

The Effect of High Stakes High School Achievement Awards: Evidence from a Group Randomized Trial*

Joshua D. Angrist
Massachusetts Institute of Technology and NBER

and

Victor Lavy
Hebrew University of Jerusalem

Revised: November 2005

*Special thanks go to Rema Hanna and Alex Levkov for outstanding research assistance in Cambridge and Jerusalem. Thanks also to Daron Acemoglu, Abhijit Banerjee, David Card, Sue Dynarksi, Ron Ehrenberg, Jinyong Hahn, Guido Imbens, Alan Krueger, Adriana Kugler, Kevin Lang, Thomas Lemieux, and seminar participants at Berkeley, Boston University, CEMFI, Harvard/MIT, Hebrew University, McMaster, Princeton, SOLE, and UCLA/RAND for helpful discussions and comments on an earlier version. The 2001 Achievement Awards program was funded by the Israel Ministry of Education and administered by the division for secondary schools. The 2000 Madarom project in Southern Israel was funded by the Sacta-Rashi Foundation. We also gratefully acknowledge funding under NIH grant 1R01HD043809-01A1. The statements in the paper reflect the views of the authors and have not been endorsed by program sponsors or funding agencies.

CONTEXT

- ☆ The most important education milestone in Israel is the Bagrut, or matriculation certificate, awarded on the basis of subject tests in grades 10-12 (mostly in 12).
- ☆ The Bagrut is required by selective universities and many higher-paying white-collar jobs
- ☆ About half of seniors get a Bagrut, but rates are much lower in some schools and groups, especially rural, predominantly Asian-African, immigrant, and Arab.
- ☆ In an effort to increase Bagrut rates, we tried demonstration projects that offered cash incentives for low-SES pupils to take and pass Bagrut exams.

PROSPECTS AND RELATED EFFORTS

- ☆ The Israeli government spends millions on remedial programs for low-achieving high school seniors, with no obvious benefit.
- ☆ Service-intensive anti-dropout interventions to date have been disappointing in the U.S. (see, e.g., the SDDAP interventions; Dynarski and Gleason, 1998).
- We hoped that direct payments would tip the scales towards school and study, provide immediate tangible rewards, and deflect negative peer pressure.
- Our program is similar to Reich's (1998) proposal to offer \$25,000 vouchers to high school grads from poor US families.

Also related: EMA, LEAP, PACES, Progres, QOP, Kenya.

THEORETICAL BACKGROUND

- ❑ Education is an excellent predictor of welfare; many government programs attempt to increase schooling for low-ed groups.
- ❑ Schooling returns appear high. So why do many students drop out of HS? Why don't more go to college?

Possible explanations:

- ♣ Low learning ability, weak motivation, disutility from schooling
- ♦ Low expectations and/or high uncertainty about returns
- ♥ Comparative advantage in non-graduate jobs
- ♠ Liquidity constraints
- ☺ High personal discount rates or hyperbolic discounting

Especially relevant in Israel:

- Bagrut status differs from US dropout/college status, but similar trade-offs and considerations apply.

- ☆ Effects of uncertainty and discounting are compounded by compulsory military service (3 years for boys; 2 for girls).

Working life seems very far off.

- ☆ Many low-achieving schools are geographically or socially isolated (e.g., “development towns” or urban slums, with few educated adult role models).

PROGRAM DESIGN AND EVALUATION ISSUES

- ❑ How should incentives be structured? Size of awards, intermediate milestones, cash vs in-kind?
- ❑ Treat individual students or entire schools?
- ❑ Evaluation strategies: simple random assignment, instrumental variables, clustered assignment?

We tried two variations:

1. *Student intervention*: A small pilot study that randomized pupils within schools through a hybrid instrumental-variables scheme.
2. *School intervention*: A Group-randomized Trial (GRT) that randomly assigned treatment to all pupils in 20 out of 40 low-achieving schools.

THE PILOT STUDY: STUDENT RANDOMIZATION

- ❑ Randomized treatment of select groups of pupils within schools. Minimal school-level participation.
- ❑ Offered choice of: \$800 cash; \$1,000 travel voucher; \$1,200 education voucher.
- ❑ Program began in March 2000; 500 seniors in T and C groups.
- ❑ “Hidden” random assignment: We used a hybrid assignment mechanism in a two-stage process to make randomization more palatable.

Results

- ❑ The program proved cumbersome to implement.
- ❑ No apparent effect, possibly signaling importance of school participation, though there were other design weaknesses.

DESIGN OF THE SCHOOLS INTERVENTION

- ❑ Award schedule (Bagrut requires 20 credit units; 2-5 per subject)

Grade	Milestone	Award (NIS)
10	Tested for at least 1 unit; enrolled in 11 th grade	500
	Passed this test	1500
11	Tested for at least 3 units; enrolled in 12 th grade	500
	Passed this/these test(s)	1500
12	Completed 14 units	1000
	Completed 20; awarded Bagrut	5000

- ❑ Total at stake was NIS10,000, about \$2,400. Real money for our population; on-par with the annual cost of programs in SDDAP.

EXPERIMENTAL DESIGN

- ❑ We identified 40 schools with low 1999 Bagrut rates, but above 3%. Treatment randomly assigned to all students in 20 schools.
- ❑ Schools were paired on the basis of their 1999 Bagrut rates, with one treatment school in each pair to improve T-C balance.
- ❑ The control school in pair 6 closed.
- ❑ The program was launched at an orientation and pep rally for school administrators in January 2001.

Descriptive statistics: Table A1.

- ❑ 10 Arab schools and 10 religious schools
- ❑ 5 treated schools are *non-compliers*.

DETAILS

- We look at first-year results for seniors, for whom NIS6,000 was at stake.

- Adverse publicity in May 2001, when 2000 scores were released. Victor and pilot participants defended.

- The Ed Ministry issued a press release indicating the program would run for the first year but would then be suspended and re-evaluated.

- Because of a budgetary crisis, bonuses not paid until Fall 2002.

WAS THERE A PROGRAM?

These problems raise the question:

Who/knew cared about Achievement Awards?

- ✎ In October 2001 the Ministry surveyed participants. Response rate was low and respondents probably negatively selected.
- ☐ 53% of responding seniors recalled specific program features; 80% of these recalled written program info or a school assembly.
- ☐ 87% of recallers said bonus was large enough to induce extra effort; about half reported they increased Bagrut prep as a result.
- ☐ Average hours of study went from 2.4/week in control schools to 2.7/week in treatment schools (consistent with: caused 25% pupils to study an *extra* 2 hours/week for 3 months).

ECONOMETRIC ISSUES

A. Improving Treatment-Control balance

- ❑ Cluster randomization may fail to balance when few clusters are randomized.
- ❑ Schools selected for treatment have higher pre-treatment Bagrut rates.

We therefore present results for 2000 as a spec check and use 2000 data for differences-in-differences type control strategies.

B. Adjusting for non-compliance

- ❑ While other adjustments are possible (e.g., IV), we focus here on intention-to-treat effects in regressions that control for individual and school-level covariates.

C. Mode of Inference for GRTs (schools vs. students)

Grouped Analyses

- ❑ A simple and conservative strategy treats school means as the unit of observation.
- ❑ Grouped estimates might be weighted by school size.

Micro-regressions

- ❑ Inference using means may be too conservative since it treats within-school obs as uninformative beyond the effect on cross-school dispersion.
- ❑ In practice, school means are highly unstable. We'd like to use individual characteristics to reduce variability.

Note: we must adjust for school-level clustering; in practice, the standard cluster estimator can be biased. We therefore use Bell and MacCaffrey's (2002) Bias-Reduced Linearization (BRL) cluster adjustment.

RESULTS

- The Figures plot school average Bagrut rates; 2001 against 2000, by treatment status. This is implicitly an unweighted grouped analysis

Clear evidence of a treatment effect.

- Table 1 reports Logit marginal effects and OLS estimates of treatment effects with alternative controls:

2001 estimates are relatively insensitive to controls; the 2000 spec check improves with controls.

- A key finding: treatment effects are for girls only. These are (marginally significant). Although there are some pre-program differences, none are significant.

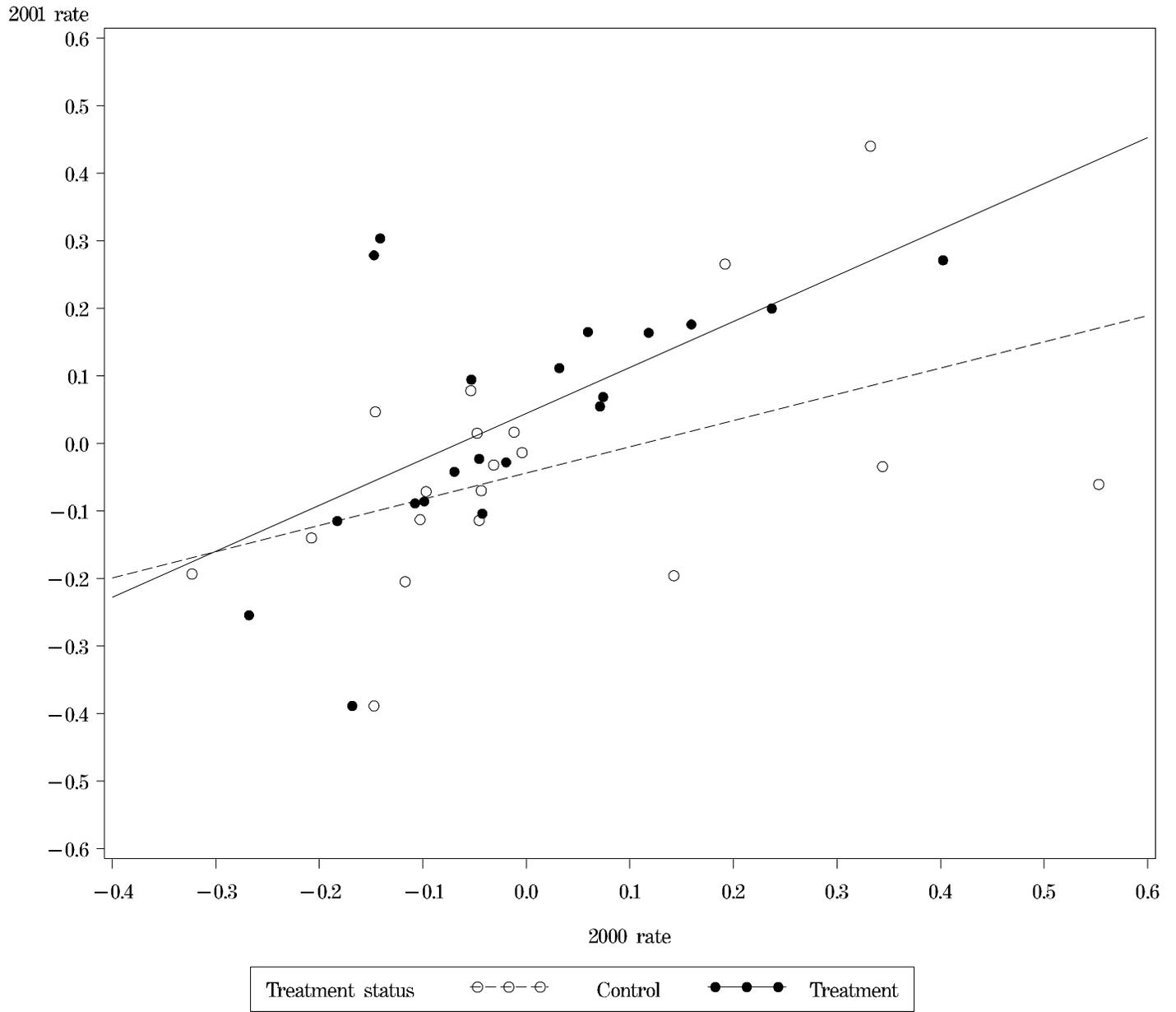


Figure 1. 2001 Bagrut rate vs. 2000 Bagrut rate by treatment status.
Residuals from regressions on school covariates.

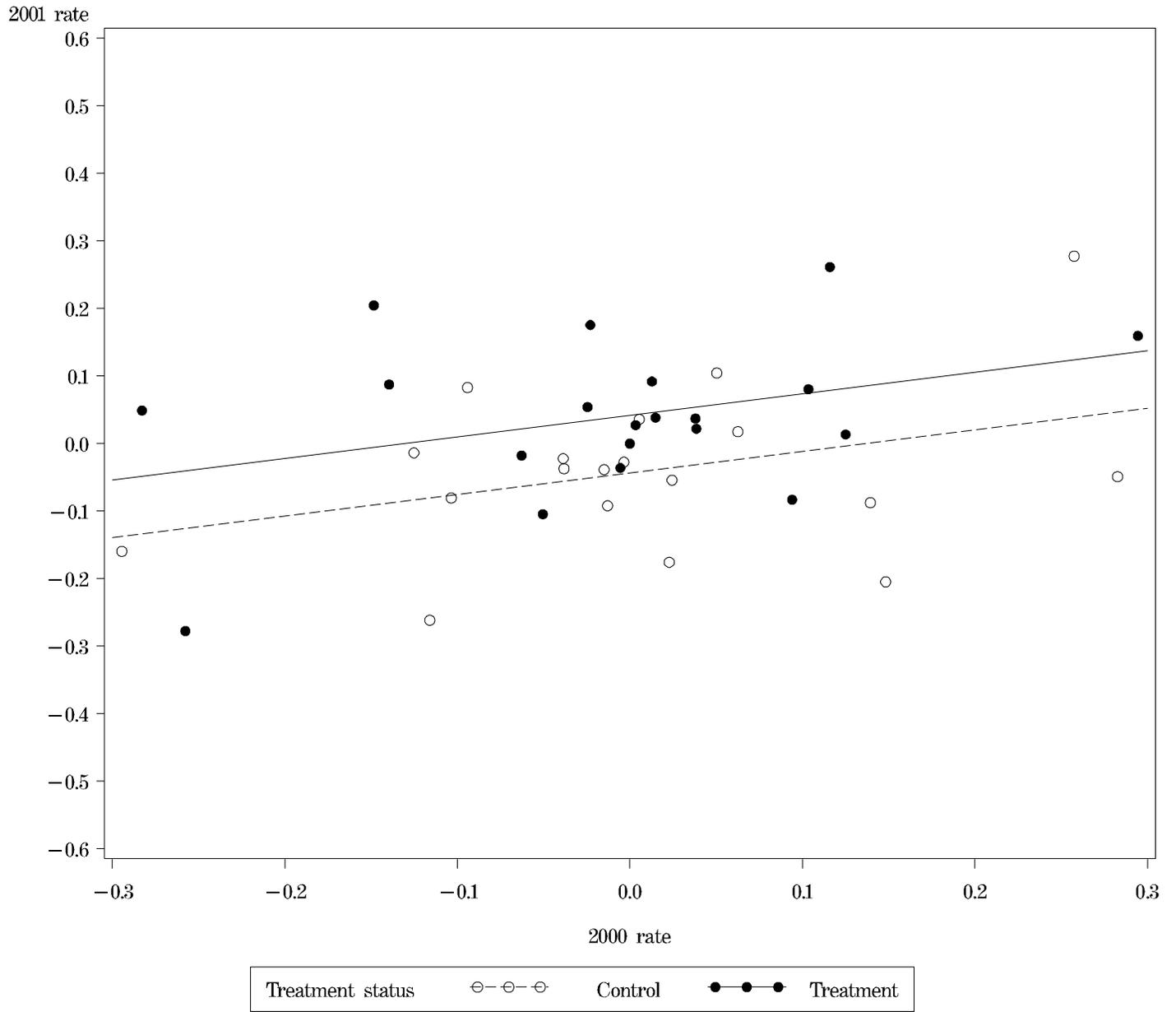


Figure 2. 2001 Bagrut rate vs. 2000 Bagrut rate by treatment status.
Residuals from regressions on school covariates and pair effects.

Table 1 - Treatment Effect for 2000 and 2001

	Pair effects	Boys + Girls		Boys		Girls	
		OLS (1)	Logit (2)	OLS (3)	Logit (4)	OLS (5)	Logit (6)
A. 2001							
<i>Dep. Mean</i>		.243		.200		.287	
Sch covs	No	.056 (.049)	.051 (.045)	-.010 (.052)	-.011 (.055)	.105 (.061)	.093 (.053)
	Yes	.052 (.047)	.054 (.043)				
Sch covs + quartiles	No	.055 (.039)	.050 (.038)	-.004 (.046)	-.004 (.047)	.095 (.047)	.093 (.046)
	Yes	.064 (.036)	.054 (.036)				
Sch covs + quartiles + micro covs	No	.052 (.039)	.047 (.039)	-.022 (.043)	-.023 (.045)	.105 (.047)	.097 (.046)
	Yes	.068 (.036)	.055 (.036)				
<i>Number of students</i>		3,821		1,960		1,861	
<i>Number of schools</i>		39		34		34	
B. 2000							
<i>Dep. Mean</i>		.224		.177		.272	
Sch covs	No	.050 (.056)	.046 (.051)	.045 (.060)	.040 (.055)	.075 (.067)	.069 (.061)
	Yes	.043 (.059)	.045 (.058)				
Sch covs + quartiles	No	.031 (.043)	.028 (.043)	.031 (.052)	.026 (.050)	.058 (.048)	.055 (.047)
	Yes	.040 (.045)	.034 (.047)				
Sch covs + quartiles + micro covs	No	.030 (.041)	.018 (.042)	.009 (.050)	.006 (.052)	.067 (.046)	.051 (.046)
	Yes	.043 (.044)	.030 (.046)				
<i>Number of students</i>		4,039		2,038		2,001	
<i>Number of schools</i>		39		33		35	

Notes: BRL standard errors are in parentheses.

Reference:

Panel A: msdavidj/BONUSES/ECTA/table2/01

Panel B: msdavidj/BONUSES/ECTA/table2/00

ANALYSIS IN SUBGROUPS

- ❑ We'd like to identify students who were most likely to have benefitted from extra effort or other responses to bonus incentives.
- ❑ For many students, it's too late.

Subgroup models

- ❑ We used two strategies to find the right group:
 - (a) Lagged-score quartiles
 - (b) Predicted Bagrut success as a function of lagged-score quartiles and family background
- ❑ Predictive models reported in Table 2; Results in Table 3 (Panels A & B).

Effects are for girls only, in the top half of the subgroup-classification distribution.

Table 2 - Determinants of Bagrut Status, 2001

	Boys+Girls		Boys		Girls	
	(1)	(2)	(3)	(4)	(5)	(6)
<i>Dep. Mean</i>	.219		.194		.239	
<i>Covariates</i>						
Father schooling	.076 (.044)	.056 (.040)	.081 (.053)	.016 (.047)	.068 (.039)	.074 (.038)
Mother schooling		.034 (.037)		.109 (.079)		-.012 (.032)
Has more than 4 siblings	-.431 (.307)	-.422 (.308)	.080 (.521)	.093 (.547)	-.682 (.254)	-.685 (.255)
Immigrated in last 5 years	.924 (.383)	.905 (.385)	1.307 (1.204)	1.351 (1.196)	1.104 (.364)	1.116 (.364)
Arab school	.987 (.540)	1.077 (.558)	.625 (.830)	.901 (.905)	1.298 (.420)	1.264 (.400)
Religious school	.627 (.579)	.632 (.578)	2.560 (.603)	2.573 (.628)	.283 (.508)	.281 (.508)
Lagged score quartile - 2	2.826 (.739)	2.838 (.738)	1.771 (.871)	1.780 (.880)	3.161 (.784)	3.158 (.785)
Lagged score quartile - 3	4.344 (.662)	4.354 (.664)	3.813 (.788)	3.855 (.776)	4.686 (.781)	4.686 (.781)
Lagged score quartile - 4	5.018 (.581)	5.014 (.583)	4.570 (.629)	4.556 (.625)	5.328 (.738)	5.333 (.735)
<i>Number of students</i>	1,876		850		1,026	
<i>Number of schools</i>	19		15		18	

Notes: The table reports logit estimates. BRL standard errors are in parentheses. Estimates were constructed using the sample of control schools only.

Reference:

msdavidj/BONUSES/ECTA/table3/PanelA

Table 3 - Estimates in Covariate Subgroups

	By lagged score						By probability of success					
	Boys + Girls		Boys		Girls		Boys + Girls		Boys		Girls	
	Top	Bottom	Top	Bottom	Top	Bottom	Top	Bottom	Top	Bottom	Top	Bottom
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
A. 2001												
<i>Dep. Mean</i>	.435	.047	.365	.035	.518	.056	.438	.047	.368	.032	.518	.056
<i>Models</i>												
Sch covs + group main effect	.103 (.064)	-.006 (.019)	-.013 (.083)	.007 (.016)	.206 (.079)	-.020 (.024)	.088 (.064)	-.005 (.019)	-.047 (.077)	.005 (.016)	.194 (.077)	-.015 (.023)
Sch covs + group linear control	.109 (.064)	-.009 (.019)	-.009 (.083)	.007 (.017)	.213 (.079)	-.021 (.022)	.097 (.064)	-.011 (.019)	-.044 (.079)	.001 (.017)	.207 (.078)	-.019 (.026)
<i>Number of students</i>	1,926	1,895	980	980	933	928	1,912	1,909	980	980	932	929
B. 2000												
<i>Dep. Mean</i>	.398	.051	.318	.035	.475	.068	.399	.049	.320	.033	.478	.066
<i>Models</i>												
Sch covs + group main effect	.058 (.065)	-.003 (.024)	.055 (.079)	-.014 (.035)	.098 (.074)	.009 (.027)	.042 (.065)	.000 (.024)	.033 (.078)	.004 (.027)	.086 (.071)	.009 (.023)
Sch covs + group linear control	.057 (.065)	-.004 (.024)	.055 (.079)	-.014 (.035)	.094 (.072)	.007 (.026)	.036 (.064)	-.004 (.025)	.010 (.077)	.000 (.028)	.089 (.070)	.007 (.024)
<i>Number of students</i>	2,020	2,019	1,022	1,016	1,004	997	2,021	2,018	1,021	1,017	1,002	999
C. Stacked 2001 and 2000												
<i>Dep. Mean</i>	.416	.056	.341	.050	.497	.074	.418	.056	.344	.048	.498	.072
<i>Models</i>												
Sch covs (t.v) + group main effect	.073 {.036}	-.008 {.005}	-.048 {.046}	-.004 {.004}	.119 {.053}	-.018 {.010}	.062 {.036}	-.007 {.006}	-.031 {.046}	-.007 {.005}	.102 {.053}	-.013 {.008}
Sch covs (t.v) + group linear control	.083 {.036}	-.005 {.003}	-.036 {.047}	-.002 {.003}	.130 {.053}	-.012 {.007}	.082 {.037}	-.010 {.006}	-.006 {.048}	-.016 {.008}	.114 {.053}	-.016 {.010}
<i>Number of students</i>	3,946	3,421	2,002	1,395	1,931	1,613	3,933	3,381	2,001	1,355	1,930	1,639

Notes: The table reports Logit marginal effects. BRL standard errors are shown in parentheses. Conventional standard errors are in braces.

Reference:

Panel A: msdavidj/BONUSES/ECTA/table4/01

Panel B: msdavidj/BONUSES/ECTA/table4/00

Panel C: V:\Josh\Bonuses\ECTA\Table4\DID

FURTHER CHECKS: DIFFERENCES-IN-DIFFERENCES

Is there a significant difference between the results for 2001 and 2000?

- We estimated DID models that control for school fixed effects.
- These models can also be thought of as capturing a (weighted) difference in treatment effects between 2001 and 2000. In other words, we measure the 2001 effect relative to the 2000 baseline.
- DID results are reported in Panel C of Table 3.
- The pattern of effects remains as before:

Substantial and significant effects on girls in upper half of the classification-variable distribution

No effect on boys, top or bottom.

HOW DID THEY DO IT?

- ❑ The Bagrut is obtained by clearing a series of hurdles

20+ credit units attempted and awarded

Distribution requirements

Composition requirement (2 units)

Math requirement (3-5 units)

English requirement (3-5 units)

(Also, 1 advanced subject for 5 units)

- ❑ Table 4 reports results for some of these mediating outcomes for students in the upper half of the classification-variable distribution only.
- ❑ Results indicated treated girls attempted more units; passed more units; were more likely to meet Math and Writing distribution requirements.

Table 4 - Mediating Outcomes

Outcome variable	Boys				Girls			
	2001 control	2000	2001	Stacked	2001 control	2000	2001	Stacked
	mean	(1)	(2)	(3)	mean	(4)	(5)	(6)
<i>Units attempted</i>								
18	.721	.059 (.055)	.049 (.065)	.012 {.039}	.822	.054 (.055)	.091 (.044)	.026 {.034}
20	.662	.050 (.063)	.063 (.066)	.037 {.042}	.757	.026 (.053)	.127 (.053)	.073 {.036}
22	.594	.066 (.073)	.052 (.063)	.003 {.050}	.690	.035 (.065)	.106 (.054)	.054 {.044}
24	.502	.073 (.075)	.042 (.069)	-.014 (.054)	.574	.027 (.080)	.065 (.054)	.024 {.051}
<i>Units awarded</i>								
18	.695	.061 (.055)	.057 (.068)	.015 {.043}	.736	.089 (.059)	.156 (.053)	.054 {.032}
20	.653	.064 (.061)	.059 (.066)	.019 {.047}	.696	.071 (.057)	.150 (.059)	.065 {.036}
22	.596	.052 (.072)	.046 (.063)	.011 {.052}	.623	.102 (.065)	.150 (.058)	.056 {.045}
24	.498	.094 (.080)	.046 (.065)	-.045 {.055}	.533	.075 (.078)	.126 (.066)	.057 {.053}
<i>Distribution requirements</i>								
Math	.563	-.007 (.082)	.004 (.063)	.025 {.055}	.625	.054 (.072)	.153 (.059)	.099 {.048}
English	.664	.107 (.062)	.082 (.057)	-.001 {.048}	.726	.131 (.069)	.111 (.048)	-.011 {.044}
Writing	.702	.015 (.058)	-.003 (.062)	.002 {.044}	.777	-.005 (.054)	.098 (.041)	.093 {.030}

Notes: The table reports Logit marginal effects estimated in models with school covariates, using data for students in the upper half of the lagged score distribution. BRL standard errors are in parentheses. Conventional standard errors are in braces. The logit estimate in the last row of column 10 failed to converge; this is an OLS estimate.

Reference:

2000+2001: msdavidj/BONUSES/ECTA/table5

Stacked: V:\Josh\Bonuses\ECTA\Table5\DID

SUMMARY AND CONCLUSIONS

- ❑ Results from the schools GRT point to an increase of 5-7% in Bagrut rates in the treatment group.
- ❑ These results are driven entirely by girls; the effect on boys is zero.
- ❑ Effects are for the girls relatively likely to obtain a Bagrut anyway (though likelihood of success is still low).
- ❑ Analysis of mediating outcomes suggests treated girls attempted and passed more units, and were also more likely to meet certain distribution requirements.
- ❑ A theoretical implication is that modest learning incentives can be important for marginal groups. The program seems very likely to pass a cost-benefit test.

APPENDIX: ACHIEVEMENT AWARDS PROGRAM RULES AND TIMING

Bagrut structure

High school includes grades 10-12, with an academic track leading to Bagrut, and a non-Bagrut diploma track. Most pupils in non-vocational schools are on the academic track. The Bagrut consists of a series of exams with the following features:

1. Elective subjects beginning in 10th
2. More tests in 11th and most in 12th
3. Tests at various proficiency levels awarding 1-5 units per subject
4. 20 units required for certificate; some subjects mandatory
5. Tests are given in June with an opportunity for retest in the winter.
6. Tests are graded externally and internally; internal grading is policed.

About 52% of all seniors received a Bagrut in 1999 and 2000. 60% of those taking at least one test received a certificate, compared with only about 20% in our sample.

Rules for School Intervention

1. Award schedule

Grade	Milestone	Reward (NIS)
10	Tested for at least 1 unit; enrolled in 11 th grade	500
	Passed this test	1500
11	Tested for at least 3 units; enrolled in 12 th grade	500
	Passed this/these test(s)	1500
12	Completed 14 credit units	1000
	Completed 20 credit units and awarded bagrut	5000

2. Tests are considered to have been passed if the external component is passed.
3. Only tests in required subjects are eligible for intermediate awards. At the time this program was introduced (January 2001), the required subjects were Bible (2 units), literature (2 units), history (2 units), civics (2 units), composition (2 units), english (3 units), mathematics (3 units). The remaining 5 units can be in any Bagrut-eligible elective subject. Many pupils, e.g. those competing for admission to selective universities, obtain more than the minimum number of credit units.
4. Awards for achievement in a given year are to be paid in the following school year.
5. All pupils in treatment schools are eligible.
6. Pupils with at least 14 units have two chances to take Bagrut exams in 12th grade. Awards will be given to those who pass on the first, second, or any subsequent try.

PROGRAM AND DATA COLLECTION TIME LINES

Program	Schools selected and principals informed	Orientation for principals; pupils informed	Baseline data	Media coverage	Bagrut tests	Student survey	Re-test	Winter re-test
School randomization (GRT)	December 2000	January 2001	January 2001	May 2001	June 2001	Aug/Sept/Oct 2001	Aug-Sept 2001 (math and english)	December 2001 - January 2002

Note: in March 2001 we called principals to check whether the program was publicized in schools.

Bonuses paid in Fall 2002; after Winter scores collated and resolution of a budgetary crisis.

Student randomization (Pilot)	March 2000	March 2000 (principals and pupils informed in writing)	March 2000	September 2000	June 2000	July/Aug/Sept 2000	--	December 2000 - January 2001
-------------------------------	------------	---	------------	----------------	-----------	--------------------	----	------------------------------

Bonuses for the June test were first paid in November 2000. Additional bonuses were paid in 2001. Payments were covered by a private donor.