more flexible relationship between treatment intensity and school size, we define 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.

$$(6) \quad INT_{st} = \beta' \bar{X}_{st} + \omega' S_{st} + \varphi_s + \delta_t + \gamma_1 \text{small}_{st} + \gamma_2 \text{med}_{st} + \gamma_3 \text{large}_{st} + \alpha_1 \text{inst1}_{st} + \alpha_2 \text{inst2}_{st} + \alpha_3 \text{inst3}_{st} + \bar{\epsilon}_{st}$$

$$(7) \quad \bar{Y}_{st} = \beta' \bar{X}_{st} + \omega' S_{st} + \varphi_s + \delta_t + \gamma_1 \text{small}_{st} + \gamma_2 \text{med}_{st} + \gamma_3 \text{large}_{st} + \alpha_1 \text{inst1}_{st} + \alpha_2 \text{inst2}_{st} + \alpha_3 \text{inst3}_{st} + \bar{\epsilon}_{st}$$

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 positive values for treated schools in 2001; small_{st} , med_{st} , and large_{st} are the three school-size variables; and the three instrumental variables are defined as follows:

$$\text{inst1}_s = \text{small}_{st} * \text{treated}_s * I(\text{year}=2001)$$

$$\text{inst2}_{st} = \text{med}_{st} * \text{treated}_s * I(\text{year}=2001)$$

$$\text{inst3}_{st} = \text{large}_{st} * \text{treated}_s * I(\text{year}=2001)$$

$I(\text{year}=2001)$ 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 8. 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 (s.e.=0.054) and the corresponding estimate based on treatment intensity in grades 10–12 is 0.202 (s.e.=0.067).

D. Allowing for Heterogeneous Treatment Effects

To allow for heterogeneous treatment effects in accordance with students' ability, we divided the distribution of the average score on matriculation exams taken before twelfth grade into quartiles and estimated the effect of the program for each quartile. To account for different time trends for each quartile, we converted the single time trend into four quartile trends.

Table 9 presents the results of the estimation of treatment effects by ability quartile based on aggregated weighted school data. We find no differences in the matriculation rates in the pre-treatment year (1999) between treated and comparison schools in any of the quartiles. This suggests that the achievements of the two groups of students not only were similar on average but were also similarly distributed. The differences-in-differences estimates are positive for all quartiles, although they are measured very imprecisely. The greatest effect seems to have occurred in the second quartile. However, we cannot reject the hypothesis of equality of coefficients in a regular joint F-test or in a joint test where each of the coefficients is weighted by the inverse average proportion of treated students in the quartile.

5. Effects on Treated Students and Spillover Effects on Non-Participants

Our next task is to identify the program effect on the treated students. The average school effect may be broken down into two parts: effect on participants and effect on participants' peers through externalities and spillover.¹² Since there is no natural comparison group for the treated students, we have to resort to non-experimental methods of identification, namely, conditioning on observed covariates. The central identifying assumption in this case is that treatment status is independent of potential achievements after conditioning on observable covariates and lagged achievements. Although this assumption is hardly satisfied in most non-experimental studies, the availability of a wealth of data on student characteristics and student achievement on matriculation exams taken

¹² 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 non-participants to participants.

before the program was implemented make this a credible approach in our case.¹³ As Smith and Todd (2004) 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 control group was selected from among the students in the comparison schools. We avoided choosing students from treated schools because they might have been exposed to the externalities of the program and might be different from program participants in observable and unobservable characteristics, as suggested by the fact that they were not selected to participate in 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 credits earned in grades 10 and 11, average weighted score on the matriculation exams taken in grades 10 and 11, 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 pre-program 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.¹⁴ An alternative approach is to implement a non-parametric 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: 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 non-parametric matching, in order to contrast the achievements for treated students with those for students from comparison

¹³ 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 pre-program 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.

¹⁴ Because some of the covariates are non-discrete, OLS estimates will be based on the whole sample even if for some values of the covariates there are only treated or comparison students.

schools and to estimate the effects of the program.¹⁵ The propensity score was estimated on the basis of all the aforementioned student covariates, pre-program achievements, and school characteristics.¹⁶

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 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 non-experimental comparison group to a sample that has the same distribution of covariates as the treated students. Table A2 in the appendix compares the characteristics of treated and comparison students matched by nearest neighbor matching.¹⁷ 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 thirty-four of the thirty-five 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 non-treated students, we applied the same methodology described for treated students. Non-treated students are compared to students from comparison schools, using linear regressions and alternative matching methods.

Empirical Results

Table 10 shows 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

¹⁵ The results reported in this paper are based on nearest neighbor matching and kernel matching. The other methods provide very similar results.

¹⁶ We tried various specifications in which we omitted one of the pre-program 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 pre-program data omitted from the propensity-score equation reinforces the credibility of the assumption that treated and comparison students would have done similarly afterwards, 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 theoretical context, see Davison and Hinkley, 1997; for an application of this procedure, see Subramanian and Deaton, 1996.

¹⁷ 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.

for non-participants 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 pre-program 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 post-program achievements. In Figures 5 and 6 we plotted the pre- and post-treatment achievements of treated students as a function of their propensity scores. The achievements are averaged according to the percentile of the propensity score. Figure 5 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 6, 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 10 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. Assuming that the program did not have spillover effects, as suggested by the results, 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.¹⁸ 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

¹⁸ 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.

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.

6. Estimates Based on Differences-in-Differences Matching Estimators

The availability of pre-program achievements allows us to apply an additional method to test for program effects. The use of differences-in-differences matching estimators lets us control for any time-invariant unobservable differences between treated and comparison students.¹⁹ An example may be fixed differences in school characteristics between treated and comparison schools. The assumption required for identification is that the difference between post-program and pre-program achievements for comparison students is the same as the difference among treated students would have been had they not been treated. This assumption may be expressed as follows:

$$E(Y_{0t'} - Y_{0t} | P(x), S=1) = E(Y_{0t'} - Y_{0t} | P(x), S=0)$$

where Y_0 denotes student achievements had the student not been treated; $P(X)$ is the propensity score; $S = 1$ if the student was treated; t denotes the pre-program period (when students were in grades 10–11); and t' denotes the post-program period (when students had completed twelfth grade). The differences-in-differences matching estimator is given by

$$\Delta_{\text{DID}} = n_T^{-1} \sum_{i \in \{CS \cap T\}} \{ [Y_{1t'}(X_i) - \hat{E}(Y_{0t'} | P(X_i), S_i=0)] - [Y_{1t}(X_i) - \hat{E}(Y_{0t} | P(X_i), S_i=0)] \}$$

where

$$\hat{E}(Y_{0t} | P(X_i), S_i=0) = Y_{0j} \text{ with } j = \underset{j \in \{T=0\}}{\text{argmin}} (|P(X_i) - P(X_j)|); n_T \text{ denotes the number of treated students;}$$

CS denotes the common support region; and T denotes the subset of treated students.

Two measures of achievements are available for the pre-program and post-program 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 credit units is required to qualify for a matriculation certificate, we should expect to find a positive effect on mean credit units earned by treated students. The second measure is the average score on matriculation exams. For this one, however, the effect may go either way.²⁰

Panel A of Table 11 presents the comparison between the treated group and the matched comparison group. The results suggest that there are no differences between the two groups in

¹⁹ For theoretical context and an application of this estimator, see Heckman et al. (1997) and Smith and Todd (2003).

²⁰ 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.

mean pre-program achievements, such as the number of credit units earned in grades 10–11. However, we find a positive difference of two credit units (s.e.=0.623) for treated students in the post-program period. The differences-in-differences estimate is almost the same: 1.951 (s.e.=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 credit units without lowering their average score. Panel B presents a similar analysis based on non-treated students. The results suggest that the program did not have any effect (positive or negative) on the achievements of this group.

7. 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 that could be used as arguments against the estimates of the effect of treatment on treated students (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 return on an increase in the matriculation rate in terms of the present value of the increment in lifetime earnings.²¹ This exercise is greatly simplified if we assume that the main benefit of having a matriculation certificate over merely having completed twelve years of schooling is the effect on the prospects of acquiring a higher education. In 2000, the annual gross earnings of labor-force participants with 11–12 years of schooling was \$15,500; for those with 13–15 years of schooling it was \$18,900. About 71 percent of members of the class of 1991 who earned matriculation certificates enrolled in post-secondary scholastic programs by 2000.²² We assume conservatively that the corresponding rate among program participants will be only half of that, 35 percent. Given that the program boosted the average school matriculation rate by 3.3 points, the annual increment in earnings due to the program is estimated as follows: $(\$18,900 - \$15,500) \times 0.35 \times 0.033 = \39 . Assuming that this increment remains constant across lifetime earnings and that the positive

²¹ 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.

²² Central Bureau of Statistics, 2003.

payoffs begin four years after the costs are incurred and remain in effect for forty-three years (until the individuals reach age sixty-five), the internal rate of return is 9 percent. 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 12 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, 2002), and a teacher-bonus program that paid teachers bonuses on the basis of their students' performance on matriculation exams (Lavy, 2002). 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 comparison. A more complete evaluation should also consider the long-term effects, including the effect on completion of higher education, employment, and even earnings.

8. Conclusions

In this paper we have studied the effects of a relatively traditional and somewhat forgotten intervention—targeted additional instruction time—on the quality of schooling. The results suggest that this remedial education program led to a 3–4 percentage point increase in the school mean matriculation rate, which represents a 11–12 percentage point increase in the matriculation rate among participating students - an improvement of 22 percent. 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 percent 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 under-performing teenagers can be helped to earn matriculation certificates, an important educational achievement with a high payoff for post-secondary schooling and success in the labor market. Remedial education targeted at adolescents while still in school appears to be a

relative successful intervention, especially given the low effectiveness of most high school dropout prevention programs in the US and also the low rate of return to a GED certification (Heckman and Rubinstein 2001). The second policy implication 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.²³ 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 not so 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 method (Cooper et al., 2000; Jacob and Lefgren, 2004). Summer schooling is generally targeted and has recently been gaining popularity in the US and Europe as a form of remedial education.²⁴ Both interventions—summer schooling and remedial education during the school year—may be viewed as ways of providing additional instruction time to disadvantaged or under-performing students. The former attempts primarily to remedy failure; the latter is designed to prevent it.

²³ Although most reductions in class size target preschool and early grades, some US states have recently initiated reforms aimed at reducing class size in high schools. Unfortunately, we were not able to find any evidence about the costs and benefits of class-size reduction in high school.

²⁴ 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>.

References

- Angrist, Joshua D. (1998). "Estimating the Labor Market Impact of Voluntary Military Service Using Social Security Data on Military Applicants," *Econometrica*, Vol. 66, No. 2, pp. 249-288.
- Angrist, Joshua D. and Victor Lavy (1999). "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement," *Quarterly Journal of Economics*, Vol. 114 No. 2, pp. 533-75.
- Angrist, Joshua D. and Victor Lavy (2001). "Does Teacher Training Affect Pupil Learning? Evidence from Matched Comparisons in Jerusalem Public Schools," *Journal of Labor Economics*, Vol. 19 No. 2, pp. 343-369.
- Angrist, Joshua D. and Victor Lavy (2002). "New Evidence on Classroom Computers and Pupil Learning," *The Economic Journal*, Vol. 112 No. 482, pp. 735-765.
- Angrist, Joshua D. and Victor Lavy (2002). "The Effects of High School Matriculation Awards: Evidence from Randomized Trials," NBER Working Paper No. 9389.
- Banerjee, Abhijit, Shawn Cole, Esther Duflo and Leigh Linden (2003). "Remedying Education: Evidence from Two Randomized Experiments in India," mimeo, MIT.
- Becker, Sascha O. and Andrea Ichino (2002). "Estimation of Average Treatment Effects Based on Propensity Scores," *The Stata Journal* Vol. 2 No. 4 pp. 338-377.
- Betts, Julian (1995). "Does School Quality Matter? Evidence from the National Longitudinal Survey of Youth," *Review of Economics and Statistics*, Vol. 77 No. 2, pp. 231-250.
- Card, David and Krueger Alan (1992). "Does School Quality Matter? Return to Education and the Characteristics of Public Schools in the United States," *Journal of Political Economy*, Vol. 100 No. 1, pp. 1-40.
- Carneiro, Pedro and James Heckman (2003). "Human Capital Policy," NBER Working Paper No. 9495.
- Card, David and Alan Krueger (1994). "Minimum Wages and Employment: A Case Study of the Fast Food Industry in New Jersey and Pennsylvania," *American Economic Review*, Vol. 84 No. 4, pp. 772-793.
- Central Bureau of Statistics, *Statistical Abstract of Israel 53*. Jerusalem: Central Bureau of Statistics, 2003.
- Cooper, Harris, Kelly Charlton, Jeffrey C. Valentine, and Laura Muhlenbruck (2000). Making the Most of Summer School: A Meta-Analytic and Narrative Review. *Monographs of the Society for Research in Child Development*, Vol. 65 No. 1, pp. 1-118.
- Currie, Janet (2001). "Early Childhood Education Programs," *Journal of Economic Perspectives*, Vol. 15 No. 2, pp. 213-238.
- Currie, Janet and Duncan Thomas (2001). "Early Test Scores, School Quality and SES: Long Run Effects on Wage and Employment Outcomes," *Research in Labor Economics*, Vol. 20, pp. 103-132.
- Davison Antony C. and David Hinkley. *Bootstrap Methods and their Application*. Cambridge University Press, 1997.
- Dynarski, Mark and Philip Gleason (1998). How Can We Help? What Have We Learned from Evaluations of Federal Dropout-Prevention Program, Princeton, NJ: Mathematica Policy Research Report 8014-140.

- Garces, Eliana, Duncan Thomas and Janet Currie (2002). "Longer Term Effects of Head Start," *American Economic Review*, Vol. 92 No. 4, pp. 999-1012.
- Heckman, James J., Ichimura Hidehiko and Petra Todd (1997). "Matching as an Econometric Evaluation Estimator," *Review of Economic Studies* Vol. 64 No. 4, pp. 605-654.
- Heckman, James and Yona Rubinstein (2001). "The Importance of Non-Cognitive Skills: Lessons from the GED Testing Program," *American Economic Review*, Vol. 91 No. 2, pp. 145-149.
- Hoxby, Caroline M. (2000). "The Effects of Class Size on Student Achievement: New Evidence from Population Variation," *Quarterly Journal of Economics*, Vol. 115 No. 3, pp. 1239-1285.
- Jacob, Brian A. and Lars Lefgren (2004). "Remedial Education and Student Achievement: A Regression Discontinuity Analysis," *Review of Economics and Statistics*, Vol. 86 No. 1, 226-244.
- Jepsen C. and S. Rivkin (2002) "What is the Tradeoff Between Smaller Classes and Teacher Quality?," NBER Working Paper No. 9205.
- Koretz, Daniel (1996). "Using Student Assessments for Educational Accountability," *Improving American Schools*, Eric A. Hanushek and Dale W. Jorgenson, eds. Washington D.C.: National Academy Press, pp. 171-195.
- Krueger, Alan B. (1999). "Experimental Estimates of Education Production Functions," *Quarterly Journal of Economics*, Vol. 114 No. 2, pp. 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*, Vo. 111 No. 468, pp. 1-28.
- Machin, Stephen, Sandra McNally and Costas Meghir (2003). "Excellence in Cities: Evaluation of an Education Policy in Disadvantaged Areas," Draft (March).
- Lavy, Victor (2002). "Evaluating the Effect of Teachers' Group Performance Incentives on Pupil Achievement," *Journal of Political Economy*, Vol. 110 No. 6, pp. 1286-1317.
- Lavy, Victor (2003). "Paying for Performance: Experimental Evidence on the Effect of Teachers' Individual Incentives on Effort and Productivity," Draft, (July).
- Rosenbaum, Paul R. and Donald B. Rubin (1983). "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, Vol. 70 No. 1, pp. 41-55.
- Schinke, S. P., Kristin Cole and Stephen R. Poulin, (2000). "Enhancing the Educational Achievement of at-Risk Youth," *Prevention Science*, Vol. 1 No. 1, pp. 51-60.
- Smith, Jeffrey and Petra Todd (2004). "Does Matching Overcome Lalonde's Critique of Nonexperimental Estimators?," *Journal of Econometrics* (forthcoming).
- Subramanian, Shankar and Angus Deaton (1996). "The Demand for Food and Calories," *Journal of Political Economy*, Vol. 104 No. 1, pp. 133-62.
- U.S. Department of Education, Office of the Under Secretary (2003), *When Schools Stay Open Late: The National Evaluation of the 21st Century Community Learning Centers Program, First Year Findings*, Washington D.C.

Figure 1: Program Participation Rate by School Size

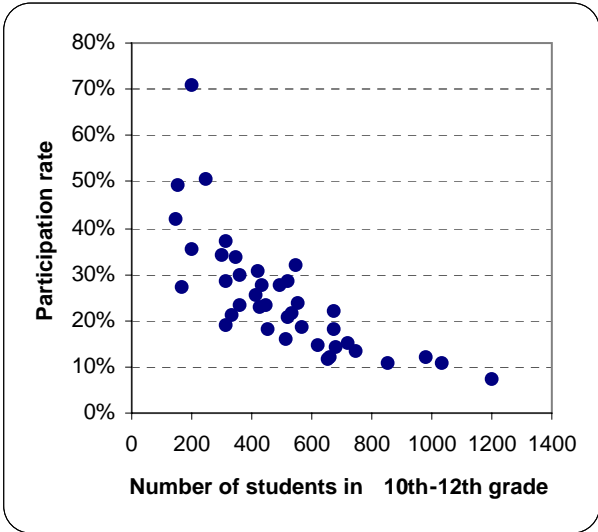


Figure 2: 12th Grade Participation Rate by Number of 12th Grade Students at School

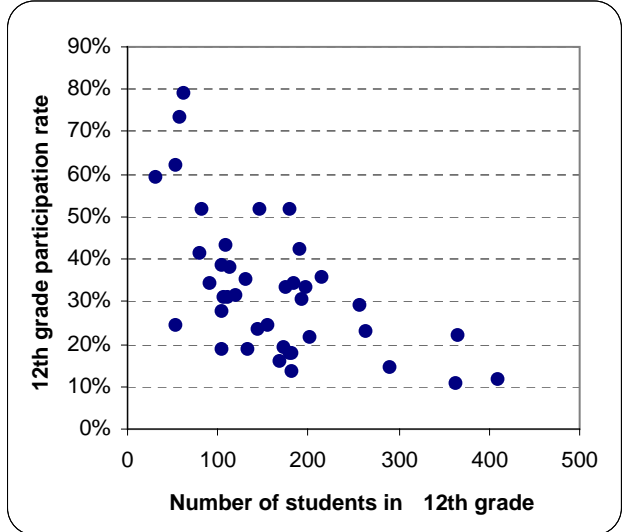


Figure 3: Program Participation Rate by Pre-Treatment Matriculation Rate

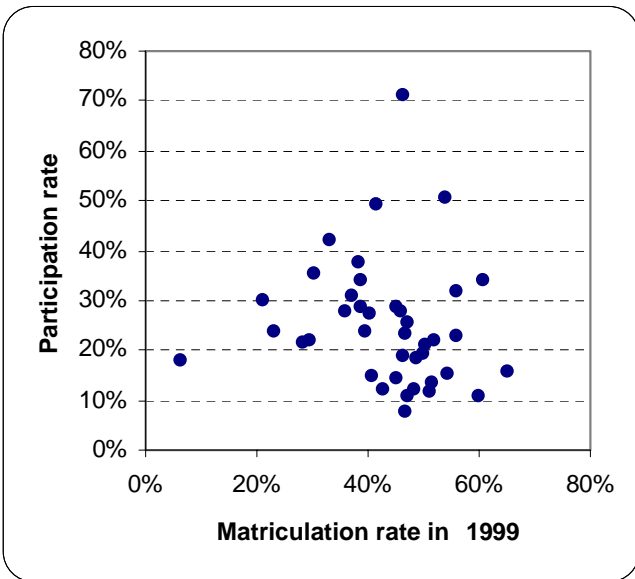


Figure 4: 12th Grade Participation Rate by Pre-Treatment Matriculation Rate

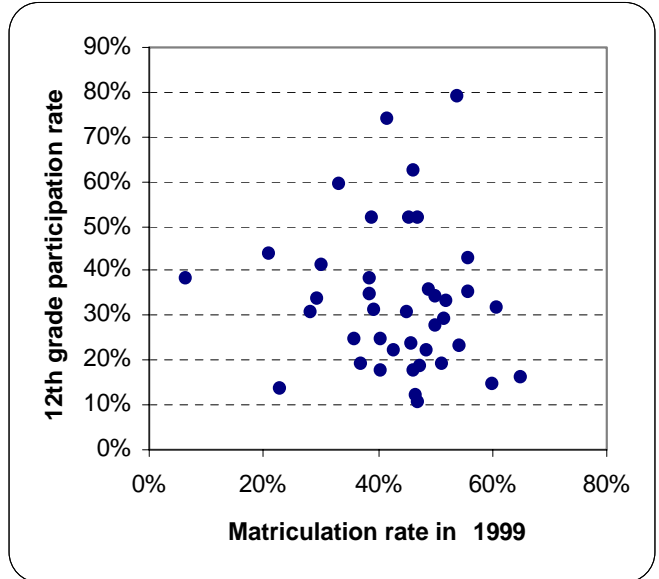


Figure 5: Lagged Average Score by Propensity Score Percentiles

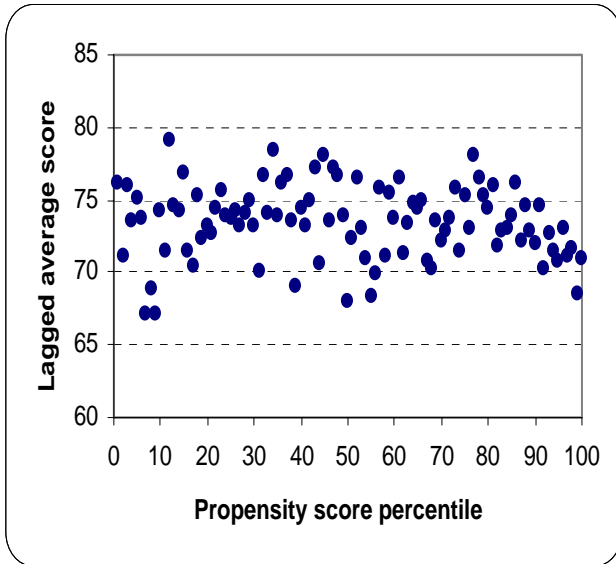


Figure 6: Matriculation Rate by Propensity Score Percentiles

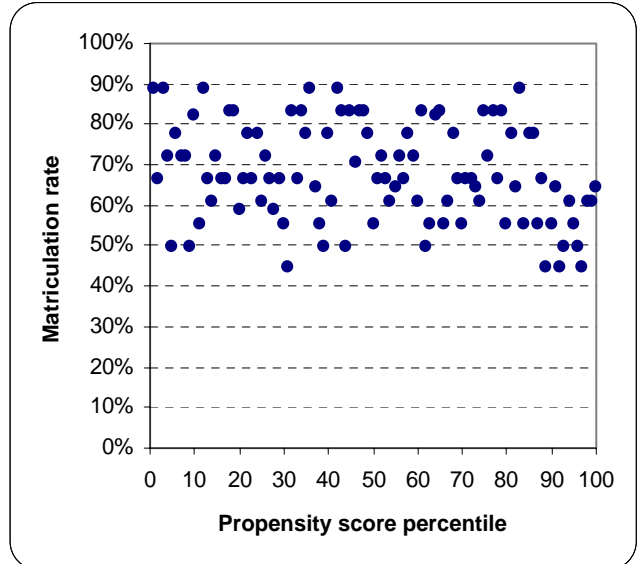


Table 1. Program time line

	June 1999	October 1999	January 2000	June 2000	October 2000	April 2001	June 2001	October 2001	January 2002
Date of program initiation for different sets of schools		10	9		23	24		34	33
		Treatment schools group (40 schools)						Comparison group	
Date of matriculation exams	+			+			+		
	Pre-treatment outcomes						Post-treatment outcomes		

Notes: Two treatment schools have been excluded from the analysis. One school is missing in the scores dataset from 1999, leaving us with no pre-treatment outcomes. The second school did not provide any data about the program implementation. As a result, we cannot identify the program participants or infer the proportion of students participating in the program in that school. The inclusion of this school in the estimation of the program effects at the school level did not change the magnitude or precision of the results.

Table 2. Participation in Bagrut 2001 program

	Number of students in treated schools	Number of program participants	Program participation rate		
			Mean	Min	Max
10 th -12 th grade	20,094	4,117	20%	7%	71%
12 th grade	6,450	1,789	28%	11%	79%

Table 3. Mean Comparison of Covariates and Outcomes for Pre-Treatment Year (1999)

	Treatment schools	Comparison schools	Difference (1-2)	Other schools	Difference (1-4)
	(1)	(2)	(3)	(4)	(5)
Student Characteristics					
Father's years of schooling	10.35	10.48	-0.13 (0.45)	12.14	-1.79 (0.34)
Mother's years of schooling	10.73	10.95	-0.21 (0.41)	12.20	-1.46 (0.28)
Number of siblings	2.58	2.25	0.33 (0.19)	2.28	0.30 (0.13)
Immigrant status (=1 if immigrated within the last 4 years)	0.03	0.06	-0.03 (0.01)	0.04	-0.01 (0.01)
Male	0.44	0.46	-0.02 (0.020)	0.48	-0.04 (0.02)
Parents Ethnic Origin					
Asia-Africa (excluding Ethiopia)	0.37	0.28	0.09 (0.03)	0.20	0.16 (0.02)
Former Soviet Union	0.19	0.22	-0.03 (0.03)	0.14	0.05 (0.02)
Ethiopia	0.02	0.05	-0.03 (0.02)	0.02	-0.00 (0.01)
Europe-America (excluding former Soviet Union)	0.07	0.09	-0.02 (0.01)	0.12	-0.06 (0.01)
Israel	0.36	0.37	-0.01 (0.03)	0.52	-0.16 (0.02)
Student Achievement					
Average lagged score in matriculation exams (prior to 12th grade)	65.38	67.18	-1.80 (1.54)	73.44	-8.06 (0.98)
Number of credit units awarded prior to 12th grade	5.09	5.64	-0.55 (0.37)	6.50	-1.40 (0.25)
School Characteristics					
Matriculation rate	0.44	0.50	-0.06 (0.04)	0.64	-0.20 (0.02)
Enrollment in religious schools					
Enrollment in 12 th grade	212.4	239.4	-27.1 (33.0)	222.9	-10.5 (23.6)
Enrollment in 10 th -12 th grade	652.8	720.6	-67.8 (101.1)	677.2	-24.4 (72.2)
Number of schools	34	40		367	
Number of students	6,725	5,508		48,710	

Notes: Mean values for treatment schools are reported in column 1. Mean values for comparison schools are reported in column 2. Mean values for all remaining Jewish matriculation-track high schools in Israel ("other schools") are reported in column 4. Difference of means between treatment and comparison schools is reported in column 3. Difference of means between treatment and all other schools is reported in column 5. Standard errors for the differences corrected for clustering at school level are reported in parentheses.

Table 4. Mean Comparison of Covariates and Outcomes for Treatment Year (2001)

	Treatment schools	Comparison schools	Difference (1-2)	Other schools	Difference (1-4)
	(1)	(2)	(3)	(4)	(5)
Student Characteristics					
Father's years of schooling	10.62	10.69	-0.07 (0.41)	11.53	-0.91 (0.34)
Mother's years of schooling	10.75	10.99	-0.25 (0.48)	11.41	-0.67 (0.37)
Number of siblings	2.52	2.07	0.45 (0.18)	2.26	0.26 (0.15)
Immigrant status (=1 if immigrated within the last 4 years)	0.02	0.04	-0.02 (0.01)	0.03	-0.01 (0.01)
Male	0.45	0.45	-0.01 (0.02)	0.48	-0.03 (0.02)
Parents Ethnic Origin					
Asia-Africa (excluding Ethiopia)	0.33	0.23	0.10 (0.03)	0.18	0.15 (0.02)
Former Soviet Union	0.21	0.24	-0.03 (0.04)	0.14	0.08 (0.03)
Ethiopia	0.02	0.05	-0.03 (0.02)	0.02	-0.00 (0.01)
Europe-America (excluding former Soviet Union)	0.06	0.08	-0.03 (0.01)	0.12	-0.06 (0.01)
Israel	0.38	0.40	-0.02 (0.03)	0.54	-0.16 (0.98)
Student Achievement					
Average lagged score in matriculation exams (prior to 12th grade)	68.92	70.66	-1.74 (1.70)	76.09	-7.17 (0.89)
Number of credit units awarded prior to 12 th grade	5.86	6.16	-0.31 (0.36)	6.95	-1.10 (0.28)
School Characteristics					
Matriculation rate	0.53	0.58	-0.05 (0.04)	0.68	-0.15 (0.02)
Enrollment in religious schools					
Enrollment in 12 th grade	206.8	230.8	-24.0 (31.9)	217.8	-11.0 (23.2)
Enrollment in 10 th -12 th 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

Notes: Mean values for treatment schools are reported in column 1. Mean values for comparison schools are reported in column 2. Mean values for all remaining Jewish matriculation-track high schools in Israel ("other schools") are reported in column 4. Difference of means between treatment and comparison schools is reported in column 3. Difference of means between treatment and all other schools is reported in column 5. Standard errors for the differences corrected for clustering at school level are reported in parentheses.

Table 5. Difference in Matriculation Rate Trends between Treatment and Comparison Schools for Pre-Treatment Years

	Unweighted	Weighted
	(1)	(2)
Treatment*1997	-0.037 (0.058)	-0.016 (0.065)
Treatment*1998	0.010 (0.056)	0.016 (0.067)
Treatment*1999	-0.023 (0.050)	0.008 (0.059)
Number of schools	67	67

Notes: Results are based on aggregate data at the school level. Number of students in 12th grade is used as weights for the weighted regression reported in column 2. The dependent variable is the school matriculation rate. Regressions include year dummies as well as the interaction between year and treated status, which is reported above. Robust standard errors are in parentheses.

Table 6. The Relationship between Program Intensity and Pre-Program Outcomes

Dependent variable	Explanatory variables: school pre-program outcomes				Explanatory variables: treated cohort pre-program outcomes			
	1999 regression		2000 regression		Lagged score regression		Lagged credit units regression	
	Matriculation rate (1)	School size (2)	Matriculation rate (3)	School size (4)	Average score (5)	School size (6)	Number of credit units (7)	School size (8)
a. All treated schools (N=40)								
Treatment intensity in 12 th grade	0.043 (0.169)	-0.00031 (0.00005)	-	-	0.003 (0.004)	-0.00033 (0.00005)	0.003 (0.013)	-0.00033 (0.00005)
Treatment intensity in 10 th -12 th grade	-0.102 (0.116)	-0.00025 (0.00004)	-	-	0.000 (0.002)	-0.00030 (0.00004)	0.006 (0.006)	-0.00029 (0.00004)
b. Schools enrolled in the program in 2001 (N=23)								
Treatment intensity in 12 th grade	0.110 (0.237)	-0.00032 (0.00008)	0.175 (0.208)	-0.00036 (0.00008)	0.004 (0.005)	-0.00032 (0.00006)	0.004 (0.015)	-0.00032 (0.00006)
Treatment intensity in 10 th -12 th grade	-0.008 (0.134)	-0.00025 (0.00004)	-0.043 (0.104)	-0.00028 (0.00005)	0.001 (0.003)	-0.00027 (0.00004)	0.004 (0.008)	-0.00026 (0.00004)

Notes: Two alternative dependent variables are used in the regressions: treatment intensity in twelfth grade and treatment intensity in tenth- to twelfth grade. Every row has four different regressions that evaluate whether pre-program achievements pertaining to the school or students, predict treatment intensity. All regressions control for school size, which is also reported. For example, the parameters reported in columns 1 and 2 are the estimates obtained from regressing treatment intensity on the school matriculation rate in 1999 and school size; the parameters reported in columns 5 and 6 are the estimates obtained from regressing treatment intensity on the average score in the matriculation exams taken by the treated cohort before they were treated (in tenth and eleventh grade) and school size. Panel A reports the results based on all treated schools. Therefore, the pre-program school outcomes are taken from year 1999 only. Panel B reports the results based on schools that were added to the program in year 2001. Therefore, school pre-program outcomes are from year 1999 and year 2000. All models are based on aggregated weighted data at the school level. The number of students in 12th grade is used as weights. Robust standard errors are in parentheses.

Table 7. The Estimated Effect of Additional Targeted Individualized Instruction on the Matriculation Rate, Treated Status Defined at the School Level

	Treatment as a Binary Variable			Treatment Intensity in 12 th Grade			Treatment Intensity in 10 th -12 th grade		
	1999 (1)	2001 (2)	DID estimates using School Fixed Effects (3)	1999 (4)	2001 (5)	DID estimates using School Fixed Effects (6)	1999 (7)	2001 (8)	DID estimates using School Fixed Effects (9)
Unweighted grouped data	-0.018 (0.022)	0.032 (0.026)	0.040 (0.021)	-0.018 (0.050)	0.123 (0.064)	0.155 (0.073)	-0.042 (0.070)	0.149 (0.085)	0.237 (0.104)
Weighted grouped data	-0.028 (0.023)	0.018 (0.025)	0.033 (0.014)	-0.045 (0.054)	0.133 (0.066)	0.117 (0.050)	-0.070 (0.068)	0.161 (0.093)	0.186 (0.078)
Microdata	-0.005 (0.021)	0.020 (0.024)	0.026 (0.010)	0.018 (0.046)	0.123 (0.059)	0.083 (0.036)	-0.015 (0.060)	0.156 (0.085)	0.127 (0.054)
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

Notes: The dependent variable for the regressions based on grouped data is the school matriculation rate. The dependent variable for the regressions based on microdata is the student matriculation status. The number of students in 12th grade is used as weights for the weighted regressions. Three alternative measures of treatment status are used in the regressions: treatment as a binary variable (columns 1-3), and two versions of treatment intensity. Columns 4-6 use the proportion of 12th grade students receiving treatment (average 0.277); columns 7-9 use the proportion of 10th-12th grade students receiving treatment (average 0.203). The regressions include additional control variables at the student level for models based on microdata or averaged at the school-year level for models based on grouped data. The following control variables are used in the regressions: gender, immigration status, parental schooling, number of siblings, two measures of pre-treatment outcomes indicating the quartile of a student's average score in matriculation exams taken before 12th grade and the quartile of the number of credit units awarded before 12th grade, school's religious status, school size, its square, and the number of students in 12th grade. The differences-in-differences models (columns 3, 6 and 9) include in addition a year effect and school fixed effects. Robust standard errors are reported in parentheses. Standard errors for the estimates based on microdata are corrected for clustering at the school-year level.

Table 8. 2SLS Estimates for Treatment Intensity

	Unweighted			Weighted		
	Treatment Intensity in 12 th grade (1)	Treatment Intensity in 10 th -12 th grade (2)	Matriculation Rate (3)	Treatment Intensity in 12 th grade (4)	Treatment Intensity in 10 th -12 th grade (5)	Matriculation Rate (6)
Instrument1	0.00229 (0.00023)	0.00186 (0.00023)	0.00056 (0.00024)	0.00220 (0.00024)	0.00178 (0.00024)	0.00051 (0.00022)
Instrument2	0.00053 (0.00008)	0.00040 (0.00005)	0.00003 (0.00004)	0.00050 (0.00008)	0.00038 (0.00005)	0.00004 (0.00003)
Instrument3	0.00028 (0.00006)	0.00016 (0.00004)	0.00003 (0.00003)	0.00024 (0.00006)	0.00013 (0.00004)	0.00003 (0.00002)
F(3,54)	50.14	50.06	2.04	41.98	49.48	3.13
2SLS	0.174 (0.074)	0.228 (0.088)		0.154 (0.054)	0.202 (0.067)	

Notes: The models are based on aggregated data at the school-year level. The number of students in 12th grade is used as weights for the regressions in columns 4-6. Three instruments are used in the regressions, which denote the interaction of school size with year 2001 dummy and treatment status at three different points of the school-size distribution: instrument1 is for the first quartile of the school-size distribution (small schools), instrument2 is for the second and third quartiles (medium-size schools) and instrument3 is for the fourth quartile (large schools). The regressions include the following control variables averaged at the school-year level: gender, immigration status, parental schooling, number of siblings, two measures of pre-treatment outcomes indicating the quartile of a student's average score in matriculation exams taken before 12th grade and the quartile of the number of credit units awarded before 12th grade, school's religious status, three variables for school size at different points of the school size distribution, the number of students in 12th grade, school fixed effects and year effects. Robust standard errors are reported in parentheses. The reported F-statistics corresponds to a test of a joint significance of the three instruments.

Table 9. Treatment Effects by Ability Quartiles

	Proportion of treated students	Treatment as a Binary Variable			Treatment Intensity in 12 th Grade			Treatment Intensity in 10 th -12 th Grade		
		1999 (1)	2001 (2)	DID estimates using School Fixed Effects (3)	1999 (4)	2001 (5)	DID estimates using School Fixed Effects (6)	1999 (7)	2001 (8)	DID estimates using School Fixed Effects (9)
Quartile 1 (lowest)	0.17	-0.020 (0.015)	0.018 (0.020)	0.020 (0.022)	-0.019 (0.038)	0.107 (0.068)	0.068 (0.069)	-0.037 (0.051)	0.197 (0.093)	0.191 (0.094)
Quartile 2	0.37	-0.021 (0.031)	0.033 (0.039)	0.041 (0.028)	-0.004 (0.081)	0.207 (0.097)	0.159 (0.075)	-0.014 (0.091)	0.223 (0.124)	0.171 (0.096)
Quartile 3	0.35	-0.001 (0.035)	0.005 (0.034)	0.021 (0.022)	0.030 (0.087)	0.077 (0.086)	0.044 (0.063)	-0.015 (0.119)	0.136 (0.123)	0.110 (0.087)
Quartile 4	0.22	-0.009 (0.021)	-0.002 (0.021)	0.015 (0.025)	-0.018 (0.062)	0.080 (0.052)	0.076 (0.067)	-0.001 (0.090)	0.061 (0.084)	0.062 (0.094)
F(3,496) ¹				0.23			0.64			0.49
F(3,496) weighted ²				0.15			0.61			0.78
Number of Schools		74	74		74	74		74	74	

Notes: The dependent variable is the school matriculation rate. The models are based on weighted school averages. The number of students in 12th grade is used as weights. Three alternative measures of treatment status are used in the regressions: treatment as a binary variable (columns 1-3), and two versions of treatment intensity. Columns 4-6 use the proportion of 12th grade students receiving treatment (average 0.277); columns 7-9 use the proportion of 10th-12th grade students receiving treatment (average 0.203). The regressions include the following control variables averaged at the school-year level: gender, immigration status, parental schooling, number of siblings, two measures of pre-treatment outcomes indicating the quartile of a student's average score in matriculation exams taken before 12th grade and the quartile of the number of credit units awarded before 12th grade, school's religious status, school-size, school-size squared, the number of students in 12th grade, school fixed effects and year effects. Robust standard errors are reported in parentheses.

¹ F-Value refers to a test of the equality of coefficients for the four quartiles.

² F-Value weighted refers to a test of equality of coefficients for the four quartiles where each coefficient is weighted by the inverse proportion of treated students in the quartile.

Table 10. The Estimated Effect of Targeted Individualized Instruction on Matriculation Rate: Student Level Analysis

	Program Participants			Non-Participants		
	OLS (1)	Nearest Neighbor Matching (2)	Kernel Matching (3)	OLS (4)	Nearest Neighbor Matching (5)	Kernel Matching (6)
Treatment Effect	0.132 (0.026)	0.139 (0.047)	0.132 (0.035)	0.007 (0.016)	0.007 (0.044)	0.002 (0.037)
Treatment effects by ability quartiles						
Quartile 1 (lowest)	0.217 (0.050)	0.215 (0.065)	0.218 (0.060)	0.001 (0.015)	-0.017 (0.030)	-0.009 (0.022)
Quartile 2	0.164 (0.037)	0.163 (0.065)	0.156 (0.046)	0.013 (0.030)	0.009 (0.050)	0.006 (0.039)
Quartile 3	0.103 (0.030)	0.101 (0.065)	0.091 (0.046)	0.016 (0.027)	0.010 (0.059)	0.007 (0.044)
Quartile 4	0.070 (0.028)	0.038 (0.058)	0.036 (0.041)	0.001 (0.018)	0.003 (0.044)	0.000 (0.030)

Notes: The dependent variable is the student matriculation status. Estimates for columns 2, 3, 5 and 6 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 pre-treatment outcomes indicating the quartile of a student's average score in matriculation exams taken before 12th grade and the quartile of the number of credit units awarded before 12th grade, school matriculation rate in 1999, school's religious status, school size, its square and the number of students in 12th grade. Standard errors (in parentheses) are corrected for clustering at 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.

Table 11. Differences-in-Differences Matching Estimators

	Pre-program outcomes in 10 th -11 th grade		Post-program outcomes in 12 th grade		DID (5)
	Mean (1)	Diff. (2)	Mean (3)	Diff. (4)	
A. Treated vs. matched comparison students					
Number of credit units awarded	6.791	0.052 (0.383)	13.357	2.003 (0.623)	1.951 (0.773)
Weighted average score	71.933	1.532 (0.973)	68.628	2.926 (1.243)	1.255 (1.124)
Number of observations: 3,578					
B. Untreated vs. matched comparison students					
Number of credit units awarded	5.447	-0.001 (0.449)	12.157	-0.827 (0.578)	-0.830 (0.559)
Weighted average score	65.737	1.389 (2.399)	64.024	-0.016 (1.975)	-1.484 (1.394)
Number of observations: 9,318					

Notes: The comparison students for treated and untreated students were selected applying nearest neighbor matching. The following variables were used for the estimation of the propensity score: gender, immigration status, parental schooling, number of siblings, two measures of pre-treatment outcomes indicating the quartile of a student's average score in matriculation exams taken before 12th grade and the quartile of the number of credit units awarded before 12th grade, school matriculation rate in 1999, school's religious status, school size, its square and the number of students in 12th grade. Standard errors (in parentheses) are robust and corrected for clustering at the school level.

Table 12. Comparison of Alternative Programs, Costs and Benefits

Program	Cost per student	Absolute gain (in percentage points)	Pre-treatment Bagrut rate
Matriculation Awards (Angrist and Lavy, 2002)	\$385	7	19%
Teacher Bonuses (Lavy, 2002)	\$170	4	41%
Bagrut 2001	\$300	3.3	44%

Table A1. The Effects of Treatment Intensity on the Matriculation Rate

		Unweighted (1)	Weighted (2)	Microdata (3)
A. Treatment Intensity in 12th Grade				
Low	<20%	0.005 (0.030)	0.022 (0.020)	0.026 (0.012)
Medium	20%-40%	0.030 (0.020)	0.031 (0.016)	0.026 (0.012)
High	>40%	0.097 (0.051)	0.064 (0.039)	0.028 (0.026)
B. Treatment Intensity in 10th-12th Grade				
Low	<15%	0.009 (0.021)	0.014 (0.016)	0.025 (0.013)
Medium	15%-30%	0.032 (0.021)	0.040 (0.017)	0.022 (0.013)
High	>30%	0.087 (0.051)	0.061 (0.040)	0.040 (0.024)

Notes: The dependent variable for the regressions based on grouped data is the school matriculation rate. The dependent variable for the regressions based on microdata is the student matriculation status. The number of students in 12th grade is used as weights for the weighted regressions. The regressions include additional control variables at the student level for models based on microdata or averaged at the school-year level for models based on grouped data. The following control variables are used in the regressions: gender, immigration status, parental schooling, number of siblings, two measures of pre-treatment outcomes indicating the quartile of a student's average score in matriculation exams taken before 12th grade and the quartile of the number of credit units awarded before 12th grade, school's religious status, school size, its square, the number of students in 12th grade, a year effect and school fixed effects. Robust standard errors are reported in parentheses. Standard errors for the estimates based on microdata are corrected for clustering at the school-year level.

Table A2. Mean Comparison of Covariates and Outcomes for Treated and Matched Students

	Treated Students	Matched Students	Difference (1-2)
	(1)	(2)	(3)
Student Characteristics			
Father's years of schooling	10.37	10.57	-0.20 (0.43)
Mother's years of schooling	10.63	10.71	-0.08 (0.51)
Number of siblings	2.51	2.45	0.07 (0.24)
Immigrant status (=1 if immigrated within the last 4 years)	0.02	0.01	0.00 (0.01)
Male	0.42	0.43	-0.01 (0.03)
Parents Ethnic Origin			
Asia-Africa (excluding Ethiopia)	0.32	0.32	0.00 (0.04)
Former Soviet Union	0.20	0.20	0.01 (0.05)
Ethiopia	0.02	0.02	-0.00 (0.01)
Europe-America (excluding former Soviet Union)	0.06	0.06	-0.00 (0.01)
Israel	0.40	0.41	-0.00 (0.04)
Student Achievement			
Average lagged score in matriculation exams (prior to 12th grade)	73.43	71.91	1.53 (0.97)
Number of credit units awarded prior to 12th grade	6.84	6.79	0.05 (0.38)
Matriculation rate	0.68	0.55	0.13 (0.04)
School Characteristics			
Enrollment at religious schools	0.21	0.20	0.00 (0.10)
Enrollment in 12 th grade	178	174	4.02 (21.32)
Enrollment in 10 th -12 th grade	544	567	-23.1 (61.9)
Number of schools	40	34	
Number of students	1,789	1,183	

Notes: Mean values for treatment schools are reported in column 1. Difference of means between treated and matched students is reported in column 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.