NBER WORKING PAPER SERIES

# TEACHERS' PAY FOR PERFORMANCE IN THE LONG-RUN: EFFECTS ON STUDENTS' EDUCATIONAL AND LABOR MARKET OUTCOMES IN ADULTHOOD

Victor Lavy

Working Paper 20983
http://www.nber.org/papers/w20983

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 2015

Excellent research assistance was provided by Boaz Abramson, Elior Cohen, and Michal Hodor. I thank Jo Altonji, Josh Angrist, Zvi Eckstein, Nathaniel Hendren, Ed Lazear, Yona Rubinstein and participants at seminars at Hebrew University, University of California Santa Barbara, IDC Herzelia, University of Warwick, NBER 2015 Summer Institute Personnel Economics conference, CEPR 2016 Public Economics Annual Symposium, Warwick-Venice 2016 Labor Economics Conference and the Barcelona 2016 Summer Forum for useful comments and suggestions. I thank Israel's National Insurance Institute (NII) for allowing restricted access to post-secondary schooling and economic and social outcomes data at adulthood in the NII protected research lab. I acknowledge financial support from the European Research Council through ERC Advance Grant 323439, the Israeli Science Foundation and the Maurice Falk Institute in Jerusalem. The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

Teachers’ Pay for Performance in the Long-Run: Effects on Students’ Educational and Labor Market Outcomes in Adulthood
Victor Lavy
NBER Working Paper No. 20983
February 2015, Revised July 2016
JEL No. J24,J3

## ABSTRACT

This paper examines the dynamic effects of teachers’ pay for performance experiment on long-term outcomes at adulthood. The program led to a gradual increase in university schooling of the high school treated students, reaching a gain of 0.25 years of schooling at age 28-30. The effects on employment and earnings were initially negative, coinciding with a higher enrollment rate in university schooling, but became positive and significant at a later age. These gains are largely mediated by the positive effect of the program on several high school outcomes, including quantity and quality gains in the high stake matriculation exams.

Victor Lavy
Department of Economics
University of Warwick
Coventry, CV4 7AL
United Kingdom
and Hebrew University of Jerusalem
and also NBER
v.lavy@warwick.ac.uk

## 1. Introduction

The long term effect of teachers' pay for performance schems is of particular interest because it is often claimed by skeptics of these schems that they are only improving student test scores by teaching to the test or by enacuoraging cheating orchestrated by teachers and schools. Therefore, it is claimed, there is no actual improvement in human capital because teachers do not respond to pay incentives by promoting broad human capital acquisition and instead focus on improving students' test taking ability, on testing preparation sessions at the expense of teaching material not included in the exam, on strageties of answering exam questions such as multiple questions and on skills and actions that raise scores on the formulas used to reward teachers.[1] Concerns about narrowly targeted gains are heightened if those gains are focused on realms where labor market rewards are due to signaling rather than human capital acquisition. To address this key issue, I examine the effect of teachers' pay for performance on long term human capital outcomes, in particular attainment and quality of higher education, and labor market outcomes at adulthood, in particular employment and earnings. For this I take advantage of a teachers' pay for performance experiment which I conducted a decade and half ago in Israel. In Lavy (2009) I analyzed the short-term effects on students' cognitive schooling outcomes of this experiment. With the passage of time, this previous research endeavor now presents an unusual opportunity to evaluate whether an intervention that offered teachers performance-based bonuses for student test achievements has a lasting and long-term impact on adult well-being. This paper provides the first evidence of links between teachers' pay for performance during high school and students' schooling and labor market outcomes during their late 20s and early 30s. I examine the impact on various long-term outcomes, including post-secondary educational attainment, employment, earnings, eligibility for unemployment benefit and marriage and fertility. Some of these outcomes, for example the later two, can also be viewed as potential mechansims for the effect of the intervention on employment and earnings.

I observe these students' outcomes every year, from high school graduation (2000-2001) until age 30 (in 2012). This permits to estimate the treatment effects for every year in this period and thus trace the dynamical evolution of the program effect. Since a high proportion of the sample was in military service for two (female) or three (male) years after high school[2], the estimates for these years (2000-2004) are not very informative because they are based on a small and selective sample of those students not enlisted into military service. However, during 2005-2008 the treated group experienced a higher enrollment rate in university schooling. As a mirror image and in perfectly synchronized timing, the treatment group had experienced during this period lower employment rate and lower earnings. By the end of this period these negative treatment effects were eliminated and the earnings treatment effect turned positive, increasing in size and becoming significantly different from zero 9-12 years after high school graduation. So just over a decade after the end of the intervention, treated students are 5 percentage points more likely to be enrolled in university education and to complete an additional 0.25 years of university schooling, a 35 percent increase relative to the control group mean. These gains are most likley explained by the improvments in high school 'Bagrut' outcomes facilitated by the teachers' PFP intervention. The higher passing rate and average score in the math and English matriculation exams (Lavy 2009), are also expressed in imporvements in average matricualtion outcomes, such as matriculation diploma certification (up by 3 percent) and the overall composite matricualtion score (up by 2.9 points). These two outcomes determine addmision to university schooling and to selective study programs, such as  medical, engineering and computer science schools. Other dimensions of the matricualtion study program that signal quality of schooling were also improved, in particular the number of science credit units which increased by 26 percent and the number of subjects studied at most advanced level which increased by 5 percent. These high school outcomes are also highly correlated with labor market outcomes at adulthood. Consistently, I find that these improvements along with the increase in university schooling led to a 1.2 perecent gain in employment rate and to 6-7 percent increase in earnings at age 28-30. The estimates suggest that the program did not effect marriage and parenthood rates by age 30. These average gains mask some heterogeneirty by family income and gender. For example, chidlren from above median family

[1] See Jacob and Levitt (2003), Glewwe, Ilias and Kremer (2010), Neal (2011) and Muralidharan and Sundararaman (2011) for a discussion of this issue.

income experience higher increase in schooling but no effect on employment while the children from below median family income have lower gain in schooling and a large positive effect on employment. The effect on earnings is the same for both groups. Children from below median family income also experienced a significant decline in marriage rate and a more modest decline in fertility.

The lessons learned from this analysis have meaningful external validity and can be easily transferable and applicable to other educational settings in developed countries. Both the high school system in Israel and its high-stakes exit exams are very similar to those in other countries. Importantly, variants of the teachers' pay for performance intervention studied here have been implemented in recent years in developed and developing countries. Therefore, this study contributes to the buildup of empirical evidence that facilitates a more complete picture of the returns to educational interventions that will provide policymakers with information needed to steer limited resources toward the most effective programs. Another important advantage of the evidence presented in this paper is that teachers' pay for performance is an intervention or a policy change that can be directly manipulated by public policy, relative to evidence based on measures such as school or teacher's quality which are not easily measured or manipulated by policy.

This paper is an addition to a recently growing literature on long term effects of educational programs. Earlier studies focused on the long-term effects of compulsory schooling laws on adult educational attainment (Angrist and Krueger, 1991) and on adult health (Lleras-Muney, 2005), for example. More recent studies have turned their attention to schooling programs aimed at improving the quality of education in addition to increasing attainment. However, most of these studies have centered on the evaluation of short-term outcomes, primarily standardized test scores, as a measure of success. Nevertheless, an equaly lingering question is the extent to which educational interventions lead to long-term improvements in well-being – measures assessed not by attainment on tests but by attainment in life. Puzzling and conflicting results from several evaluations make this a highly salient issue. Three small-scale, intensive preschool experiments produced large effects on contemporaneous

[2] Israelis begin a period of compulsory military service after high-school graduation. Boys serve for three years and girls for two (longer if they take a commission). Arabs and Jewish Orthodox religious girls are exempted from military service.

test scores that quickly faded (Schweinhart et al., 2005; Anderson, 2008). Non-experimental evaluations of Head Start, a preschool program for poor children, revealed a similar pattern, with test-score effects gone by middle school. But in each of these studies, treatment effects re-emerged in adulthood in the form of increased educational attainment, enhanced labor market attachment, and reduced crime (Deming, 2009; Garces et al., 2002; Ludwig and Miller, 2007). Other studies have shown evidence for the effect of investments in childhood on postsecondary attainment (Krueger and Whitmore 2001, Dynarski et al 2011). Very recently, Chetty et al (2011 and 2014) examined the longer-term effect of value-added measures of teachers' quality in a large urban school district in the United States, and reported significant effects on earnings at age 27 even though the effect on test scores had faded away much earlier. Dustmann et al (2012), however, found that attending a better school in Germany had no effect on school attainment or labor-market outcomes. However, there are no studies that focus on the long term effect of teachers' performance pay programs even though this question is of utmost importance because the ultimate goal of education is to improve lifetime well-being and there is much uncertainty about the long term gains from such programs. Determining which interventions are more effective in achieving better long-term outcomes is critical for improving the efficiency of education and school resource allocation.

The remainder of this paper is as follows. Section 2 describes the Pay for Performance Experiment and the Identification and econometric model. Section 3 describes the data and Section 4 presents the empirical findings. Section 5 concludes.

## 2. The Pay for Performance Experiment

Teacher incentive programs are increasing in popularity. Performance-related pay for teachers is being introduced in many countries, amidst much controversy and opposition from teachers and unions. The rationale for these programs is the notion that incentive pay may motivate teachers to improve their performance (Lazear 2000 and 2001, Lavy 2002 and 2007, Neal 2011, Duflo et al. 2012). Opponents of teachers' incentive programs argue that schools may respond to test score-based incentives in perverse ways such as cheating in grading and teaching to the test (Glewwe et al, 2010, Neal 2011) so that any gain in performance is short-lived without real long-term accumulation of human capital. Even though there is some evidence that performance pay for teachers have significant short term learning benefits, their critique is focused on their harmful long term effects although the evidence is scanty. The evidence presented in this paper fills in this gap with results on a wide array of lifetime outcomes.

***The Teachers' Incentive Experiment***

In early December 2000, the Ministry of Education unveiled a new teachers' bonus experiment in 49 Israeli high schools that I designed and helped to implement. The main feature of the program was an individual performance bonus paid to teachers on the basis of their own students' achievements. The experiment included all English, Hebrew, Arabic, and mathematics teachers who taught classes in grades 10 through 12 in advance of matriculation exams in these subjects in June 2001. The ranking was based on the difference between the actual outcome and a value predicted on the basis of a regression that controlled for the students' study program and socioeconomic characteristics, and a fixed school level effect. Separate regressions were used to compute the predicted pass rate and mean score, and each teacher was ranked twice – once for each outcome – using the size of the residual from the regressions. All teachers whose students' mean residual was positive in both outcomes were divided into four ranking groups, from first place to fourth. The first place award was $7,500; second place, $5,750; third place, $3,500; and fourth place, $1,750.

Schools eligible for the program met two criteria: 1) they had a recent history of relatively poor performance in the mathematics or English matriculation exams, and 2) the most recent school-level matriculation rate was equal to or lower than the national mean of 45 percent. Though 99 schools met the first criterion, only 49 met the second criterion on the matriculation rate. The program included 629 teachers. Nearly half of the teachers, 302 of them, won awards—94 in English, 124 in math, 67 in Hebrew and Arabic, and 17 in other subjects. Although the program was designed as an experiment, schools were not assigned to it at random. Nevertheless, the design of the program enables the implementation of a randomized trial identification strategy which I outline below. The results of the short term effect, presented in Lavy (2009), suggested that teachers' incentives increased students' achievements by increasing the test taking rate as well as the conditional pass rate and test scores in math and English exams. The improvement in these conditional outcomes, which were estimated based on tests and grading external to schools, accounted for more than half of the increase in the unconditional outcomes in math and somewhat less in English. These improvements appear to result from changes in teaching methods, after-school teaching, and increased responsiveness to students' needs, and not from artificial inflation or manipulation in test scores. The evidence that incentives induced improved effort and pedagogy is important in the context of the recent concern that incentives may have unintended effects, such as "teaching to the test" or cheating and manipulation of test scores, and that they do not generate real learning. However, more conclusive evidence about whether teachers' pay for performance schemes improve real human capital accumulation can only be based on longer term effects, in particular on completed post-secondary schooling, employment, wages, welfare dependency and crime, which I intend to investigate in this proposal.

***Identification and econometric model: measurement error in the assignment variable***

The program rules limited assignment to schools with a 1999 matriculation rate equal to or lower than 45 percent (43 percent for religious and Arab schools). However, the matriculation rate used for assignment was an inaccurate measure of this variable. This measurement error could be useful for identification of the program effect. In particular, conditional on the true matriculation rate, program status may be virtually randomly assigned due to mistakes in the preliminary file. Most (80 percent) of the measurement errors were negative, 17 percent were positive, and the rest were free of error. Identification based on the random measurement error can be presented formally as follows:

Let $S = S^* + \varepsilon$ be the error-affected 1999 matriculation rate used for the assignment, where $S^*$ represents the correct 1999 matriculation rate and $\varepsilon$ the measurement error. $T$ denotes participation status, with $T = 1$ for participants and $T = 0$ for non-participants. Since $T(S) = T(S^* + \varepsilon)$, once we control for $S^*$, assignment to treatment is random ("random assignment" to treatment, conditional on the true value of the matriculation rate). The presence of a measurement error creates a natural experiment, where treatment is assigned randomly, conditional on $S^*$, in a sub-sample of the 98 eligible schools. Eighteen of the eligible schools had a correct 1999 matriculation rate above the threshold line. Thus, these schools were "erroneously" chosen for the program. For each of these schools, there is a school with an identical correct matriculation rate but with a draw from the (random) measurement error distribution which is not large (and negative) enough to drop it below the assignment threshold. Such pairing of schools amounts to non-parametrically matching schools on the basis of the value of $S^*$. Therefore, the eighteen untreated schools may be used as a control group that reflects the counterfactual for identification of the effect of the program. The group of 18 treated and 18 control schools are perfectly balanced in student and school characteristics (see Table A1 in the online appendix which is reproduced from Lavy 2009 Table 3). The following model is used as the basis for regression estimates using the NE sample:

$$Y_{ijt} = \alpha + X_{ijt}{}' \beta + Z_{jt}{}' \gamma + \delta T_{jt} + \Phi_j + \eta D_t + \varepsilon_{ijt}$$

where $i$ indexes individuals; $j$ indexes schools; t indexes years 2000 and 2001, and $T$ is the assigned treatment status. $X$ and $Z$ are vectors of individual and school level covariates and $D_t$ denotes year effects with a factor loading $\eta$. The treatment indicator $T_{jt}$ is equal to the interaction between a dummy for treated schools and a dummy for the year 2001. The regressions will be estimated using pooled data from both years (the two adjacent cohorts of 2000 and 2001), stacked as school panel data with fixed school-level effects ($\Phi_j$) included in the regression. The resulting estimates can be interpreted as an individual-weighted difference-in-differences procedure comparing treatment effects across years. The estimates are implicitly weighted by the number of students in each school. The introduction of school fixed effects controls for time-invariant omitted variables and also provides an alternative control for school-level clustering.

***Identification based on a Regression Discontinuity model***

To check the robustness of the results based on the NE sample, I also use two additional alternatives methods to identify the effect of the teacher bonus program. The first is an RD design. Given that the rule governing selection to the program was simply based on a discontinuous function of a school observable, the probability of receiving treatment changes discontinuously as a function of this observable. The discontinuity in our case is a sharp decrease (to zero) in the probability of treatment beyond a 45 percent school matriculation rate for nonreligious Jewish schools and beyond 43 percent for Jewish religious schools and Arab schools. I exploit this sharp discontinuity to define a treatment sample that included schools that were just below (up to -5 percent) the threshold of selection to the program and a comparison group that included untreated schools that were just above (up to +5 percent) this threshold. The time series on school matriculation rates show that the rates fluctuate from year to year for reasons that transcend trends or changes in the composition of the student body. Some of these fluctuations are random. Therefore, marginal (in terms of distance from the threshold) participants may be similar to marginal nonparticipants. The degree of similarity depends on the width of the band around the threshold. Sample size considerations exclude the possibility of a bandwidth lower than 10 percent, and a wider band implies fluctuations of a magnitude that are not likely to be related to random changes. Therefore, a bandwidth of about 10 percent seems to be a reasonable choice in our case. The main drawback of this approach is that it produces an estimate from marginally (relative to the threshold) exposed schools only. However, this sample may be of particular interest because the threshold schools could be representative of the schools that such programs are most likely to target.

There are 13 untreated schools with matriculation rates in the 0.46–0.52 range and 14 treated schools in the 0.40–0.45 range. The 0.40–0.52 range may be too large, but I can control for the value of the assignment variable (the mean matriculation rate) in the analysis. Note, also, that there is some overlap between this sample and the natural experiment sample. Eleven of the 14 treated schools and 8 of the 11 control schools in the RD sample are also part of the natural experiment sample, leaving only six schools (3 control and 3 treated), which are included in the former but not in the latter. However, there are 17 schools in the NE sample (7 treated and 10 control) that are not included in the RD sample, which suggests that there is enough “informational value added” in each of the samples.

Table A2 in the online appendix (reproduced from Lavy 2009 Table 6) is similar to Table A1 but for the RD sample. The treatment-control differences and standard errors in the student background variables (columns 3 and 6) reveal that the two groups are very similar in both years in all characteristics, except the ethnicity variable in year 2000 and number of siblings in 2001. However, both estimated differences are only marginally significant. The third panel reveals some treatment-control differences: in math lagged credits and in the average score for the 2001 cohort, and in English lagged credits for the 2000 cohort. But the control-treatment gaps in lagged credits are opposite in sign in math (negative) and English (positive) and in each subject they are significant for only one of the two cohorts.

To interpret similarity or differences between the estimates of the two methods, it is important to note that the main conceptual difference between the NE and the RD methods is that the latter does not control for S* and, if there were no measurement errors, the RD design would have compared, in addition, pupils or schools with different S*. The two methods would therefore yield similar results if either S* is weakly related to the outcomes or if the variance of the measurement error is large relative to the variance of S* around the cutoff point (which means that those above and below the critical S have approximately the same S*). The first condition is met, as S* has very small positive correlations with the high school and post-secondary schooling and labor market outcomes. The second condition is not met because within the range (0.40-0.52) around the cutoff point of the assignment variable (S), the two relevant variances are very similar, 0.075 for the measurement error and 0.078 for S*.

**3. Data**

In this study I use data available in administrative files for the sample of participants in the treatment and control group of the experiment. The students in the sample graduated high school between 2000 and 2001, and in 2013 they are adults aged 30-31. I use several panel datasets available from Israel's National Insurance Institute (NII). The NII is responsible for social security and mandatory health insurance in Israel. NII allows restricted access to this data in their protected research lab. The underlying data sources include: (1) the population registry data, which contain information on marital status, number of children and their birth dates; (2) NII records of post-secondary enrollment from 2000 through 2013 based on annual reports submitted to NII every fall term by all of Israel's post-secondary education institutions. Based on this annual enrollment data I computed the number of years of post-secondary schooling[3]; (3) Israel Tax Authority information on income and earnings of employee and self-employed individuals for each year during 2000-2012. This file includes information both for the students and their father and mother; (4) NII records on

[3] The NII which is responsible for the mandatory health insurance tax in Israel tracks post-secondary enrollment because students pay a lower health insurance tax rate. Post-secondary schools are therefore required to send a

unemployment benefits for the period 2009-2012, and marriage and fertility information as of 2012. The NII linked these data to students' background data that I used in Lavy (2009) to study the effect of the teachers' incentive experiment on high school academic outcomes. This information comes from administrative records of the Ministry of Education on the universe of Israeli primary schools during the 1997-2002 school years. In addition to individual identifier, and a school and class identifier, it also included the following family-background variables: parental schooling, number of siblings, country of birth, date of immigration if born outside of Israel, ethnicity and a variety of high school and high school achievement measures. This file also included a treatment indicator, school ID and cohort of study. I had restricted access to this data in the NII research lab at the NII headquarters in Jerusalem.

**The post high school academic schooling system in Israel**

The post high school academic schooling system in Israel includes seven universities (one of which confers only graduate and PhD degrees), and over 50 colleges that confer academic undergraduate degrees (some of these also give master's degrees)[4]. All universities require a *bagrut* diploma for enrollment. Most academic colleges also require a *bagrut*, though some look at specific *bagrut* diploma components without requiring full certification. For a given field of study, it is typically more difficult to be admitted to a university than to a college. The national university enrollment rates for the cohort of graduating seniors in 1995 (through 2003) was 27.6 percent and the respective rate for academic colleges was 8.5 percent.[5]

The post-high school outcome variables of interest here are indicators of ever having enrolled in a university and in an academic college as of the 2013 school year, and the number of years of schooling completed in these two type of academic institutions by this date. I measure these two outcomes for the 2000 and 2001 12th grade students. Even after accounting for compulsory military

list of enrolled students to the NII every year. For the purposes of this project, the NII Research and Planning Division constructed an extract containing the 2001–2013 enrollment status of students in our study.

[4] A 1991 reform sharply increased the supply of post-secondary schooling in Israel by creating publicly funded regional and professional colleges.

[5] These data are from the Israel Central Bureau of Statistics, Report on Post-Secondary Schooling of High School Graduates in 1989–1995 (available at: http://www.cbs.gov.il/publications/h_education02/h_education_h.htm).

service[6], we expect that most students who enrolled in academic post-high school education, including those who continued schooling beyond undergraduate studies, to have graduated by the 2013 academic year.

**Definitions of Outcomes in Adulthood**

In this subsection, I describe the outcomes in adulthood for students in the analysis sample:

***Post-Secondary Academic Schooling***

In the NII data, I observe two sets of post-secondary outcomes for each of the students in the sample. First, I observe year-by-year post-secondary schooling attainment, including the type of post-secondary schooling institution attended –if any- and the number of years of schooling completed in each type of institution. The post-high school outcome variables of interest are indicators of ever having enrolled in a university or in an academic college, and the number of years of schooling completed in these two types of academic institutions. Even after accounting for compulsory military service, I expect that most students who enrolled in academic post-high school education, including those who continued schooling beyond undergraduate studies, to have graduated by age 30. I measure enrollment and years of schooling completed for each year from 2000 to 2013, adjusting for years since graduating high school.

***Labor Market Outcomes***

I observe year-by-year labor market outcomes from high school graduation to 2012, including employment status and annual earnings. Individual earnings data come from the Israel Tax Authority (ITA). Filing tax forms in Israel is compulsory only for individuals with non-zero self-employment earnings but ITA has information on annual gross earnings from salaried and non-salaried employment and they transfer this information annually to NII, including number of months of work in a given year. NII produces an annual series of total annual earnings from salaried and self-employment. Following NII practice, individuals with positive (non-zero) number months of work and zero or missing value for earnings are assigned zero earnings. 14.1% of individuals (students)

[6] Boys serve for three years and girls for two (longer if they take a commission). Ultra-orthodox Jews are exempt from military service as long as they are enrolled in seminary (Yeshiva); orthodox Jewish girls are exempt upon request; Arabs are exempt, though some volunteer.

have zero earnings at age 30-31 in our basic sample and 16.6% have zero earnings in this sample. To account for earnings data outliers I dropped from the sample all observations who are away from the mean in the sample by six or more standard deviations. Very few observations are dropped from the sample in each of the years and the results are not qualitatively affected by this sample selection procedure. To account for age differences of the different cohorts included in the sample the outcomes are adjusted for years since graduating high school. The same earnings data is also available for the parents of the students in our sample, for the years 2000-2002 and 2008-2012. I compute the average earnings of each parent and of the household for 2000-2002 and will use it as an additional control in a robustness check of the evidence presented in this paper. These data were not available for the analysis of the effect of the program on short term outcomes. I also use as additional outcomes the NII indicator of being *Eligible for Unemployment Benefit* and the annual amount of *Unemployment Benefit Compensation*.

***Personal Status Outcomes***

The Population Registry is available to us only for 2012. However, the date of each marriage and birth event are reported in the data and therefore I can adjust the demographic outcomes for years since graduating high school. These outcomes include indictors for *Marriage Status* and *for Having Children.*

**Descriptive Statistics**

Table 1, columns 1-2, presents detailed summary descriptive statistics for the outcome variables for 2012 by treatment and control group, for the pre-program cohort who graduated high school in 2000 (pre-treatment) for the NE full sample and treatment-control balancing tests (column 3). In columns 4-6, I present the respective evidence for the first three quartiles sample (3Q) of the ability distribution of students. I present results based separately on the 3Q sample because these are the students that their math and English matriculation outcomes were mostly affected by the teachers' bonuses program (Lavy 2009). The balancing tests extend the analysis presented in Lavy (2009) for the similarity between the treatment and the control group in terms of family characteristics and pre-program high school outcomes. These results are reproduced here for convenience in Table A1 in the online appendix. In this section I show that the treatment-control similarities observed in Table A1 is evident also when comparing the pre-program long-term outcomes presented in Table 1.

Post-secondary academic schooling enrollment rates statistics are presented in panel A, and post-secondary completed years of schooling statistics are presented in panel B. The ever enrolled rate in university schooling until 2012 in the treatment group for the pre-treatment cohort (2000) in the full sample is 21.6 percent, and in the control group it is 19.1 percent. The treatment-control difference is 0.026 (se-0.046), not statistically different from zero. The respective enrollment rates in academic colleges are 14.5 and 14.3, practically equal for the two groups.[7] This perfect balancing is also evident when the treatment-control differences are measured based on the 3 quartile sample (columns 4-6): the university and academic-colleges enrollment rate differences between the two groups are 0.001 (se=0.034) and 0.009 (se=0.035). Note that the means of these outcomes show clearly that in the full sample the mean of enrollment rate in university schooling is higher than in the 3 quartiles sample while the respective means in academic colleges in the two samples are identical. This is expected as the students in the fourth quartile have the highest mean high school outcomes.

Similar treatment-control similarities are observed with respect to completed years of university and academic colleges schooling. These evidence are presented in panel B: all four differences (in the natural experiment full and three quartile samples) are small and not statistically different from zero. For example, the mean years of academic college schooling in the full sample is 0.379 in the treatment sample and 0.365 in the control sample and the difference 0.014 is not statistically different from zero. In summary, of the 8 schooling related difference estimates presented in panel A and B, none suggest any pre-program control-treatment gaps.

Summary statistics and balancing tests for the pre-program cohort labor market outcomes in 2012 are presented in panel C of Table 1. The employment rate in 2012 among the treatment and control groups is 84.0 and 84.6 percent, the difference (-.006) is small and insignificant. Average annual earnings of these two groups are 67,214 New Israel Shekels (NIS), equivalent to $17,690, and 69,478 NIS ($18,300), respectively, and the difference is not statistically different from zero (-2,265, se=3945). The unemployment rate for these two groups is 8.0 percent and 8.7 percent (very similar to the national unemployment rate in 2012), and again the groups' difference is not statistically different from zero. Correspondingly there is no difference between the two groups' mean annual unemployment benefits received during 2012. The respective evidence from the 3 quartile sample that are presented in columns 4-6 show the same basic similarity in outcomes means between the treatment and control groups. None of the eight differences presented in panel C of Table 1 are statistically different from zero.

Panel D presents the means and balancing tests for marriage and fertility outcomes. About 56 percent of the individuals in the sample are married by 2012 and the treatment-control difference (two percent) is not different from zero. 45 percent of the sample have children, suggesting almost all who are married have children by 2012. However, there are no differences in these two outcomes between treatment and control group and in the mean number of children which is just below one. The evidence about the means and balancing tests based on the three quartile sample are identical to those based on the full sample.

Panel E presents the means and balancing tests for father and mother average earnings in 2000-2002, the years that students in the sample were in high school. Note that this information became available only recently through access to data at the NII lab and I therefore add them now to the treatment-control balancing analysis. The treatment control differences in these variables are positive in the full sample but they are relatively small and not significantly different from zero. In the three quartile sample the differences are even smaller and in the case of maternal earnings the differences change signs and is negative. Here as well the small differences are not statistically different from zero. The evidence on balancing of parental earnings when using the treated cohort sample are very similar without any noticeable treatment-control differences in the full or three quartile sample. These results are not presented in Table 1 which presents balancing tests for the pre-treatment cohort.

Table 2 presents the descriptive statistics and balancing tests based on the Regression Discontinuity (RD) sample. This sample is smaller than the NE sample, especially the treated group is smaller by abut 30 percent. Yet the means of the treatment and control groups are very similar to

---

[7] Note that very few students ever enroll in both university schooling and academic colleges schooling.

those based on the NE sample and the respective balancing tests show a very similar pattern to that obtained from the NE sample, namely no detectable differences between treatment and control groups and all t-tests allow rejecting the hypothesis that the differences are different from zero. However, it is worth noting that the sign of some of the differences are opposite of the signs of the NE sample differences. Of particular interest and importance is the treatment-control earnings gap in the pre-program cohort. In the NE sample (panel C in Table 1) this gap is negative though small and insignificant, while in the RD sample it is positive but here as well it is small and insignificant (panel C in Table 2). I will return to this distinction when discussing the treatment estimates of the effect on earnings which as will be shown are very similar in the NE and RD samples.

## 4. Empirical Evidence

### *Effect on Post-Secondary Academic Schooling Attainment*

The program had positive and significant short term effects on high school English and Math outcomes at the end of high school (Lavy 2009). Since the program increased the exam participation, the average score and the passing rate in the math and English matriculation exams, we should expect also positive effect on overall summary outcomes of the matriculation exams such as the matriculation rate, total number of matriculation credits, and the average score in the all matriculation exams. Evidence on improvement in these summary achievement measures should lead to an increase in post-secondary academic schooling because they are used as admission criterions to various academic institutions and study programs.

Table A3 presents results for the short-run impact of the pay for performance experiment summary measures of the matriculation exams' program, including the average matriculation score, the matriculation rate and other related end of high school outcomes. In the full sample, the average matriculation score is up by 2.8 points (se=1.017), and the matriculation rate went up by 2.6 percentage points which amounts to a 7 percent improvement, a relatively large gain. The average number of credit units increased by 0.8, the number of credits in science credits increased by 0.6 unit (a 25% percent increase), and the number of subjects studied at the most advanced level (5 credits) increased by over a tenth subject. The respective estimated effects based on the three quartile sample are similar to those based on the full sample with one exception, the effect on the matriculation rate is 5.5 percentage point increase, larger than the effect estimated in the full sample. This difference is consistent with and can be explained by the evidence presented in Lavy (2009) that indicated that the effect on the math and English tests' passing rate where also marginally larger in the three quartile sample, an improvement that allowed marginally failing students to qualify for the matriculation diploma.

We next present graphical presentation of the effect of the pay for performance program on post-secondary schooling. We focus on the two segments of the academic post-secondary education in Israel. The first includes the seven research universities in Israel that confer BA, MA and PhD degrees. These schools require a matriculation diploma for admission, an intermediate or advanced matriculation study level in English (note that to qualify for a matriculation diploma a basic study program in English is sufficient) and at least one studied matriculation subject at an advanced level; about 35% of all students are enrolled in one of the seven universities. The second type of post-secondary schooling includes over 50 academic colleges that confer mostly a BA degree and offer mostly social sciences, business and law degrees.

In figure 1, we measure the treatment effect for each year since high school graduation and trace the dynamic pattern for university enrollment for the NE sample. In order to do so, we run a separate regression for each of the outcomes and for each of the years since high school graduation. We then plot the coefficients of these regressions around a 90% confidence interval. Note that both the ever-enrolled variable and the years of schooling are cumulative variables. Hence, we expected the effects to be either flat or increasing over time.

This treatment effect becomes positive from year three after graduating high school and it reaches it is highest at 5 percent from the eight year after high school graduation, remaining flat afterwards.[8] This pattern likely reflects the fact that students who do not enroll in post-secondary schooling in the first eight years are very unlikely to return to school later in life. In contrast, the effect on years of schooling accumulates over time (Figure 1A). Although most of the increase

[8] The emergence of treatment effect after three years is reasonable given that most of the female students are in military service for the two years following high school graduation and for boys this period is three years.

happens in the first eight years, the effect seems to be increasing even after 12 years since graduation, reaching a pick at 0.25 years. The fact that the increase keeps accumulating even 12 years after high school graduation suggests that focusing on outcomes immediately after graduation might underestimate the long-term effects. Note that the effect on the intensive margin seems to operate beyond the increase in enrollment. Given a 5 percentage-points increase in enrollment and a typical duration of 3-4 years, we should expect schooling to increase by only 0.15-0.20 years. The fact that the effect on years of schooling is larger than 0.20 years suggests that the program induced treated students to stay longer and complete longer programs.

The effect size on enrollment and years of schooling can be assessed in comparison to the mean enrollment rate for the treated group which increases gradually from year one and is highest at 20 percent thirteen years later. The mean of university years of schooling in the treatment group is 0.8.[9] Figures 2 and 2A present the estimated effects on academic college enrollment and attainment and the pattern revealed in these figures is very different, as the effect is negative and small, practically close to zero.

In Figures 3-3A and 4-4A I replicated this graphical analysis based on the RD sample and the results are identical to those obtained based on the NE sample. Figures 5-5A and 6-6A present the respective graphs for the NE 3Q sample and Figures 7-7A and 8-8A do the same for the RD 3Q sample. The evidence in these figures is practically the same as those obtained based on the full sample.

Table 3 presents the point estimates and their standard errors for the impact of the pay for performance experiment on university and academic colleges schooling attainment as measured at the end of the period of study. The table presents results based on the NE sample (columns 1-4) and the RD sample (columns 5-8). Effects on enrollment are presented in columns 1-2 and 5-6. Evidence on years of schooling are presented in columns 3-4 and 7-8. The results are based on the full sample are presented in panel A.

[9] The yearly university enrollment rate is highest at year 4-7 and then starts to decline until practically leveling at close to zero at year 13 (this result is not shown in the paper for space consideration).

Enrollment in university schooling increased by 4.8 percent and this effect is precisely measured (se=0.013). This gain, relative to the pre-program mean of students in treatment schools (21.6 percent) is a 22 percent increase. This increase in enrolment led to a 0.250 increase in completed years of university schooling, reflecting a 30 percent increase relative to the baseline mean of 0.825 years of university schooling. The relative gain in university enrollment and in the respective completed years of schooling are of similar magnitude, both being large relative to the impact of some other educational interventions or policy change, for example in comparison to the gain due to an increase in compulsory schooling. The effect on academic-colleges enrollment and years of schooling are negative but close to zero and not very precisely measured. This pattern may suggest some compositional change in the academic schooling but the offsetting decline in academic college schooling are too small to be economically meaningful.

The estimates based on the RD sample are consistent with the estimates based on the NE sample. Enrollment in university education increased by 6 percentage point (se=0.014) and completed years of university schooling increased by 0.242 years (se=0.073). The relative magnitude of these gains are identical to those presented above based on the NE sample. Obtaining similar results based on two different identification strategies, randomized assignment to treatment based on a natural experiment and assignment to treatment based on a threshold of an observed criterion, lend credibility to the causal interpretation of the finding reported in panel A of Table 3. The comparability of the NE and RD evidence with respect to the treatment effect of the program on long term educational outcome at adulthood are similar to the respective comparability between the results obtained from the NE and RD samples regarding the program's effect on high school educational outcomes.

The results based on the three quartile sample (presented in panel B) are similar to the results based on the full sample. Based on the NE sample, university enrollment increased by 5.2 percent and completed years of university schooling increased by 0.24 years. The estimated effects on academic-colleges enrollment and years of schooling are also negative but somewhat smaller than those estimated based on the full sample. The results based on the RD three quartile sample are fully consistent with the respective estimates based on the NE sample.

*Effect on Employment and Earnings*

We expect that the increase in both the quality and the quantity of high school and post-secondary education to result in better labor market outcomes in adulthood. In figures 9-12, I repeat the year-by-year analysis but now focusing on labor market outcomes. The figures show the estimated effects by years since graduation from high school. The employment and earnings data are available until year 2012 so eleven years is the longest period since graduating high school for which I examine the effect of the program. Overall I find an increasing pattern in both employment and earnings. The effects become significantly different from zero after about 9-11 years after high school graduation. Perhaps reflecting the higher enrollment in post-secondary schooling, the effect on employment is initially negative and increases thereafter. The effects on earnings follow a similar pattern. As treated students spent more years on average in the schooling system and appear on average to start working later, we expect the effect on earnings to be initially negative and to increase as students complete their university schooling and accumulate labor market experience. Indeed, we find that the effects are initially negative and become significantly different from zero by the end of our sample period. In the following paragraph I describe these results in more details.

Figures 9 and 9A present the yearly estimates on employment for the NE and RD samples, respectively. As already noted, the estimates for the first three years are not meaningful because most of the students in our sample were still in military service. In the fourth year since high school graduation, about 70 percent of the individuals in the sample were employed (according to our definition of employment which is employed at least for one month during the year and had positive earnings). From year four until year eight since high school graduation, the treatment effect estimates on employment are negative. The largest effect is about -0.02 in the NE sample and -0.03 in the RD sample. The NE estimates are not precisely estimated, none of these negative estimates are different from zero. The RD estimates are more precisely measured and 2 of the four estimates are statistically different from zero. When stacking the data for these four years (from the fourth to the seventh year since high school graduation), both the NE estimates (-0.014 se=0.008) and the RD estimates (-0.026

se=0.010) are statistically different from zero.[10] From year seven after high school graduation the treatment effect on employment is positive, statistically significant in some years and marginally so in others, both in the NE and RD samples. The highest employment treatment effect estimates about 3 percent. The average employment rate from the seventh year from high school graduation is about 87 percent and this rate is stable until the end of the period. The respective evidence based on the NE and RD 3 quartiles samples are very similar to the evidence from the respective full sample.

The year by year estimated treatment effects on annual earnings are presented in Figures 11-11A. These estimates are negative from the fourth to the seventh year since graduation and then they turn positive and remain so until the end of the period studied. The lowest estimate based on the NE sample is about IS -4,000 shekels relative to mean earnings of IS 25,000 in the same year. The estimates based on the RD sample are very similar. The period with negative earnings' effect coincides with the years with negative employment effect and with the period when the treatment effect on university enrollment becomes positive and increasing. This inverse mirror image of the treatment effect on employment and on university enrollment explains the negative effect on earnings. The treatment effect on earnings turns positive and significant from year seventh on but it fluctuates in size, not surprisingly because earnings is a noisier outcome than university enrollment. The evidence based on the RD sample and the two respective samples, full and three quartiles samples, reveal similar patterns to those based on the NE full sample.

Table 4 columns 1-2 and 5-6, present the point estimates and their standard errors for the various labor market outcomes at the end point of the period we study. I also present in columns 3-4 and 7-8 estimates from stacked regressions where I pool the data of the last three years of the studied period, namely nine to eleven years after graduating high school. The stacked regression yields average treatment effect for this period and allow more precise estimation of the effect on labor market outcomes. Focusing on the stacked regression results, the teachers' incentive experiment increased employment of treated students by 1.3 percent. The respective average employment rate is 83 percent. Based on the three quartiles sample the effect is higher, about 2 percent higher

[10] However, it is likely that these estimates are biased downward because the employment measure I use does not distinguish between part time and full time employment and students usually have lower labor supply while

employment rate, and more precisely measured (se=0.010). Similar estimates are obtained based on the RD sample. The estimated effect on eligibility for unemployment benefits are negative but small and not significantly different from zero. The average unemployment rate in the treated group before treatment is low, only 6.9 percent, perhaps the reason why there is no discernable effect on this outcome. This rate is very similar to the national unemployment rate in 2010-2012 (7.1 percent) for the closest age group (25-34).

The average annual earnings in the pre-treatment cohort in 2009-2011 NE sample is NIS 55,311 ($14,555 based on an exchange rate of 3.8 Israeli Shekels to one US Dollar). The average estimated effect of the PFP program on annual earnings for this three years' period is NIS 4,711 (se=1519) ($1240). The estimated effect based on the RD sample is similar, also amounting to a 9 percent annual increase relative to the pre-treatment treatment group mean.[11] The effect on annual earnings based on the three quartile sample is lower, 7 percent in the NE sample and 6 percent in the RD sample. We note again that the similar earnings effect estimates from two different identification strategies is lending support and credibility to the causal interpretation of these estimates. Furthermore, we note that the pre-program treatment-control mean difference in earnings was small and insignificant in both samples and it was of different signs, negative in the NE sample and positive in the RD sample, yet both samples yield the same positive earnings gain.

A natural question about the above estimated effect on earnings is whether it captures the permanent long term effect. First, note that I measure the effect on earnings at about age 30-31, when individuals already completed their post-secondary schooling. Second, based on a sample of older cohorts, I find that earnings at age 30-35 is a strong predictor of earnings at an older age. Yet, it is important to note that earnings have larger variation over time than other personal outcomes. To get a better indication about the permanency of the effect on earnings, I estimated the effect on the percentile rank of individuals in the respective distribution of their cohort. There is no direct evidence that suggests that rank forecast is more stable than earnings or log earnings. However, recent papers in

in school.

[11] The fact that the treatment effect on earnings stays positive over several consecutive years is perhaps an indication that this gain reflect real productivity differences and not signaling of the higher schooling outcomes that resulted from 'teaching to the test'.

the intergenerational mobility literature provide some indirect evidence that are relevant to this issue. These studies have shown that movements across ranks in the income distribution are uncorrelated with parental income conditional on rank at age 30; in contrast, movement in log earnings are correlated with parental income conditional on log income at age 30 — in particular, rich kids have higher earnings growth so that age 30 measurements are biased predictors of later-life earnings. But, the rank forecasts appear to be less biased. For example, Nybom and Stuhler (2016) show with data from Sweden that the relationship between a child's income rank and their parental income rank stabilizes by around age 30; in contrast, the relationship in log earnings is less stable. Chetty et al (2015) find a similar pattern in the US tax data, reporting that percentile ranks predict well where children of different economic backgrounds will fall in the income distribution later in life. Using instead log earnings leads to inferior predictions because of the growth path expansions at the top of the income distribution.

Table 5 presents estimates of the effect of the program on percentile rank of earnings, where the rank is computed separately for each cohort. The estimates are fully consistent with the estimated effects on earnings that are presented in Table 4. After nine to eleven years from high school graduation, the program moves treated individuals by 2 percentile ranks (column 4, first row) and this effect is relatively precisely measured. The rest of the estimates presented in the table suggest similar findings.

***How Robust Are the Results to Controlling for Parental or Family Earnings?***

In this section I present a robustness check of the PFP program effect when I add father, mother, or family earnings as an additional control in the DID estimation of treatment effect. Each of these variables is the respective average earnings in 2000-2002, the years just before and during the program implementation. I prefer to use a three-year average earnings instead of a specific year earning because this measure is likely to be more correlated with the permanent level of family resources. These results for university schooling based on the NE sample are presented in Tables A4. The respective estimates for the effect on employment and earnings are presented in Table A5.

In columns 1-3 of Table A4, I present the treatment estimates on university enrollment when a control for father's earnings (row 1) or mother earnings (row 2) is added to the DID regressions. The three columns correspond to estimates after 10, 11 and 12 years since high school graduation. These estimates of treatment effect of the program are very similar to the respective estimates presented in Table 3. Columns 4-6 present the respective treatment estimates on years of schooling. Again these estimates are practically identical to those presented in Table 3.

In Table A5 I present the treatment estimates on earnings and employment when a controls are added for parental or family earnings. The three columns correspond to estimates after 9, 10 and 11 years since high school graduation. These estimates of treatment effect of the program are very similar to the respective estimates presented in Table 4. Columns 5-7 present the treatment estimates on employment. These estimates suggest somewhat larger effect on employment in comparison to the estimates presented in Table 4.

The obvious conclusion is that adding control for parental earnings do not affect at all the treatment estimates of the effect on PFP program on university schooling attainment. This result is not unexpected given that the parental or family earnings controls are quite balance between treatment and control, both in the pre and post-treatment cohorts (see estimates presented in panel E of Table 1).

***Treatment Heterogeneity by Family Earnings and Gender***

Next I estimate program treatment heterogeneity in university schooling, employment and earnings, by baseline family income. The possibility of different program effect by family income has policy implications with respect to the targeted versus universal implementation of teachers' incentives programs and for the external validity of our findings with respect to different socio-economic background of treated students. The sample is divided by the median of family income in 2000-2002. The estimates based on these two samples are presented in Table 6 for the NE sample. Columns 1-2 present the estimates for sub-samples by family income, panel A for the above median sample and panel B for the below median sample. The effect on university schooling is positive and significant for both groups but the effect in the high income sample is twice as large the respective effect in the lower income sample. The effect on earnings however is just about the same, marginally higher for the lower income sample. The much larger effect on employment for the low family income group is what makes the income effect shortfall due to the lower education gain of this group. The estimated increase in employment is 4.9 percent versus no employment effect at all for the higher family income sample. However, it is interesting to note that the increase in earnings for a unit gain in university schooling is the same for the high and low family income samples. The higher income sample had no gain in employment, therefore the earnings gain per a tenth of a year of schooling gain is 1492 NIS (4,447/2.98). The earnings gain due to the increase in employment for the sample of low family income is 3,445 shekels ((54,885/80.4) x 4.9)) and therefore the earnings gain per a tenth of a year increase in schooling is 1,483 NIS (1958/1.32), remarkably identical to 1,492 NIS, the respective figure for the high income sample. This means that the return to an increase in years of university schooling does not differ by family income which imply that the increase in quality of schooling, both at high school and at university, is similar for both groups.

The estimated effects by gender are presented in columns 3-4, panel C for boys and panel D for girls. The effect on schooling is positive and significant for both gender, but it is higher for girls, a gain of a third of university year of schooling versus about half of that for boys. However, boys have a larger increase in earnings, by about 2,000 shekels a year, which is almost 30 percent of the total gain. Two explanation related to employment pattern can account for this difference. First, there is a positive effect of 1.7 percent on men employment (though this effect is not precisely measured) while the respective estimate for women is zero. This effect on male employment will account for a large part of the gender difference in earnings. Secondly, the lower gain in female earnings most likely reflects a much higher propensity among women to work part-time during this period in life. I cannot provide direct evidence on this second explanation because the data I use does not include information on hours of work. However, based on data from the 2012 Israeli Labor Force Survey, I estimated that in the age group 29-34 the rate of part time employment is 0.25 among women versus 8 percent among men.

***Comparing the Effect on Earnings to Related Evidence***

No other study provided to date evidence on the effect of teachers' pay for performance on student earnings at adulthood. However, it is still useful to compare our results to effect of other childhood and schooling interventions on earnings at adulthood. Andersson et al. (2013) estimated that the effects of living in public or voucher housing on later earnings are positive, substantial, and significant for non-Hispanic Black female teenagers, but living in public or voucher housing has no effect on the later earnings of non-Hispanic Black male teenagers. The point estimates suggest that females earn 21 percent more if they ever resided in voucher housing and 18 percent more if they ever resided in public housing. The corresponding estimates when treatment is measured as number of years indicate that each additional year of voucher-supported housing participation increases earnings by 7 percent for females, while each additional year of public housing also increases female earnings by 7 percent. The overall estimated treatment effects for males suggests that each year of public housing participation as a teenager increases adult earnings by 5 percent for males. Gertler et al. (2014) report substantial effects on the earnings of participants in a randomized intervention conducted in 1986–1987 that gave psychosocial stimulation to growth-stunted Jamaican toddlers. The intervention consisted of weekly visits from community health workers over a 2-year period that taught parenting skills and encouraged mothers and children to interact in ways that develop cognitive and socioemotional skills. Twenty years later the intervention participants' earnings increased by 25 percent. Chetty et al. (2011) have shown that having a kindergarten teacher with over ten years of experience increased students' average annual earnings at ages 25 to 27 by 6.9 percent ($1,093) between 2005 and 2007. Similarly, an improvement in class quality increased average annual income earned between ages 25 and 27. Johnson et al. (forthcoming QJE) show that for children from low-income families, increasing per-pupil spending by 10 percent in all 12 school-age years increase adult hourly wages by 13 percent. Schweinhart et al. analyze the long term effect of the High/Scope Perry Preschool experiment and find that students in treatment had significantly higher median annual earnings than the no-program group, by 20 percent at age 27 and by 36 percent at age 40. Finally, Chetty, Hendren and Katz (2016) find that moving to a lower-poverty neighborhood (MTO) significantly improves college attendance rates by 2.5% and earnings by 31%, for children who were young (below age 13) when their families moved. Clearly our estimated effects on earnings are not unusually high relative to estimates surveyed above. For example, the teachers' pay experiment raised college going rate by 5%, twice that of the MTO effect, and increased earnings 10-12 years after high school graduation by 7-9%, a fourth of the MTO effect.

***Mechanisms for the Effect Earnings***

Several factors explain the 6 to 9 percent increase in earnings as a result of the teachers' pay for performance program. First to note is the contribution of the positive effect on employment to the increase in earnings. In the NE full sample, the employment gains accounts for 2 percent of the 9 percent increase and in the NE three quartile sample it accounts for 2 percent of the 7 percent increase in earnings. Similarly, in the RD three quartile sample it accounts for 1 percent of the 6 percent increase in earnings.

A second factor explaining earning growth is the increase in university years of schooling. Recent estimates of the rate of return to a year of university schooling in Israel ranges from 12 to 16 percent.[12] The lowest estimate (12%) implies that the 0.25 increase in years of university schooling contributed 3 percent to the gain in earnings. The highest estimate (16%) implies that the increase in university schooling accounts for 4% of the increase in earnings.

Another factor that accounts for part of the increase in earnings is the direct effect of the improved matriculation outcomes on earnings, independently of the effect they have on university years of schooling. Particularly important is the matriculation rate which increased by the program by 2-5 percent. The evidence suggests that having a matriculation diploma is rewarded in the labor market by a return beyond its effect on post-secondary schooling. For example, Angrist and Lavy (2009) estimate that *bagrut* holders earn 13 percent more than other individuals with exactly 12 years of schooling. Therefore, the gain the matriculation rate accounts for 0.3 percent on the earnings gain in the NE full sample and 0.65 percent of the earnings gain in the NE three quartile sample. Similarly, the quality improvements in the matriculation study program (as reflected in the composite score, number of credit units and credits in honor and science subjects) are also rewarded in the labor market beyond their effect on post-secondary schooling (Caplan et al (2009)).[13] The implied mechanism is that the range of improvements in high school educational outcomes that resulted from the pay for

[12] Frish (2009) exploit changes in compulsory schooling laws and obtain IV estimates that are much larger than the OLS Mincerian estimates. Navon (2005) estimate that the return to an MA degree (two years of schooling) is 30 percent.

[13] Caplan et al (2009) demonstrate that earnings in Israel is highly positively correlated with the quality of post-secondary schooling (colleges versus universities and higher versus lower quality universities). For example, this study shows that earnings are much higher for graduates of Tel Aviv, Jerusalem and the Technion Universities relative to graduates from the other four universities in the country. Admission to the top universities is of course positively correlated with the high school matriculation outcomes.

performance intervention allowed affected students better access to higher quality post-secondary education, mainly by facilitating enrollment in more selective programs that lead to higher return to schooling. Examples of such study programs are computer science, engineering, and law schools in the top 2-3 universities in the country. We can partially assess this channel by checking the correlations between *bagrut* outcomes that measure quality and earnings. We therefore estimate OLS regressions of annual earnings on the various high school *bagrut* outcomes while including in the regressions controls for student's parental and demographic characteristics. We use here only the sample of pre-program students in control schools. The results for the NE sample are presented in Table 7. Clearly each of the high school outcomes is highly and positively correlated with earnings at adulthood (panel A column 2). When all four high school outcomes are included jointly in the regression they still have positive coefficients but only two of them are statistically significant, the average matriculation score and the number of honor level subjects (column 3). In panel B, I present similar evidence for the correlation between two post-secondary outcomes, enrollment in academic college or university and number of completed years of schooling in any of these institutions. Clearly both of these outcomes have high correlation with earnings at adulthood (panel B column 2). When included jointly in the regression they have still a positive coefficient but only the first two are statistically significant (panel B column 3). In column 4, I present the estimates from a regression when the high school and the post-secondary schooling outcomes are included jointly in the regression. All estimates are positive but the outcomes that are significant are the average matriculation score and the enrollment indicator in university and academic colleges. In Table A6, I present similar regressions for the two sub-samples by family income. The estimates in these two sub-samples are qualitatively similar.

***Effect on marriage and children***

I next examine teacher pay for performance treatment impacts on children's marriage and fertility outcomes in Table 8. These outcomes are defined in terms of 11 years since graduating high school because the data for these variables are available to us only for 2012. Therefore, the two outcomes we examine are an indicator of being married 11 years after high school graduation and an indicator of having children 11 years after high school graduation. About 58 percent of the pre-treatment treated schools sample are married by 2011. The treatment effect on marriage rate based on the full NE sample is negative but not significantly different from zero. This is evident in the full sample (NE sample) and in sub-samples by family income and by gender. Similarly, the estimated effect on having children, based on the full sample, is negative but very small and not statistically different from zero. In panel B I report results based on sub-samples by family income. For the sub-sample of low family income, the estimates on marriage and children are negative and the first effect is large and significant, a decline of 3.3 percent in the marriage rate and decline of 1.7 percent in the probability of have children 11 years after high school graduation, but the latter effect is not precisely measured. The two respective estimates for the high family income sample are small and not statistically significant. In panel C, I report the results by gender and clearly the treatment effect on the two demographic outcomes are small and not distinguishable from zero. The implied conclusion from these evidence is that the teachers' pay for performance program lowered marriage rate, probably by delaying it to a later age, among students from low income families but not for others.

**5. Conclusions**

In this paper I study the long term effect of an experiment that paid teachers additional income based on their students' performance in high stakes exams at the end of high school. All studies of teachers' incentive programs and of the vast majority of published research on the impact of other school interventions has examined their effects on short-run outcomes, primarily by looking at their impact on standardized test scores. This study is the first study to use a long horizon follow up of students from high school to adulthood at age 30 to examine impact of a teachers pay for performance scheme on long-run life outcomes. Such an analysis can address the critical question of whether a public educational intervention can achieve the ultimate goal of improving lifetime well-being. This research makes also an important contribution to the recently growing literature on the long-term effects of education quality by providing evidence on an intervention that changes a specific input that can improve the attainment and quality of students. Focusing on an intervention that can be expanded or implemented elsewhere, such as teachers' pay for performance, provides explicit guidance for educational policy making. This research direction is a natural follow up to recent studies that estimated positive effect of teaching quality using teachers' fixed effect and value added models. However, explicit policy prescriptions of how to improve teaching quality do not follow immediately from these important evidence and the results presented in this paper helps in this regard by unraveling 'wires' in the 'black box' of teachers' quality.

The evidence presented above show that over a decade after the initial intervention, treated individuals experience sizable gains in schooling attainment and quality, large increase in annual earnings, some of which reflect a return to education quality beyond the return to years of schooling. These gains are very small relative to the cost of the program. The average cost of the program was $170 per student versus more than $1,000 annual earnings gain starting at about age 30. Of course a complete cost-benefit analysis should also take into account the forgone earnings while in university schooling and the respective tuition fees. However, given that individuals will benefit for many years from the increase in earnings, the present value of of benefits surely outweigh the cost, suggesting a high private rate of return. A social rate of return analysis of this project should take into acount the cost of university schooling not recovered by tution fees and the benefit for additional taxes collected from the higher earnings. Clearly, these adjustments will still yield a high social internal rate of return on this project.

The evidence presented in this paper are relevant and important for education policy in developed and developing countris as merit pay and incentive based pay for teachers is being implemented or contemplated in many countries. In U.S. education policy, for example, teachers' merit pay reforms has recently re-emerged at the top of the education policy agenda. In his first major education policy speech, President Obama promoted merit pay for teachers, and created in 2009 the $4.4 billion in federal funds Race to the Top fund to encourage States to implement performance pay system for teachers.[14] Another example is the UK reform where from September 2014, teachers in England have their annual salary rises tied to performance, replacing a system where almost all teachers automatically moved up a point on the pay scale every year. The move has been hugely controversial. For example, on March 26 2014 the National Union of Teachers struck in protest of the

[14] Merit-Based Pay For Teachers | eduflow: https://eduflow.wordpress.com/2013/10/08/merit-based-pay-for-teachers.

overhall in pay structures that is due to begin later in the year.[15] However, it should be noted that the evidence presented in this paper relates to long-term effects of a pay-for-performance scheme that targeted improvements in the high-stake exams that play a key role in determing university and college admission and hence, it is conceivable that in this particular setting, the pay-for-performace programme produced long-term results. This evidence is of course relevant for the many countries who also use such high stake high school exams for the same purpose[16], but perhaps less relevant in other settings. For example, if a similar program were introduced in primary or middle school, gain in test scores in this case may not have necessairily similar long-term effects. Another point to note is that a pay for performance program when implemented at scale, for example nation-wide, will have general equibiliruim implications and the scope of university enrollment expansion estimated above will be feasible only if the increase in demand generated by the program can be met by the current supply of unviersity schooling as was the case in the context of this study.

[15] The Economist, March 29 2014.

[16] Examples of high school exit exams that play a similar role for university admission include the A level exams in the UK, the Matura exam in various countries, including Albania, Austria, Bosnia and Herzegovina, Bulgaria, Croatia, the Czech Republic, Hungary, Italy, Poland, Serbia, Slovakia, Slovenia, Switzerland, Ukraine, the Gaokao National Matriculation Examination in China, the General Exiting Exam (EGEL) in Mexico.

## 6. References


Abramitzky R., and V. Lavy. 2014. “How Responsive is Investment in Schooling to Changes in Returns? Evidence from an Unusual Pay Reform in Israel’s Kibbutzim”, *Econometrica*, Vol. 82, No. 4 (July), 1241–1272.

Andersson, Fredrik, John Haltiwanger, Mark Kutzbach, Giordano Palloni, Henry Pollakowki, and Daniel Weinberg. “Childhood Housing and Adult Earnings: A Between-Siblings Analysis of Housing Vouchers and Public Housing”. Discussion Papers, CES 13-48, U.S. Census Bureau, Center for Economic Studies September 2013.

Angrist Josh and Victor Lavy, 1999, "Using Maimonides' Rule to Estimate the Effect of Class Size on Children's Academic Achievement." *Quarterly Journal of Economics*, Vol. 114 No. 2 (May), 533-575.

Anderson, Michael L. 2008. “Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects," *Journal of the American Statistical Association*, 103 (484), 1481-1495.

Angrist, J. D. and A.B. Krueger. 1991. “Does Compulsory School Attendance Affect Schooling and Earnings?," *Quarterly Journal of Economics*, 106 (4), 979-1014.

Black, Dan, Kermit Daniel, and Seth Sanders. 2002. "The Impact of Economic Conditions on Participation in Disability Programs: Evidence from the Coal Boom and Bust." *American Economic Review*, 92(1): 27-50.

Caplan Tom, Orly Furman, Dmitri Romanov, Noam Zussman. 2009. “The Quality of Israeli Academic Institutions: What the Wages of Graduates Tell About It?” Central Bureu of Statistics, Israel , Working Paper Series NO. 42, May.

Card, David and Krueger, Alan B. 1991. “Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States.” *Journal of Political Economy, pp:*1-40.

Chetty, R., J. Friedman, N. Hilger, E. Saez, D. Whithmore Schanzenbach, and D. Yagan. 2011. “How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project Star," *Quarterly Journal of Economics* 126(4): 1593-1660.

Chetty, Raj, John N. Friedman, and Jonah Rockoff. 2013. “Measuring the Impact of Teachers I: Evaluating Bias in Teacher Value-Added Estimates” *American Economic Review*, June.

Chetty, Raj, John N. Friedman, and Jonah Rockoff. 2013. “Measuring the Impact of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood” *American Economic Review*, June.

Chetty, Raj, Nathaniel Hndren and Lawrence Katz “The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment”, *American Economic Review* 106(4): 855-902, 2016.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez, "Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States" ***Quarterly Journal of Economics*** 129(4): 1553-1623, 2014.

Deming, David. 2009. "Early Childhood Intervention and Life-Cycle Skill Development: Evidence from Head Start," *American Economic Journal: Applied Economics*, 1 (3), 111-134.

Deming David, S. Cohodes, J. Jennings, and C. Jencks. 2013. "School Accountability, Postsecondary Attainment and Earnings." NBER Working Paper. w19444.

Dustmann C., P. Puhani and U. Schonberg. 2012. "The Long-Term Effects of School Quality on Labor Market Outcomes and Educational Attainment", Draft, UCL department of economics, January.

Dynarski, S., J. Hyman, and D. Whitmore Schanzenbach. 2013. "Experimental Evidence on the Effect of Childhood Investments on Postsecondary Attainment and Degree Completion" *Journal of Policy Analysis and Management,* 32(4).

Duflo E., R. Hanna, and S. Ryan. 2012. "Incentives Works: Getting Teachers to Come to School" *American Economic Review*, pp: 1241-78.

Frish Roni, 2009, "The Economic Returns to Schooling In Israel" Israel Economic Review Vol. 7 No. 1, 113–141.

Fryer, R. G. 2011. "Teacher Incentives and Student Achievement: Evidence from New York City Public Schools," National Bureau of Economic Research Working Paper 16850.

Garces, E., D. Thomas, and J. Currie. 2002. "Longer-Term Effects of Head Start," *American Economic Review*, pp: 999-1012.

Gertler, Paul J., James Heckman, Rodrigo Pinto, Arianna Zanolini, Christel Vermeersch, Susan Walker, Susan M. Chang, and Sally Grantham-McGregor. "Labor Market Returns to an Early Childhood Stimulation Intervention in Jamaica" Science, 344(6187): 998–1001, May 2014.

Glewwe, Paul, N. Ilias and M. Kremer. 2010. "Teacher Incentives*," American Economic Journal: Applied Economics,* pp: 205-27.

Gould E., V. Lavy and D. Paserman. 2011. "Fifty-five Years after the Magic Carpet Ride: The Long-Run Effect of the Early Childhood Environment on Social and Economic Outcome", *Review of Economic Studies*, July: 77, 1164–1191.

Jacob, B. A., and S. D. Levitt. 2003.: "Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating," *Quarterly Journal of Economics* 118, 843-77.

Johnson, Rucker C., C. Kirabo Jackson and Claudia Persico "The Effects of School Spending on Educational and Economic Outcomes: Evidence from School Finance Reforms," *The Quarterly Journal of Economics* (forthcoming).

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*, CXI: 1–28.

Lazear, E. “Performance Pay and Productivity,” *American Economic Review*, December, 2000.

Lazear, E. 2001. “Paying Teachers for Performance: Incentives and Selection,” Draft, August.

Lavy, V. 2002. “Evaluating the Effect of Teachers’ Group Performance Incentives on Students Achievements.” *Journal of Political Economy*, 10 (6), December: 1286–1318.

Lavy, V. 2007. “Using Performance-Based Pay to Improve the Quality of Teachers”, *The Future of Children*, 87-110.

Lavy, V. 2009. “Performance Pay and Teachers’ Effort, Productivity and Grading Ethics”, *American Economic Review*, Vol. 99, No. 5, December: 1979-2011.

Lleras-Muney, Adriana. 2005. “The Relationship between Education and Adult Mortality in the United States," *Review of Economic Studies*, 72, 189-221.

Ludwig, Jens and Douglas L. Miller. 2007. “Does Head Start Improve Children's Life Chances? Evidence from a Regression Discontinuity Design," *The Quarterly Journal of Economics*, 122 (1), 159-208.

Ludwig, Jens, Greg J. Duncan, Lisa A. Gennetian, Lawrence F. Katz, Ronald C. Kessler, Jeffrey R. Kling and Lisa Sanbonmatsu. 2013. “Long-Term Neighborhood Effects On Low-Income Families: Evidence From Moving To Opportunity”, NBER Working Paper 18772.

Muralidharan Karthik and Venkatesh Sundararaman. 2011. "Teacher Performance Pay: Experimental Evidence from India", Journal of Political Economy, Vol. 119, No. 1, pp. 39-77.

Navon, Guy, 2006. "Human Capital Heterogeneity: University Choice and Wages," MPRA Paper 9708, University Library of Munich, Germany.

Neal, D. 2011. "The Design of Performance Pay in Education," Handbook of the Economics of Education, Vol. 4, pp. 495-550.

Nybom Martin and Jan Stuhler. “Biases in Standard Measures of Intergenerational Income Dependence”. Draft, April 1, 2016.

Schweinhart, L., J. Montie, Z. Xiang, W.S. Barnett, C.R. Belfeld, and Milagros Nores. 2005. Lifetime effects: The High/Scope Perry Preschool study through age 40, Ypsilanti: High/Scope Press.

**Table 1: Post-Secondary Schooling, Employment and Income, and Demographic Statistics in 2012: The 2000 Cohort: The Natural Experiment Sample**

| | 4 Quartiles Sample | | | 3 Quartiles Sample | | |
|---|---|---|---|---|---|---|
| | Treated schools | Non treated Schools | Difference | Treated schools | Non treated Schools | Difference |
| Dependent variable | (1) | (2) | (3) | (4) | (5) | (6) |
| **A. Enrollment in Post High School Education** | | | | | | |
| University | 0.216 | 0.191 | 0.026 | 0.119 | 0.118 | 0.001 |
| | | | (0.046) | | | (0.034) |
| Academic College | 0.145 | 0.143 | 0.001 | 0.142 | 0.133 | 0.009 |
| | | | (0.029) | | | (0.035) |
| **B. Post High School Years of Schooling** | | | | | | |
| University | 0.825 | 0.716 | 0.109 | 0.343 | 0.380 | -0.037 |
| | | | (0.211) | | | (0.122) |
| Academic College | 0.379 | 0.365 | 0.014 | 0.372 | 0.328 | 0.044 |
| | | | (0.080) | | | (0.093) |
| **C. Employment Outcomes in 2012** | | | | | | |
| Employed (1 = Yes, 0 = No) | 0.840 | 0.846 | -0.006 | 0.837 | 0.860 | -0.023 |
| | | | (0.015) | | | (0.016) |
| Average Annual Earnings (NIS) | 67,214 | 69,478 | -2,265 | 63,456 | 66,565 | -3,108 |
| | | | (3,945) | | | (4,657) |
| Received Unemployment Insurance Benefits (1 = Yes, 0 = No) | 0.080 | 0.087 | -0.007 | 0.083 | 0.092 | -0.009 |
| | | | (0.011) | | | (0.014) |
| Total Unemployment Insurance Benefits Received (NIS) | 911 | 941 | -30 | 870 | 937 | -67 |
| | | | (138) | | | (158) |
| **D. Demographic Outcomes** | | | | | | |
| Married | 0.563 | 0.542 | 0.020 | 0.557 | 0.549 | 0.008 |
| | | | (0.042) | | | (0.037) |
| Children (1 = Yes, 0 = No) | 0.451 | 0.454 | -0.003 | 0.455 | 0.468 | -0.013 |
| | | | (0.051) | | | (0.046) |
| Number of Children | 0.836 | 0.792 | 0.044 | 0.845 | 0.834 | 0.011 |
| | | | (0.127) | | | (0.125) |
| Age at Marriage | 24.338 | 24.591 | -0.253 | 24.314 | 24.467 | -0.153 |
| | | | (0.379) | | | (0.383) |
| Age at First Birth | 25.302 | 25.467 | -0.165 | 25.236 | 25.269 | -0.033 |
| | | | (0.357) | | | (0.353) |
| **E. Parental Outcomes** | | | | | | |
| Average of Father Earnings in 2000-2002 | 102,212 | 96,212 | 6,001 | 86,975 | 85,776 | 1,199 |
| | | | (16,693) | | | (12,666) |
| Average of Mother Earnings in 2000-2002 | 47,715 | 45,484 | 2,231 | 39,336 | 41,351 | -2,014 |
| | | | (7,798) | | | (6,096) |
| Number of Observations | 2,424 | 2,703 | 5,127 | 1,704 | 2,020 | 3,724 |
| Weighted Number of Observations | 3,980 | 4,171 | 8,151 | 3,058 | 3,087 | 6,145 |

Notes: The table reports means and standard deviations for different post-secondary schooling, employment, income, and dempographic variables for the natural experiment sample described in the paper. Columns 1-3 report results for all four quartiles, and columns 4-6 report results for students who are in the lowest three quartiles of test grades. Panel A is comprised of binary variables indicating whether the individual has been enrolled or not to a specific type of post-secondary institution by 2012. The categories are not mutually exclusive and overlapping is possible. Panel B reports the number of years of education an individual has attained by 2012 in each type of the post-secondary institutions described in panel A. Panel C reports different employment and income variables for the individual in the year 2012. Panel D reports different demographic variables for the year 2012 in addition to the age at marriage and the age at first birth. Panel E reports different parental variables. Standard errors in parenthesis are adjusted for school level clustering. Number of observations does not apply to the the age at marriage and the age at first birth variables, as these are computed on a sub-sample of individuals that are married/have children.

**Table 2: Post-Secondary Schooling, Employment and Income, and Demographic Statistics in 2012: The 2000 Cohort: The Regression Discontinuity Sample**

| | 4 Quartiles Sample | | | 3 Quartiles Sample | | |
|---|---|---|---|---|---|---|
| | Treated schools | Non treated Schools | Difference | Treated schools | Non treated Schools | Difference |
| Dependent variable | (1) | (2) | (3) | (4) | (5) | (6) |
| **A. Enrollment in Post High School Education** | | | | | | |
| University | 0.209 | 0.148 | 0.061 | 0.116 | 0.084 | 0.032 |
| | | | (0.039) | | | (0.023) |
| Academic College | 0.160 | 0.127 | 0.033 | 0.162 | 0.112 | 0.050 |
| | | | (0.032) | | | (0.035) |
| **B. Post High School Years of Schooling** | | | | | | |
| University | 0.793 | 0.523 | 0.270 | 0.333 | 0.249 | 0.084 |
| | (0.155) | (0.100) | (0.184) | | | (0.080) |
| Academic College | 0.423 | 0.317 | 0.106 | 0.422 | 0.274 | 0.148 |
| | | | (0.084) | | | (0.093) |
| **C. Employment Outcomes in 2012** | | | | | | |
| Employed (1 = Yes, 0 = No) | 0.843 | 0.844 | -0.001 | 0.842 | 0.858 | -0.016 |
| | | | (0.017) | | | (0.019) |
| Average Annual Earnings (NIS) | 70,091 | 67,389 | 2,701 | 66,275 | 64,249 | 2,026 |
| | | | (3,841) | | | (4,744) |
| Received Unemployment Insurance Benefits (1 = Yes, 0 = No) | 0.080 | 0.091 | -0.011 | 0.086 | 0.100 | -0.014 |
| | | | (0.013) | | | (0.015) |
| Total Unemployment Insurance Benefits Received (NIS) | 890 | 967 | -77 | 889 | 1,026 | -137 |
| | | | (165) | | | (178) |
| **D. Demographic Outcomes** | | | | | | |
| Married | 0.548 | 0.558 | -0.010 | 0.551 | 0.556 | -0.005 |
| | | | (0.050) | | | (0.047) |
| Children (1 = Yes, 0 = No) | 0.427 | 0.487 | -0.060 | 0.439 | 0.493 | -0.054 |
| | | | (0.053) | | | (0.053) |
| Number of Children | 0.760 | 0.830 | -0.070 | 0.789 | 0.864 | -0.075 |
| | | | (0.138) | | | (0.150) |
| Age at Marriage | 24.586 | 24.495 | 0.090 | 24.521 | 24.349 | 0.172 |
| | | | (0.409) | | | (0.415) |
| Age at First Birth | 25.523 | 25.405 | 0.117 | 25.451 | 25.165 | 0.287 |
| | | | (0.415) | | | (0.402) |
| **E. Parental Outcomes** | | | | | | |
| Average of Father Earnings in 2000-2002 | 103,816 | 81,924 | 21,892 | 91,632 | 75,312 | 16,321 |
| | | | (16,668) | | | (13,093) |
| Average of Mother Earnings in 2000-2002 | 49,082 | 39,383 | 9,699 | 43,912 | 36,496 | 7,416 |
| | | | (6,945) | | | (5,935) |
| Number of Observations | 1,697 | 2,471 | 4,168 | 1,257 | 1,844 | 3,101 |
| Weighted Number of Observations | 2,843 | 3,064 | 5,907 | 2,281 | 2,246 | 4,527 |

Notes: The table reports means and standard deviations for different post-secondary schooling, employment, income, and dempographic variables for the natural experiment sample described in the paper. Columns 1-3 report results for all four quartiles, and columns 4-6 report results for students who are in the lowest three quartiles of test grades. Panel A is comprised of binary variables indicating whether the individual has been enrolled or not to a specific type of post-secondary institution by 2012. The categories are not mutually exclusive and overlapping is possible. Panel B reports the number of years of education an individual has attained by 2012 in each type of the post-secondary institutions described in panel A. Panel C reports different employment and income variables for the individual in the year 2012. Panel D reports different demographic variables for the year 2012 in addition to the age at marriage and the age at first birth. Panel E reports different parental variables. Standard errors in parenthesis are adjusted for school level clustering. Number of observations does not apply to the the age at marriage and the age at first birth variables, as these are computed on a sub-sample of individuals that are married/have children.

**Table 3: Differences-in-Differences Estimates of the Effect of Teachers' Bonuses Program on Post-Secondary Schooling**

| | **The Natural Experiment Sample** | | | | **The Regression Discontinuity Sample Sample** | | | |
|---|---|---|---|---|---|---|---|---|
| | **Enrollment in Post-Secondary Schooling** | | **Post-Secondary Years of Schooling** | | **Enrollment in Post-Secondary Schooling** | | **Post-Secondary Years of Schooling** | |
| | **12 Years After High-School Graduation Outcomes** | | **12 Years After High-School Graduation Outcomes** | | **12 Years After High-School Graduation Outcomes** | | **12 Years After High-School Graduation Outcomes** | |
| | 2000 Cohort in Treated Schools | Estimate | 2000 Cohort in Treated Schools | Estimate | 2000 Cohort in Treated Schools | Estimate | 2000 Cohort in Treated Schools | Estimate |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| **A. 4 Quartiles Sample** | | | | | | | | |
| University | 0.216 | 0.048 | 0.825 | 0.250 | 0.209 | 0.060 | 0.793 | 0.242 |
| | (0.412) | (0.013) | (1.877) | (0.067) | (0.407) | (0.014) | (1.856) | (0.073) |
| Academic College | 0.145 | -0.026 | 0.379 | -0.072 | 0.160 | -0.017 | 0.423 | -0.047 |
| | (0.352) | (0.014) | (1.051) | (0.037) | (0.367) | (0.018) | (1.107) | (0.046) |
| Number of Observations | 2,703 | 10,077 | 2,703 | 10,077 | 2,471 | 8,230 | 2,471 | 8,230 |
| Weighted Number of Observations | 4,171 | 15,903 | 4,171 | 15,903 | 3,064 | 11,561 | 3,064 | 11,561 |
| **B. 3 Quartiles Sample** | | | | | | | | |
| University | 0.119 | 0.052 | 0.343 | 0.239 | 0.116 | 0.061 | 0.333 | 0.226 |
| | (0.324) | (0.012) | (1.115) | (0.043) | (0.320) | (0.014) | (1.130) | (0.053) |
| Academic College | 0.142 | -0.011 | 0.372 | -0.028 | 0.162 | -0.014 | 0.422 | -0.019 |
| | (0.349) | (0.013) | (1.046) | (0.030) | (0.369) | (0.017) | (1.106) | (0.036) |
| Number of Observations | 2,020 | 7,382 | 2,020 | 7,382 | 1,844 | 6,161 | 1,844 | 6,161 |
| Weighted Number of Observations | 7,382 | 11,952 | 7,382 | 11,952 | 2,246 | 8,801 | 2,246 | 8,801 |

*Notes* : This table presents the differences-in-differences estimates of the effect of the School Choice program on post-secondary schooling 12 years after high-school graduatoin. Panel A and Panel B report the results for the three quartile and four quartile samples described in the paper. Columns 1-4 report the results for the natural experiment sample described in the paper, and columns 5-8 report the results for the regression discontinuity sample described in the paper. Columns 1,3,5, and 7 represent the mean and standard error for the 2000 cohort in the treated schools. These cohorts did not receive the treatment so it is useful to compare their averages as a benchmark for the treatment effect. Columns 2,4,6, and 8 report the Differences-in-Differences estimate for each of the dependent variables. Standard errors are clustered at the school year level.

**Table 4: Differences-in-Differences Estimates of the Effect of The Teachers' Bonuses Program on Employment and Income**

| | The Natural Experiment Sample | | | | The Regression Discontinuity Sample | | | |
|---|---|---|---|---|---|---|---|---|
| | 11 Years After High-School Graduation Outcomes | | 9-11 Years After High-School Graduation Outcomes Stacked Outcomes | | 11 Years After High-School Graduation Outcomes | | 9-11 Years After High-School Graduation Outcomes Stacked Outcomes | |
| | 2000 Cohort in Treated Schools | Estimate | 2000 Cohort in Treated Schools | Estimate | 2000 Cohort in Treated Schools | Estimate | 2000 Cohort in Treated Schools | Estimate |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| **A. 4 Quartiles Sample** | | | | | | | | |
| Employment Indicator (1 = Yes, 0 = No) | 0.841 | 0.010 | 0.830 | 0.013 | 0.842 | 0.008 | 0.832 | 0.012 |
| | (0.366) | (0.009) | (0.376) | (0.008) | (0.365) | (0.007) | (0.373) | (0.007) |
| Total Annual Earnings (NIS) | 64,993 | 5,851 | 55,311 | 4,714 | 66,903 | 6,731 | 56,655 | 4,860 |
| | (56,317) | (1,993) | (49,396) | (1,519) | (57,493) | (2,481) | (50,032) | (1,857) |
| Received Unemployment Insurance Benefits Indicator (1 = Yes, 0 = No) | 0.068 | 0.000 | 0.070 | -0.002 | 0.069 | -0.004 | 0.071 | -0.005 |
| | (0.252) | (0.011) | (0.255) | (0.006) | (0.254) | (0.014) | (0.257) | (0.008) |
| Total Unemployment Insurance Benefits Received (NIS) | 693 | 37 | 597 | 28 | 725 | -112 | 602 | -14 |
| | (3,076) | (114) | (2,699) | (60) | (3,193) | (134) | (2,694) | (82) |
| Number of Observations | 2,703 | 10,077 | 8,109 | 30,231 | 2,471 | 8,230 | 7,413 | 24,690 |
| Weighted Number of Observations | 4,171 | 15,903 | 12,513 | 47,709 | 3,064 | 11,561 | 9,192 | 34,683 |
| **B. 3 Quartiles Sample** | | | | | | | | |
| Employment Indicator (1 = Yes, 0 = No) | 0.827 | 0.024 | 0.819 | 0.020 | 0.831 | 0.019 | 0.825 | 0.014 |
| | (0.378) | (0.013) | (0.385) | (0.010) | (0.375) | (0.012) | (0.380) | (0.010) |
| Total Annual Earnings (NIS) | 61,919 | 4,982 | 53,438 | 3,869 | 64,158 | 5,190 | 55,187 | 3,567 |
| | (52,965) | (1,513) | (46,647) | (1,435) | (54,107) | (1,457) | (47,487) | (1,524) |
| Received Unemployment Insurance Benefits Indicator (1 = Yes, 0 = No) | 0.068 | 0.003 | 0.073 | -0.002 | 0.057 | -0.003 | 0.057 | -0.007 |
| | (0.252) | (0.012) | (0.260) | (0.008) | (0.231) | (0.015) | (0.231) | (0.010) |
| Total Unemployment Insurance Benefits Received (NIS) | 700 | 114 | 619 | 37 | 520 | -73 | 429 | -18 |
| | (3,137) | (133) | (2,750) | (67) | (2,570) | (147) | (2,133) | (90) |
| Number of Observations | 2,020 | 7,382 | 6,060 | 22,146 | 1,844 | 6,161 | 5,532 | 18,483 |
| Weighted Number of Observations | 3,087 | 11,952 | 9,261 | 35,856 | 2,246 | 8,801 | 6,738 | 26,403 |

*Notes* : This table presents the differences-in-differences estimates of the effect of theTeachers' Bonuses program on different employment and income outcomes. Panel A and Panel B report the results for the three quartile and four quartile samples described in the paper. Columns 1-4 report the results for the natural experiment sample described in the paper, and columns 5-8 report the results for the regression discontinuity sample described in the paper. Columns 1-2 and 5-6 report results for 11 years after high-school graduation, and columns 3-4 and 7-8 report results for the stacked outcomes of 9-11years after high-school graduation. The variable "Employment Indicator" receives the value of 1 is the individual has any work record for the given year and 0 otherwise. The variable "Received Unemployment Insurance Benefits Indicator" Receives the value of 1 if the individual has any record indicating that he received any amount of unemployment benefits in the given year, and 0 otherwise. The variable "Total Unemployment Insurance Benefits Received" describes the total NIS amount of unemployment benefits the individual received in the given year. Average Annual Earnings measure the total NIS amount of earnings the individual received in the given year. Columns 1,3, 5, and 7 report the mean and standard error for the 2000 cohort in the treated schools. This cohort did not receive the treatment so it is useful to compare their averages as a benchmark for the treatment effect. Columns 2,4, 6, and 8 report the Differences-in-Differences estimate for each of the dependent variables listed above. Standard errors are clustered at the school year level.

**Table 5: Differences-in-Differences Estimates of the Effect of The Teachers' Bonuses Program on Percentile Ranking of Income**

| | The Natural Experiment Sample | | | | The Regression Discontinuity Sample | | | |
|---|---|---|---|---|---|---|---|---|
| | 11 Years After High-School Graduation Outcomes | | 9-11 Years After High-School Graduation Outcomes Stacked Outcomes | | 11 Years After High-School Graduation Outcomes | | 9-11 Years After High-School Graduation Outcomes Stacked Outcomes | |
| | 2000 Cohort in Treated Schools | Estimate | 2000 Cohort in Treated Schools | Estimate | 2000 Cohort in Treated Schools | Estimate | 2000 Cohort in Treated Schools | Estimate |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| **A. 4 Quartiles Sample** | | | | | | | | |
| Total Annual Earnings (NIS) | 48.777 | 2.638 | 48.3 | 2.041 | 49.814 | 3.018 | 49.0 | 1.972 |
| | (30.337) | (0.863) | (30.3) | (0.794) | (30.547) | (1.019) | (30.5) | (0.899) |
| Number of Observations | 2,703 | 10,077 | 8,109 | 30,231 | 2,471 | 8,230 | 7,413 | 24,690 |
| Weighted Number of Observations | 4,171 | 15,903 | 12,513 | 47,709 | 3,064 | 11,561 | 9,192 | 34,683 |
| **B. 3 Quartiles Sample** | | | | | | | | |
| Total Annual Earnings (NIS) | 47.164 | 2.181 | 47.5 | 1.533 | 48.174 | 2.334 | 48.5 | 1.256 |
| | (29.975) | (0.836) | (30.3) | (0.892) | (30.113) | (0.832) | (30.3) | (0.938) |
| Number of Observations | 2,020 | 7,382 | 6,060 | 22,146 | 1,844 | 6,161 | 5,532 | 18,483 |
| Weighted Number of Observations | 3,087 | 11,952 | 9,261 | 35,856 | 2,246 | 8,801 | 6,738 | 26,403 |

*Notes* : This table presents the differences-in-differences estimates of the effect of theTeachers' Bonuses program on percentile ranking income outcomes. Percentile ranking of income are assigned within each cohort and are lagged to be age-adjusted. Panel A and Panel B report the results for the three quartile and four quartile samples described in the paper. Columns 1-4 report the results for the natural experiment sample described in the paper, and columns 5-8 report the results for the regression discontinuity sample described in the paper. Columns 1-2 and 5-6 report results for 11 years after high-school graduation, and columns 3-4 and 7-8 report results for the stacked outcomes of 9-11years after high-school graduation. Average Annual Earnings measure the total NIS amount of earnings the individual received in the given year. Columns 1,3, 5, and 7 report the mean and standard error for the 2000 cohort in the treated schools. This cohort did not receive the treatment so it is useful to compare their averages as a benchmark for the treatment effect. Columns 2,4, 6, and 8 report the Differences-in-Differences estimate for each of the dependent variables listed above. Standard errors are clustered at the school year level.

**Table 6: Differences-in-Differences Estimates of the Effect of The Teachers' Bonuses Program by Family Income and Gender, 11 Years After High-School Graduation - The Natural Experiment Sample**

| | 2000 Cohort in Treated Schools | Estimate | 2000 Cohort in Treated Schools | Estimate |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| | **A. High Family Income** | | **C. Boys** | |
| University Enrollment | 0.246 | 0.078 | 0.176 | 0.028 |
| | (0.431) | (0.027) | (0.381) | (0.015) |
| University Years of Schooling | 0.953 | 0.298 | 0.647 | 0.161 |
| | (1.969) | (0.103) | (1.663) | (0.060) |
| Employment Indicator (1 = Yes, 0 = No) | 0.868 | 0.002 | 0.860 | 0.017 |
| | (0.339) | (0.012) | (0.348) | (0.015) |
| Total Annual Earnings (NIS) | 69,678 | 4,477 | 73,557 | 6,778 |
| | (57,684) | (1,930) | (59,466) | (3,416) |
| Number of Observations | 1,481 | 5,891 | 1,353 | 4,953 |
| | **B. Low Family Income** | | **D. Girls** | |
| University Enrollment | 0.127 | 0.027 | 0.209 | 0.066 |
| | (0.333) | (0.014) | (0.407) | (0.028) |
| University Years of Schooling | 0.451 | 0.132 | 0.804 | 0.333 |
| | (1.423) | (0.041) | (1.852) | (0.117) |
| Employment Indicator (1 = Yes, 0 = No) | 0.804 | 0.049 | 0.819 | 0.004 |
| | (0.397) | (0.015) | (0.386) | (0.013) |
| Total Annual Earnings (NIS) | 54,885 | 5,403 | 52,400 | 4,681 |
| | (50,950) | (1,654) | (48,378) | (2,226) |
| Number of Observations | 1,222 | 4,186 | 1,350 | 5,124 |

*Notes*: This table presents the differences-in-differences estimates of the effect of theTeachers' Bonuses program on employment, income, university enrollment and university years of schooling by family income and by gender, for the natural experiment sample described in the paper. Columns 1-2 report the results for the high family income sample and the low family income sample. Individuals included in the high family income sample are those who come from families in which the average household income in 2000-2002 is higher than the mean average household income in the same years. Columns 3-4 report the results for boys and for girls. The variable "Employment Indicator" receives the value of 1 is the individual has any work record for the given year and 0 otherwise. Columns 1 and 3 report the mean and standard error for the 2000 cohort in the treated schools. This cohort did not receive the treatment so it is useful to compare their averages as a benchmark for the treatment effect. Columns 2 and 4 report the Differences-in-Differences estimate for each of the dependent variables listed above. Standard errors are clustered at the school year level.

**Table 7: OLS Relationships between High School Matriculation Outcomes, College Schooling, and Earnings at Adulthood: : The Natural Experiment Sample - Treated Schools**

| | | Annual Earnings - 11 Years After High-School Graduation | | |
|---|---|---|---|---|
| | 2000 Cohort in Treated Schools | Separate Estimate | Joint Estimate Panel A\B | Joint Estimate Panels A + B |
| | (1) | (2) | (3) | (4) |
| **A. High School Matriculation Outcomes** | | | | |
| Average Matriculation Score | 73 | 321 | 221 | 205 |
| | (23) | (62) | (65) | (67) |
| Received High School Matriculation (1 = Yes, 0 = No) | 0.538 | 9,099 | 4,360 | 3,787 |
| | (0.499) | (2,352) | (2,628) | (2,611) |
| Number of Credit Units in Matriculation Exams | 22 | 600 | 41 | 86 |
| | (10) | (131) | (159) | (155) |
| Number of Honor Level Subjects | 2.517 | 3,579 | 1,744 | 1,040 |
| | (1.820) | (916) | (1,211) | (1,222) |
| **B. Post Secondary Schooling** | | | | |
| Enrollment in University or Academic College Throughout 11 Years After High-School Graduation (1 = Yes, 0 = No) | 0.318 | 10,577 | 8,431 | 7,073 |
| ` | (0.466) | (2,419) | (3,462) | (3,212) |
| Completed Years of University or Academic College Throughout 11 Years After High-School Graduation | 1.138 | 2,213 | 648 | 397 |
| | (1.984) | (770) | (1,177) | (1,166) |
| Number of Observations | 2,703 | 5,344 | 5,344 | 5,344 |
| Number of Weighted Observations | 4,171 | 7,995 | 7,995 | 7,995 |

*Notes*: This table presents OLS relationships between high school matriculation outcomes, college schooling, and earnings at adulthood for the natural experiment four-quartile treated schools sample described in the paper. Column 1 reports means and standard deviations for the 2000 cohort in treated schools. Column 2 represents the OLS estimate of a regression where the dependent variable is the annual wage 11 years after high-school graduation, and the independent variables include the same variables as reported in the paper in addition to the outcome variable described in the table. Column 3 reports the OLS estimate when all the variables that appear in Panel A\B are controlled for in the wage regression in addition to the rest of the control variables described in the paper. Column 4 reports the OLS estimate from a wage regression where all the explanatory variables in the table are controlled simultaneously. Standard errors are clustered at the school year level.

**Table 8: Differences-in-Differences Estimates of the Effect of the Teachers' Bonuses Program on Demographic Outcomes 11 Years After High-School Graduation - The Natural Experiment Sample**

| | Married | | Children | |
|---|---|---|---|---|
| | 2000 Cohort in Treated Schools | Estimate | 2000 Cohort in Treated Schools | Estimate |
| | (1) | (2) | (3) | (4) |
| **A. Full Sample** | | | | |
| | 0.584 | -0.011 | 0.451 | -0.003 |
| | (0.493) | (0.013) | (0.498) | (0.011) |
| **B. By Family Income** | | | | |
| High Family Income | 0.554 | -0.003 | 0.407 | 0.006 |
| | (0.497) | (0.019) | (0.491) | (0.015) |
| Low Family Income | 0.621 | -0.033 | 0.506 | -0.017 |
| | (0.485) | (0.017) | (0.500) | (0.018) |
| **C. By Gender** | | | | |
| Boys | 0.493 | -0.005 | 0.338 | -0.003 |
| | (0.500) | (0.020) | (0.473) | (0.018) |
| Girls | 0.677 | -0.007 | 0.569 | -0.001 |
| | (0.468) | (0.016) | (0.495) | (0.016) |

*Notes* : This table presents the differences-in-differences estimates of the effect of the Teachers' Bonuses program on different demographic rate outcomes 11 years after high-school graduation for the natural experiment sample described in the paper. Panel A reports the results for the full sample, Panel B reports the results for the high and low family income samples, and Panel C reports the results by gender. Columns 1-2 report the results for the variable "Married"", which receives the value of 1 is the individual is married 11 years after graduation, 0 otherwise. Columns 3-4 report the results for the variable "Children", which receives the value of 1 is the individual has any children by 11 years after graduation, 0 otherwise. Columns 1 and 3 report the mean and standard error for the 2000 cohort in the treated schools. This cohort did not receive the treatment so it is useful to compare its' average as a benchmark for the treatment effect. Columns 2 and 4 report the Differences-in-Differences estimate for each of the dependent variables listed above. Standard errors are clustered at the school year level.

**Table A1** - Treatment-Control Balancing Tests: The Natural Experiment Sample

| | 2000 | | | 2001 | | |
|---|---|---|---|---|---|---|
| | Treatment | Control | Difference | Treatment | Control | Difference |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| | | | A. School characteristics | | | |
| Religious school | 0.330 | 0.219 | 0.110 | 0.324 | 0.214 | 0.110 |
| | | | (0.163) | | | (0.164) |
| Arab school | 0.158 | 0.000 | 0.158 | 0.155 | 0.000 | 0.155 |
| | | | (0.088) | | | (0.087) |
| Lagged "Bagrut" rate | 0.467 | 0.509 | -0.042 | 0.474 | 0.475 | -0.001 |
| | | | (0.032) | | | (0.053) |
| Two-years lagged | 0.490 | 0.519 | -0.029 | 0.527 | 0.528 | -0.002 |
| Bagrut rate | | | (0.049) | | | (0.034) |
| | | | B. Student background | | | |
| Father education | 10.685 | 10.586 | 0.100 | 10.539 | 10.332 | 0.207 |
| | | | (0.821) | | | (0.838) |
| Mother education | 10.624 | 10.764 | -0.140 | 10.519 | 10.539 | -0.020 |
| | | | (0.849) | | | (0.947) |
| Number of siblings | 3.009 | 2.026 | 0.983 | 2.912 | 1.662 | 1.250 |
| | | | (0.410) | | | (0.384) |
| Gender (male=1) | 0.513 | 0.414 | 0.098 | 0.556 | 0.431 | 0.125 |
| | | | (0.066) | | | (0.061) |
| Immigrant | 0.016 | 0.029 | -0.013 | 0.025 | 0.012 | 0.013 |
| | | | (0.017) | | | (0.018) |
| Asia-Africa ethnicity | 0.218 | 0.325 | -0.107 | 0.235 | 0.276 | -0.041 |
| | | | (0.062) | | | (0.054) |
| | | | C. Student lagged outcomes | | | |
| Math credits gained | 0.337 | 0.277 | 0.061 | 0.256 | 0.453 | -0.197 |
| | | | (0.172) | | | (0.118) |
| English credits gained | 0.155 | 0.077 | 0.078 | 0.107 | 0.079 | 0.028 |
| | | | (0.051) | | | (0.061) |
| Total credits attempted | 5.251 | 4.594 | 0.657 | 5.322 | 5.342 | -0.020 |
| | | | (0.674) | | | (0.498) |
| Total credits gained | 4.308 | 3.761 | 0.547 | 4.218 | 4.482 | -0.264 |
| | | | (0.601) | | | (0.393) |
| Average score | 63.131 | 64.774 | -1.643 | 62.121 | 67.710 | -5.589 |
| | | | (2.591) | | | (2.217) |
| # obs | 2,654 | 2,369 | 5,023 | 2,598 | 2,236 | 4,834 |
| # obs, weighted | 4,095 | 3,818 | 7,913 | 3,812 | 3,679 | 7,491 |
| # schools | 18 | 18 | 36 | 18 | 18 | 36 |

*Notes:* Standard errors in parenthesis are adjusted for school level clustering.

* Observations were weighted with frequency weights in order to have similar number of students in control and treatment schools within each group of schools with close true matriculation rate.

* The schools status of nationality and religiosity does not change. Therefore, any change in the means across years reflects relative changes in the number of students in a cohort.

* This table is based on the math sample.

**Table A2** - Treatment-Control Balancing Tests: The Regression Discontinuity Sample

| | 2000 | | | 2001 | | |
|---|---|---|---|---|---|---|
| | Treatment | Control | Difference | Treatment | Control | Difference |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| | | | A. School characteristics | | | |
| Religious school | 0.100 | 0.301 | -0.201 | 0.095 | 0.290 | -0.195 |
| | | | (0.142) | | | (0.140) |
| Arab school | 0.131 | 0.000 | 0.131 | 0.132 | 0.000 | 0.132 |
| | | | (0.094) | | | (0.096) |
| Lagged "Bagrut" rate | 0.448 | 0.495 | -0.047 | 0.458 | 0.470 | -0.012 |
| | | | (0.017) | | | (0.041) |
| | | | B. Student background | | | |
| Father education | 11.027 | 10.219 | 0.808 | 10.835 | 10.081 | 0.753 |
| | | | (0.591) | | | (0.643) |
| Mother education | 11.095 | 10.526 | 0.570 | 11.027 | 10.527 | 0.501 |
| | | | (0.659) | | | (0.711) |
| Number of siblings | 2.622 | 2.288 | 0.335 | 2.605 | 1.902 | 0.703 |
| | | | (0.352) | | | (0.383) |
| Gender (male=1) | 0.493 | 0.425 | 0.068 | 0.499 | 0.451 | 0.048 |
| | | | (0.058) | | | (0.052) |
| Immigrant | 0.014 | 0.045 | -0.031 | 0.013 | 0.009 | 0.004 |
| | | | (0.021) | | | (0.007) |
| Asia-Africa ethnicity | 0.215 | 0.313 | -0.097 | 0.214 | 0.273 | -0.060 |
| | | | (0.052) | | | (0.054) |
| | | | C. Student lagged outcomes | | | |
| Math credits gained | 0.185 | 0.364 | -0.180 | 0.185 | 0.452 | -0.267 |
| | | | (0.131) | | | (0.128) |
| English credits gained | 0.207 | 0.053 | 0.155 | 0.183 | 0.101 | 0.083 |
| | | | (0.061) | | | (0.088) |
| Total credits attempted | 4.788 | 4.944 | -0.156 | 5.064 | 5.346 | -0.283 |
| | | | (0.476) | | | (0.489) |
| Total credits gained | 4.008 | 4.066 | -0.058 | 4.188 | 4.394 | -0.206 |
| | | | (0.376) | | | (0.384) |
| Average score | 61.671 | 64.548 | -2.877 | 61.797 | 65.770 | -3.973 |
| | | | (2.932) | | | (1.973) |
| # obs | 2,471 | 1,638 | 4,109 | 2,401 | 1,519 | 3,920 |
| # schools | 14 | 13 | 27 | 14 | 13 | 27 |

*Notes:* Standard errors in parenthesis are adjusted for school level clustering.
* The schools status of nationality and religiosity does not change. Any change in the means across years reflects relative changes in the number of students in a cohort.
* This table is based on the math sample.

**Table A3: Differences-in-Differences Estimates of the Effect of Teachers' Bonuses Program on High School Education Outcomes**

| Sample | 4 Quartiles | | 3 Quartiles | |
|---|---|---|---|---|
| | Mean 2000 Cohort in Treated Schools | Treatment Estimate | Mean 2000 Cohort in Treated Schools | Treatment Estimate |
| | (1) | (2) | (3) | (4) |
| Average Matriculation Score | 72.926 | 2.779 | 66.537 | 2.868 |
| | (23.098) | (0.892) | (22.599) | (1.017) |
| Received High School Matriculation (1 = Yes, 0 = No) | 0.532 | 0.026 | 0.423 | 0.055 |
| | (0.499) | (0.020) | (0.494) | (0.023) |
| Number of Credit Units in Matriculation Exams | 22.199 | 0.803 | 20.205 | 0.669 |
| | (10.257) | (0.334) | (10.238) | (0.329) |
| Number of Science Credit Units in Matriculation Exams | 2.339 | 0.589 | 1.340 | 0.343 |
| | (3.813) | (0.181) | (2.910) | (0.154) |
| Number of Honor Level Subjects | 2.491 | 0.128 | 2.034 | 0.092 |
| | (1.821) | (0.062) | (1.631) | (0.065) |
| Number of Observations | 4,162 | 16,031 | 3,072 | 11,921 |

*Notes* : This table presents the differences-in-differences estimates of the effect of the Teachers' Bonuses program on high-school educational outcomes fo the three and four quartiles samples described in the paper. Columns 1 and 3 report the means and standard deviations for the 2000 cohort in the treated schools. This cohort did not receive the treatment so it is useful to compare its' average as a benchmark for the treatment effect. Columns 2 and 4 report the differences-in-differences estimates for each of the dependent variables. Standard errors are clustered at the school year level.

**Table A4: Estimates of Parental Earnings Controls Added to the Basic University Enrollment and Years of Schooling Difference-in-Difference Model - The Natural Experiment Sample**

| | **University Enrollment, Years After High School Graduation** | | | **University Years of Schooling, Years After High School Graduation** | | |
|---|---|---|---|---|---|---|
| | **10** | **11** | **12** | **10** | **11** | **12** |
| | Main Effect | Main Effect | Main Effect | Main Effect | Main Effect | Main Effect |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| **A. Main Effect** | | | | | | |
| Father's Average Earning 2000-2002 | 0.048 | 0.045 | 0.044 | 0.229 | 0.237 | 0.237 |
| | (0.016) | (0.015) | (0.014) | (0.072) | (0.073) | (0.075) |
| Mother's Average Earning 2000-2002 | 0.048 | 0.044 | 0.042 | 0.228 | 0.234 | 0.233 |
| | (0.013) | (0.012) | (0.012) | (0.058) | (0.059) | (0.061) |
| Household's Average Earning 2000-2002 | 0.048 | 0.044 | 0.042 | 0.220 | 0.227 | 0.225 |
| | (0.015) | (0.014) | (0.013) | (0.066) | (0.067) | (0.069) |

Notes: This table presents the differences-in-differences estimates of the effect of the Teachers' Bonuses program on post-secondary education outcomes for the natural experiment sample described in the paper, when controling for parental income. Panel A reports the estimates of the main treatment effect from the differences-in-differences model describes in Table 4, to which each of these parental income controls are separately added. Columns 1 and 5 report results for 10 years after high-school graduation, columns 2 and 6 report results for 11 years after high-school graduation, and columns 3 and 7 report results for 12 years after high-school graduation. Standard errors are clustered at the school year level.

**Table A5: Estimates of Parental Earnings Controls Added to the Basic Earnings and Employment Difference-in-Difference Model - The Natural Experiment Sample**

| | Earnings, Years After High School Graduation | | | | Employment, Years After High School Graduation | | |
|---|---|---|---|---|---|---|---|
| | 9 | 10 | 11 | 9-11 Years After Graduation - Stacked | 9 | 10 | 11 |
| | Main Effect | Main Effect | Main Effect | Main Effect | Main Effect | Main Effect | Main Effect |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| **A. Main Effect** | | | | | | | |
| Father's Average Earning 2000-2002 | 5,203 | 3,974 | 6,174 | 5,105 | 0.025 | 0.010 | 0.018 |
| | (1,377) | (1,852) | (2,137) | (1,588) | (0.011) | (0.009) | (0.009) |
| Mother's Average Earning 2000-2002 | 4,837 | 2,970 | 5,676 | 4,532 | 0.025 | 0.001 | 0.012 |
| | (1,289) | (1,965) | (2,027) | (1,558) | (0.011) | (0.009) | (0.009) |
| Household's Average Earning 2000-2002 | 5,220 | 3,706 | 6,055 | 4,983 | 0.026 | 0.009 | 0.017 |
| | (1,415) | (1,846) | (2,187) | (1,627) | (0.011) | (0.009) | (0.009) |

Notes: This table presents the differences-in-differences estimates of the effect of the Teachers' Bonuses program on earnings and employment outcomes, when controling for parental income. Panel A reports the estimates of the main treatment effect from the differences-in-differences model describes in Table 4, to which each of these parental income controls are separately added. Columns 1 and 5 report results for 9 years after high-school graduation, columns 2 and 6 report results for 10 years after high-school graduation, and columns 3 and 7 report results for 11 years after high-school graduation. Column 4 report results for the stacked earnings of 9-11 years after high-school graduation. The variable "Employment Indicator" receives the value of 1 is the individual has any work record for the given year and 0 otherwise. Standard errors are clustered at the school year level.

**Table A6: OLS Relationships between High School Matriculation Outcomes, College Schooling, and Earnings at Adulthood by Family Income: The Natural Experiment Sample - Treated Schools**

| | High Family Income | | | | Low Family Income | | | |
|---|---|---|---|---|---|---|---|---|
| | | Annual Earnings - 11 Years After High-School Graduation | | | | Annual Earnings - 11 Years After High-School Graduation | | |
| | 2000 Cohort in Treated Schools | Separate Estimate | Joint Estimate Panel A\B | Joint Estimate Panels A + B | 2000 Cohort in Treated Schools | Separate Estimate | Joint Estimate Panel A\B | Joint Estimate Panels A + B |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| **A. High School Matriculation Outcomes** | | | | | | | | |
| Average Matriculation Score | 78 | 377 | 305 | 295 | 68 | 292 | 172 | 150 |
| | (20) | (113) | (117) | (124) | (25) | (68) | (69) | (67) |
| Received High School Matriculation (1 = Yes, 0 = No) | 0.608 | 6,911 | 2,751 | 2,348 | 0.452 | 11,199 | 5,608 | 4,887 |
| | (0.488) | (2,702) | (3,466) | (3,560) | (0.498) | (2,792) | (3,065) | (2,883) |
| Number of Credit Units in Matriculation Exams | 24 | 436 | -104 | -94 | 20 | 747 | 184 | 286 |
| | (09) | (208) | (238) | (232) | (11) | (139) | (245) | (228) |
| Number of Honor Level Subjects | 2.869 | 2,773 | 1,560 | 1,266 | 2.078 | 4,359 | 1,827 | 590 |
| | (1.804) | (1,310) | (1,749) | (1,811) | (1.743) | (1,020) | (1,522) | (1,411) |
| **B. Post Secondary Schooling** | | | | | | | | |
| Enrollment in University or Academic College Throughout 11 Years After High-School Graduation (1 = Yes, 0 = No) | 0.409 | 6,861 | 7,267 | 6,429 | 0.204 | 15,926 | 8,126 | 6,182 |
| ` | (0.492) | (2,982) | (3,948) | (3,907) | (0.403) | (4,096) | (7,073) | (6,630) |
| Completed Years of University or Academic College Throughout 11 Years After High-School Graduation | 1.506 | 1,164 | -120 | -477 | 0.677 | 4,140 | 2,476 | 2,323 |
| | (2.186) | (1,064) | (1,569) | (1,595) | (1.582) | (1,087) | (1,856) | (1,823) |
| Number of Observations | 1,481 | 2,958 | 2,958 | 2,958 | 1,222 | 2,386 | 2,386 | 2,386 |
| Number of Weighted Observations | 2,317 | 4,464 | 4,464 | 4,464 | 1,854 | 3,531 | 3,531 | 3,531 |

*Notes* : This table presents OLS relationships between high school matriculation outcomes, college schooling, and earnings at adulthood for the natural experiment four-quartile treated schools sample described in the paper by family income. Columns 1-4 reports the results for individuals who come from high income families. Individuals included in this sample are those who come from families in which the average household income in 2000-2002 is higher than the mean average household income in the same years. Columns 5-8 report the results for individuals who come from low income families. Individuals included in this sample are those who come from families in which the average household income in 2000-2002 is lower than the mean average household income in the same years. Columns 1 and 5 report means and standard deviations for the 2000 cohort in treated schools. Columns 2 and 6 represents the OLS estimate of a regression where the dependent variable is the annual wage 11 years after high-school graduation, and the independent variables include the same variables as reported in the paper in addition to the outcome variable described in the table. Columns 3 and 7 report the OLS estimate when all the variables that appear in Panel A\B are controlled for in the wage regression in addition to the rest of the control variables described in the paper. Columns 4 and 8 report the OLS estimate from a wage regression where all the explanatory variables in the table are controlled simultaneously. Standard errors are clustered at the school year level.

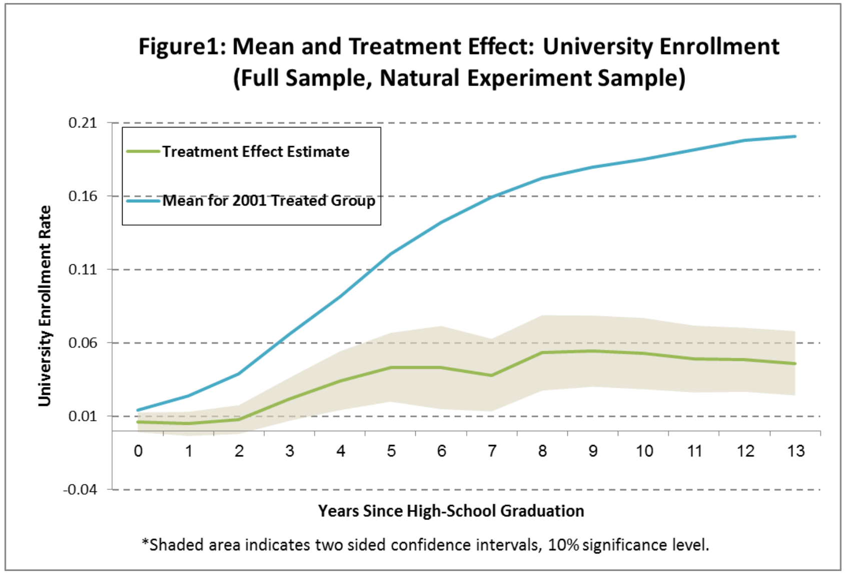


Figure1: Mean and Treatment Effect: University Enrollment
(Full Sample, Natural Experiment Sample)
Treatment Effect Estimate
Mean for 2001 Treated Group
University Enrollment Rate
0.21
0.16
0.11
0.06
0.01
-0.04
0 1 2 3 4 5 6 7 8 9 10 11 12 13
Years Since High-School Graduation
*Shaded area indicates two sided confidence intervals, 10% significance level.

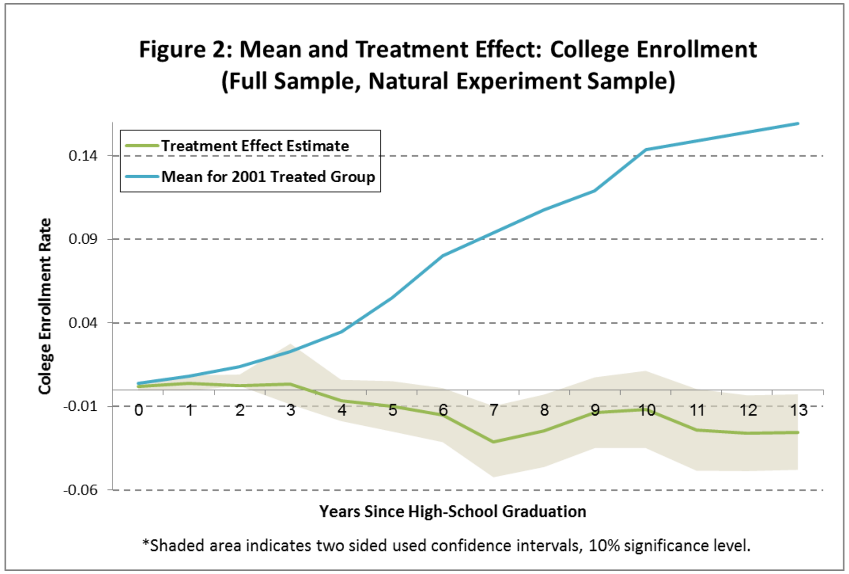


Figure 2: Mean and Treatment Effect: College Enrollment
(Full Sample, Natural Experiment Sample)
Treatment Effect Estimate
Mean for 2001 Treated Group
Colege Enrollment Rate
0.14
0.09
0.04
-0.01
-0.06
0 1 2 3 4 5 6 7 8 9 10 11 12 13
Years Since High-School Graduation
*Shaded area indicates two sided used confidence intervals, 10% significance level.

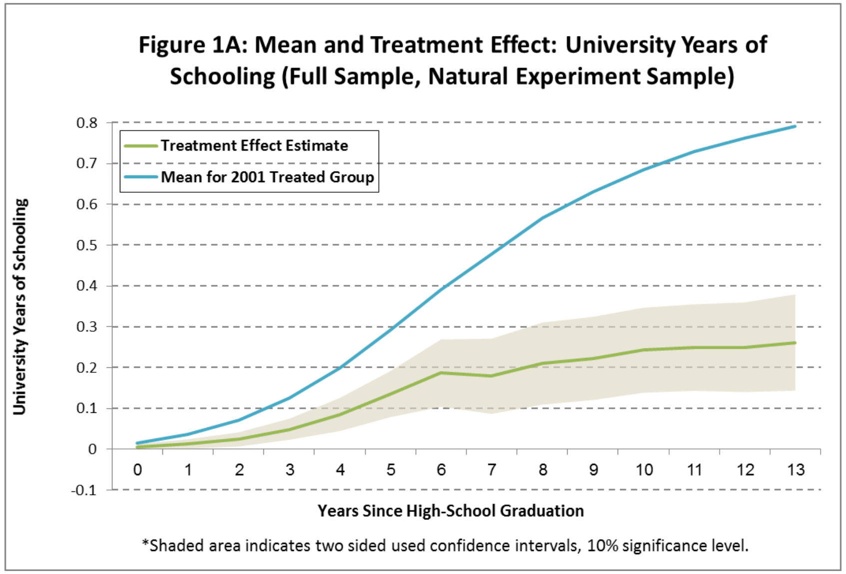


Figure 1A: Mean and Treatment Effect: University Years of Schooling (Full Sample, Natural Experiment Sample)
Treatment Effect Estimate
Mean for 2001 Treated Group
University Years of Schooling
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.1
0 1 2 3 4 5 6 7 8 9 10 11 12 13
Years Since High-School Graduation
*Shaded area indicates two sided used confidence intervals, 10% significance level.

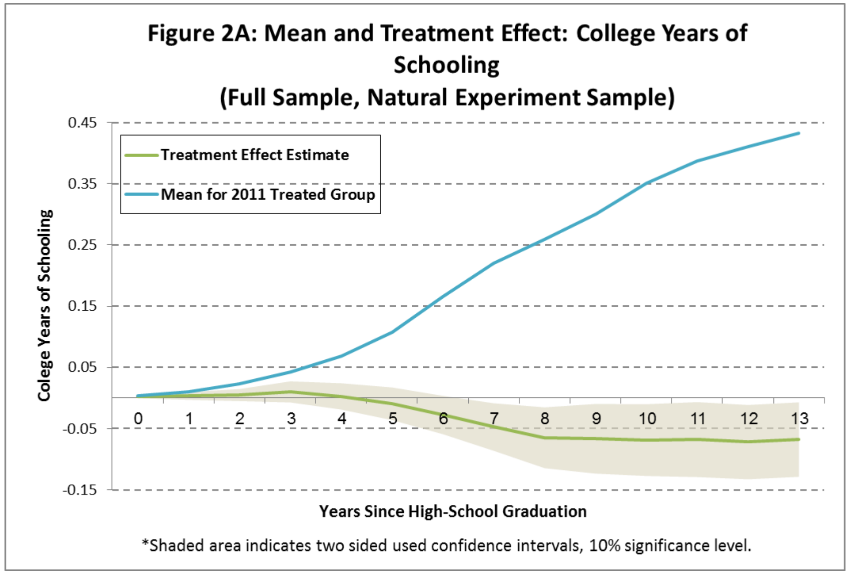
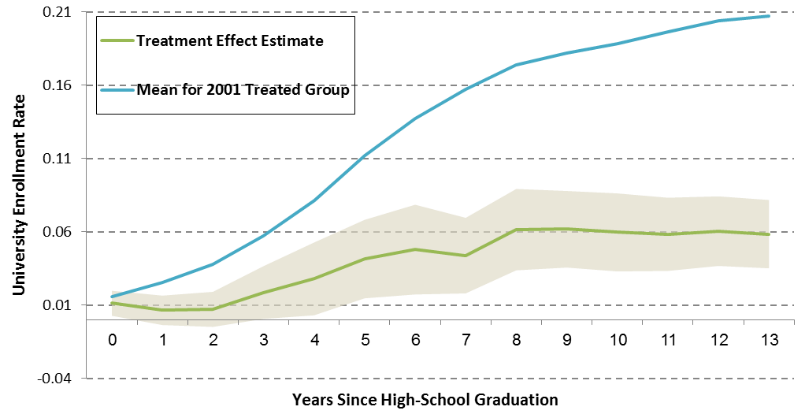


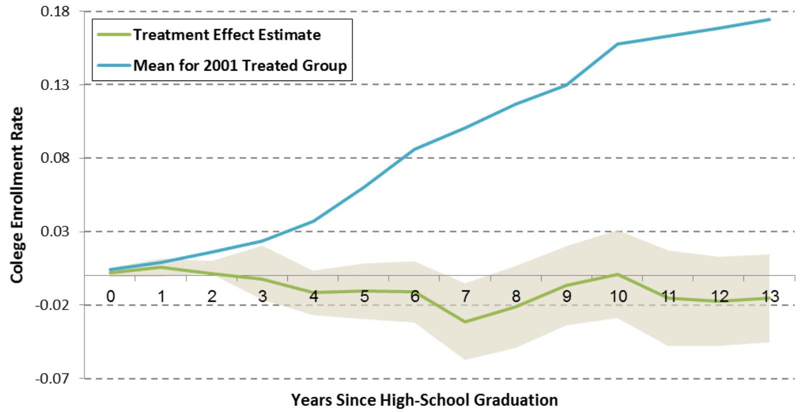
Figure 2A: Mean and Treatment Effect: College Years of Schooling
(Full Sample, Natural Experiment Sample)
Treatment Effect Estimate
Mean for 2011 Treated Group
Colege Years of Schooling
0.45
0.35
0.25
0.15
0.05
-0.05
-0.15
0 1 2 3 4 5 6 7 8 9 10 11 12 13
Years Since High-School Graduation
*Shaded area indicates two sided used confidence intervals, 10% significance level.

**Figure 3: Mean and Treatment Effect: University Enrollment (Full Sample, Regression Discontinuity Sample)**



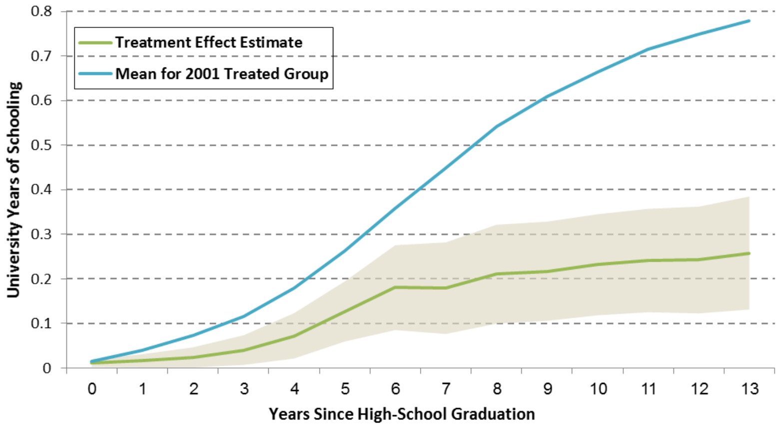

*Shaded area indicates two sided confidence intervals, 10% significance level.

**Figure 4: Mean and Treatment Effect: College Enrollment (Full Sample, Regression Discontinuity Sample)**



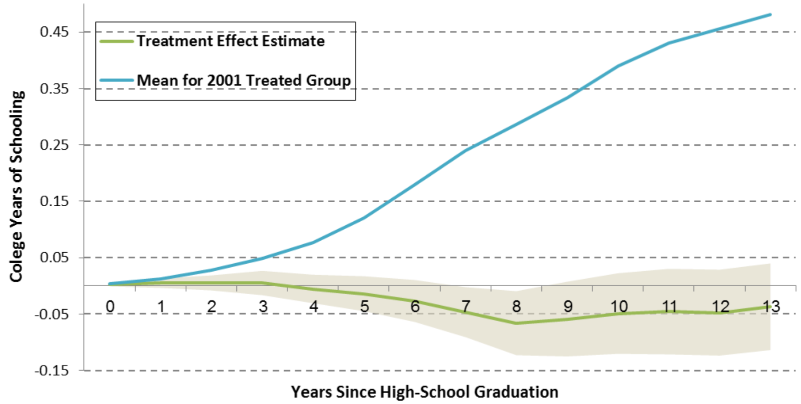

*Shaded area indicates two sided used confidence intervals, 10% significance level.

**Figure 3A: Mean and Treatment Effect: University Years of Schooling (Full Sample, Regression Discontinuity Sample)**



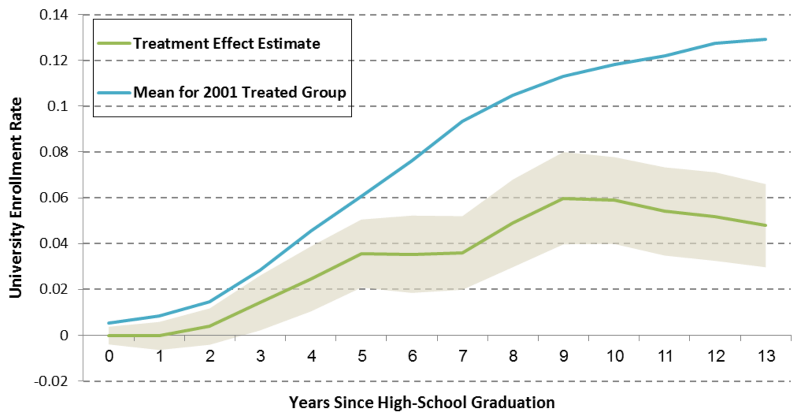

*Shaded area indicates two sided used confidence intervals, 10% significance level.

**Figure 4A: Mean and Treatment Effect: College Years of Schooling (Full Sample, Regression Discontinuity Sample)**



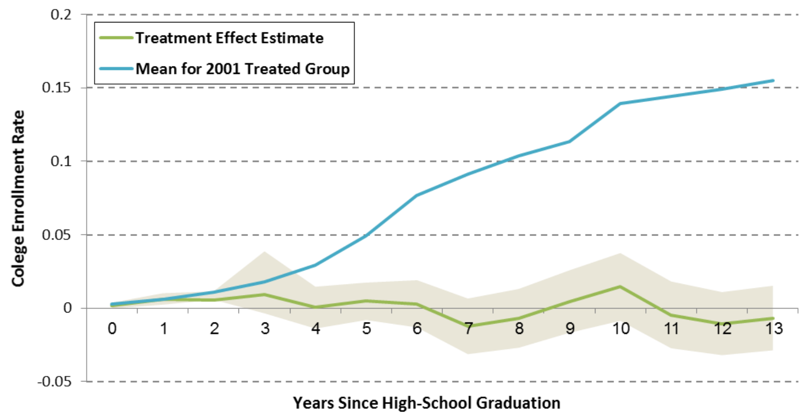

*Shaded area indicates two sided used confidence intervals, 10% significance level.

**Figure 5: Mean and Treatment Effect: University Enrollment (First Three Quartiles, Natural Experiment Sample)**



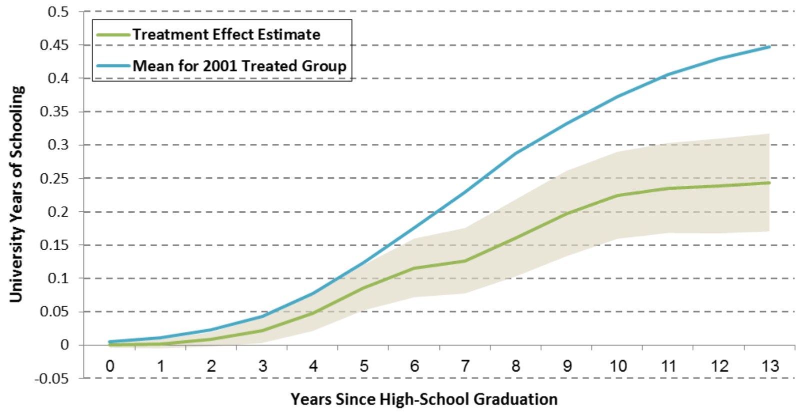

*Shaded area indicates two sided confidence intervals, 10% significance level.

**Figure 6: Mean and Treatment Effect: College Enrollment (First Three Quartiles, Natural Experiment Sample)**



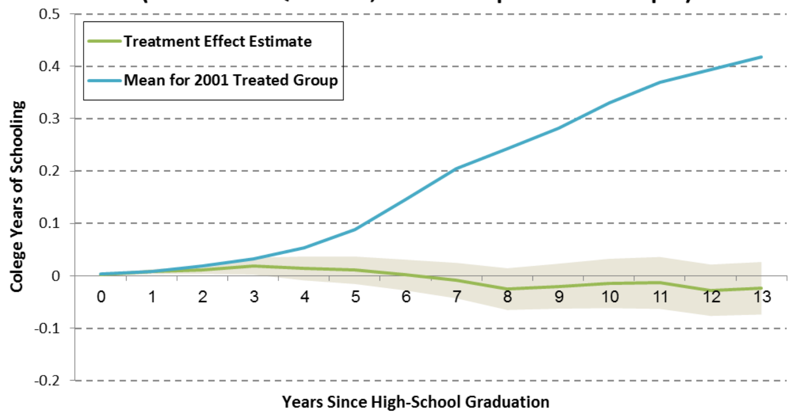

*Shaded area indicates two sided used confidence intervals, 10% significance level.

**Figure 5A: Mean and Treatment Effect: University Years of Schooling (First Three Quartiles, Natural Experiment Sample)**




*Shaded area indicates two sided used confidence intervals, 10% significance level.

**Figure 6A: Mean and Treatment Effect: College Years of Schooling (First Three Quartiles, Natural Experiment Sample)**




*Shaded area indicates two sided used confidence intervals, 10% significance level.

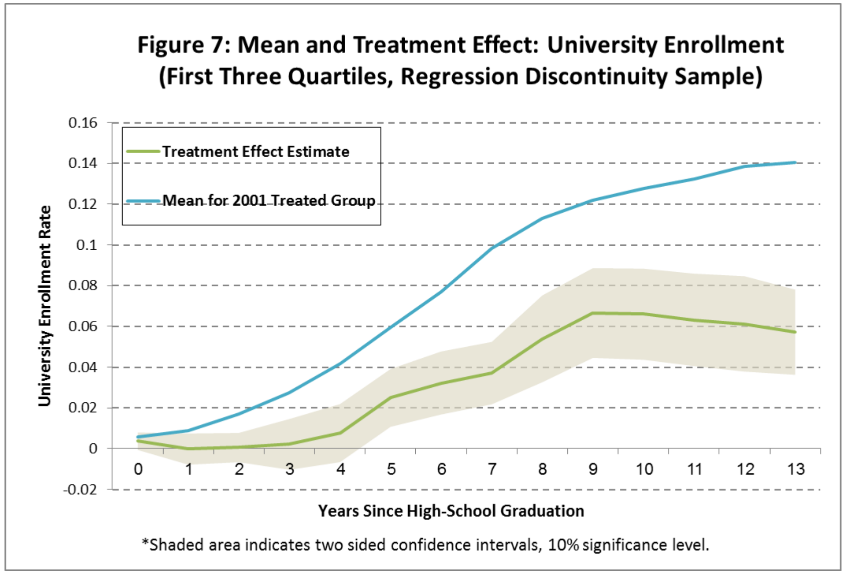


Figure 7: Mean and Treatment Effect: University Enrollment
(First Three Quartiles, Regression Discontinuity Sample)
Treatment Effect Estimate
Mean for 2001 Treated Group
University Enrollment Rate
Years Since High-School Graduation
*Shaded area indicates two sided confidence intervals, 10% significance level.

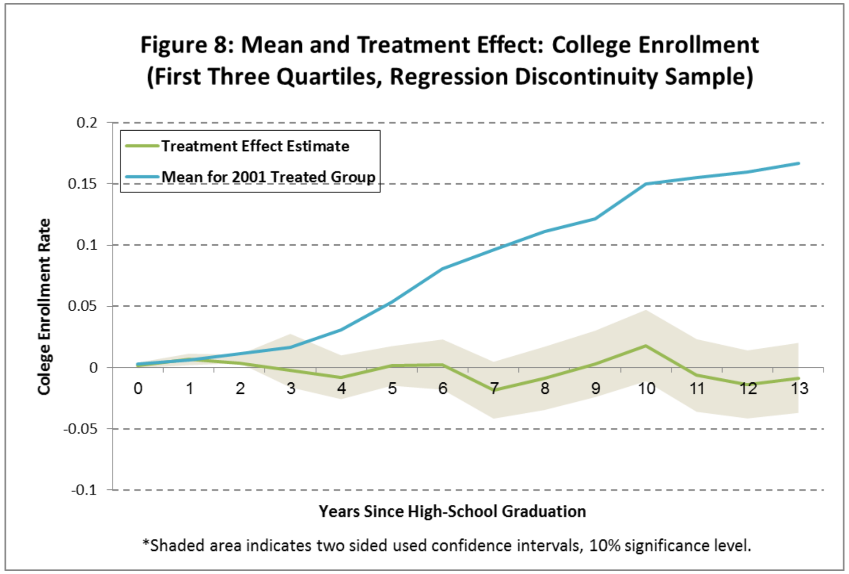


Figure 8: Mean and Treatment Effect: College Enrollment
(First Three Quartiles, Regression Discontinuity Sample)
Treatment Effect Estimate
Mean for 2001 Treated Group
Colege Enrollment Rate
Years Since High-School Graduation
*Shaded area indicates two sided used confidence intervals, 10% significance level.

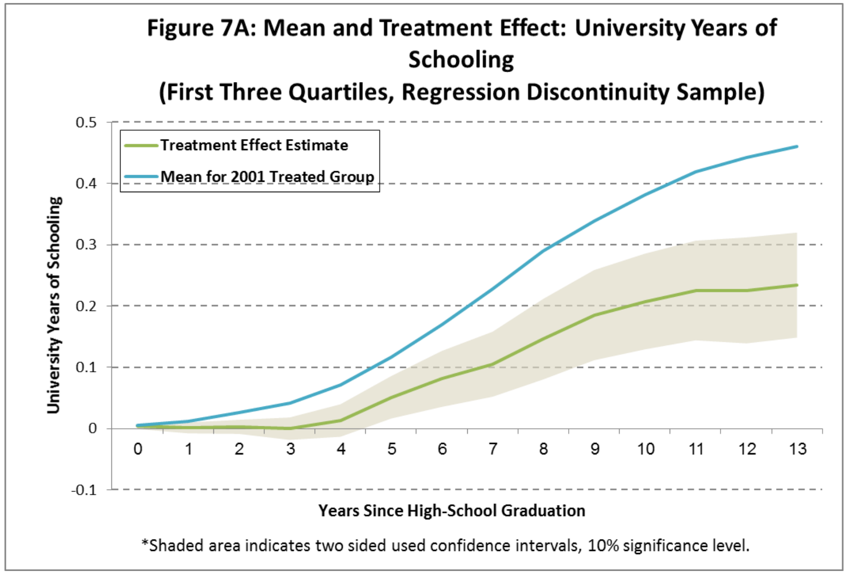


Figure 7A: Mean and Treatment Effect: University Years of Schooling
(First Three Quartiles, Regression Discontinuity Sample)
Treatment Effect Estimate
Mean for 2001 Treated Group
University Years of Schooling
Years Since High-School Graduation
*Shaded area indicates two sided used confidence intervals, 10% significance level.

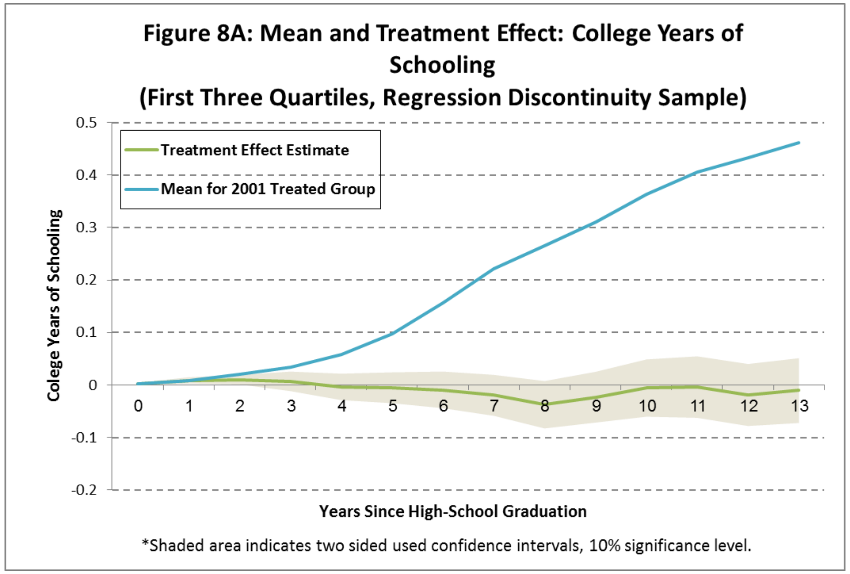
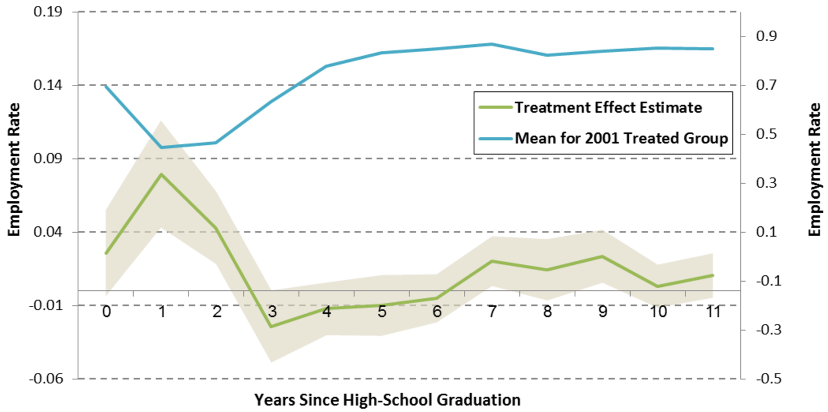


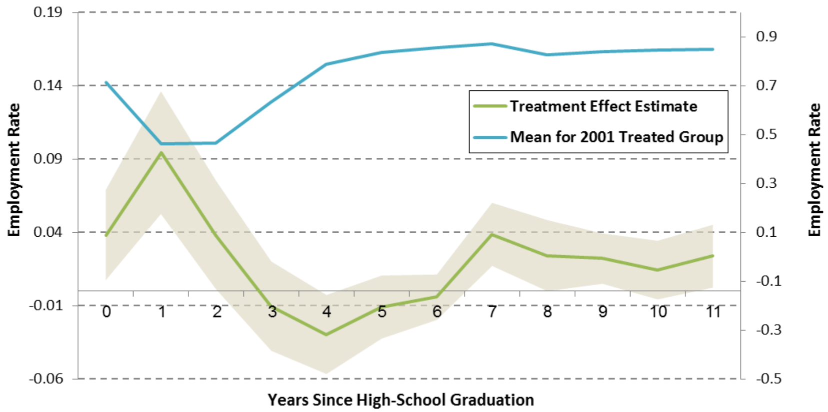
Figure 8A: Mean and Treatment Effect: College Years of Schooling
(First Three Quartiles, Regression Discontinuity Sample)
Treatment Effect Estimate
Mean for 2001 Treated Group
Colege Years of Schooling
Years Since High-School Graduation
*Shaded area indicates two sided used confidence intervals, 10% significance level.

**Figure 9: Mean and Treatment Effect: Employment Rate (Full Sample, Natural Experiment Sample)**



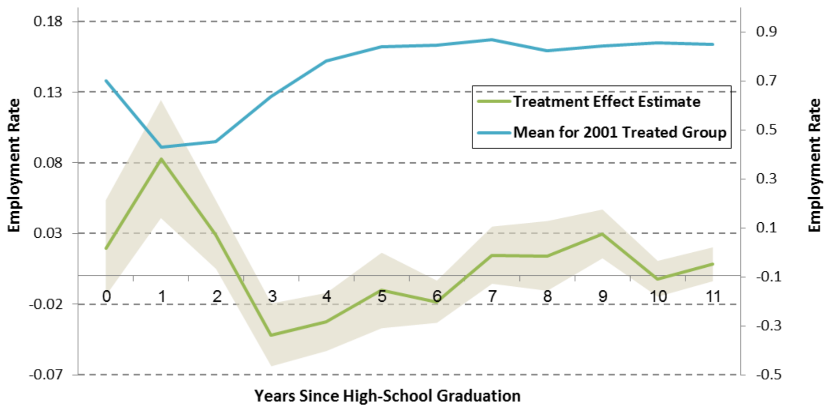

*Shaded area indicates two sided confidence intervals, 10% significance level.

**Figure 10: Mean and Treatment Effect: Employment Rate (First Three Quartiles, Natural Experiment Sample)**



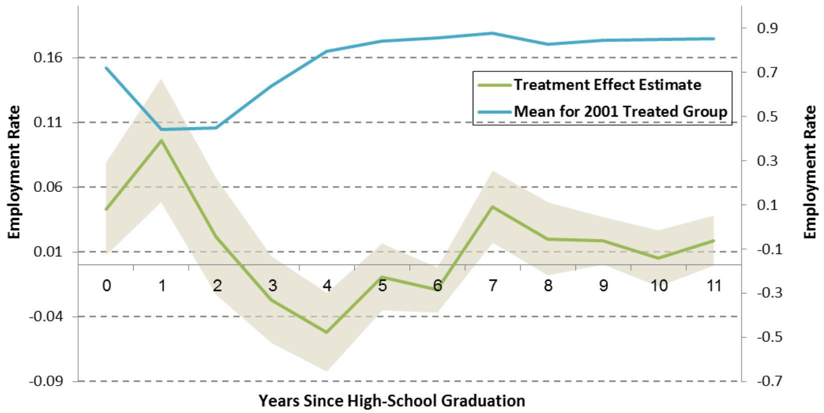

*Shaded area indicates two sided confidence intervals, 10% significance level.

**Figure 9A: Mean and Treatment Effect: Employment Rate (Full Sample, Regression Discontinuity Sample)**




*Shaded area indicates two sided confidence intervals, 10% significance level.

**Figure 10A: Mean and Treatment Effect: Employment Rate (First Three Quartiles, Regression Discontinuity Sample)**




*Shaded area indicates two sided confidence intervals, 10% significance level.

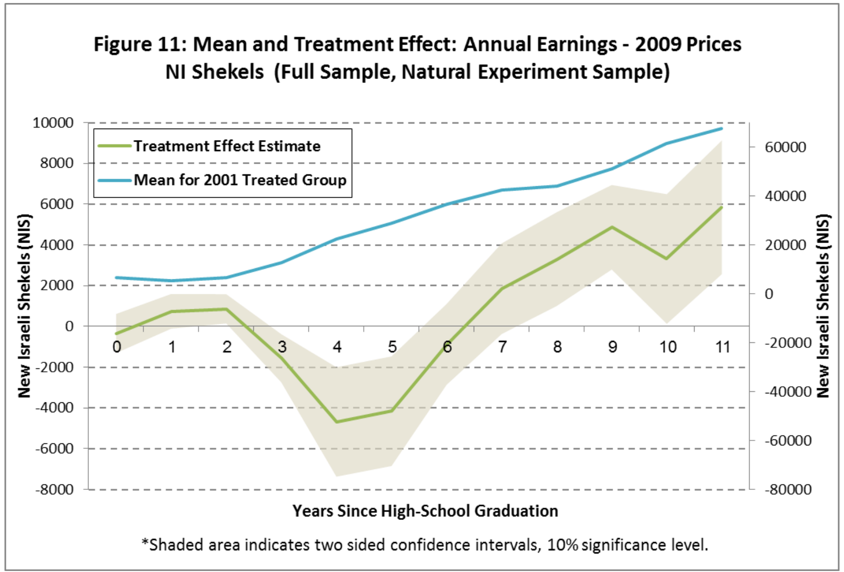


Figure 11: Mean and Treatment Effect: Annual Earnings - 2009 Prices NI Shekels (Full Sample, Natural Experiment Sample)
Treatment Effect Estimate
Mean for 2001 Treated Group
New Israeli Shekels (NIS)
Years Since High-School Graduation
*Shaded area indicates two sided confidence intervals, 10% significance level.

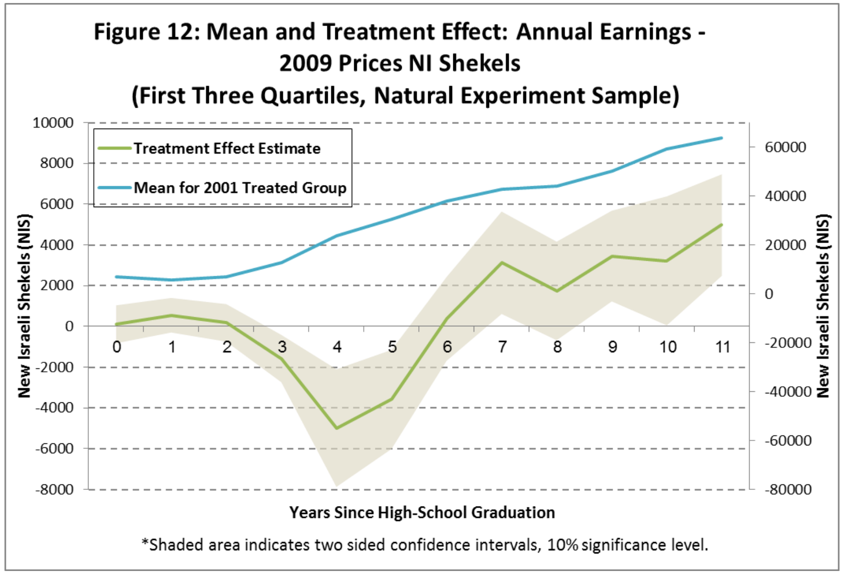



Figure 12: Mean and Treatment Effect: Annual Earnings - 2009 Prices NI Shekels
(First Three Quartiles, Natural Experiment Sample)
Treatment Effect Estimate
Mean for 2001 Treated Group
New Israeli Shekels (NIS)
Years Since High-School Graduation
*Shaded area indicates two sided confidence intervals, 10% significance level.

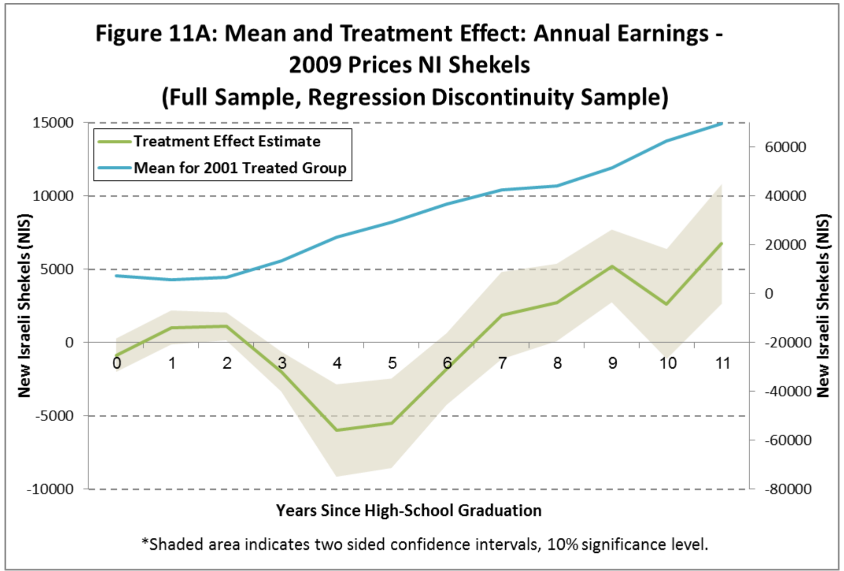



Figure 11A: Mean and Treatment Effect: Annual Earnings - 2009 Prices NI Shekels
(Full Sample, Regression Discontinuity Sample)
Treatment Effect Estimate
Mean for 2001 Treated Group
New Israeli Shekels (NIS)
Years Since High-School Graduation
*Shaded area indicates two sided confidence intervals, 10% significance level.

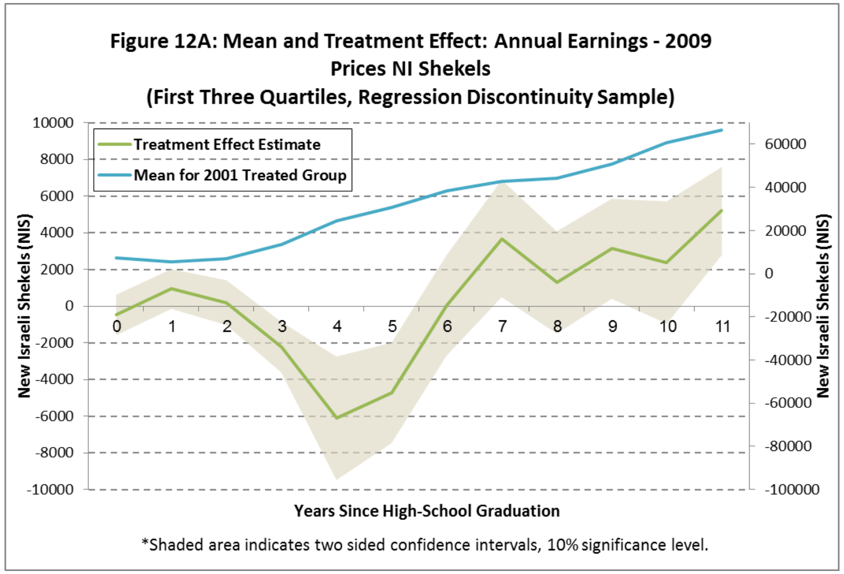



Figure 12A: Mean and Treatment Effect: Annual Earnings - 2009 Prices NI Shekels
(First Three Quartiles, Regression Discontinuity Sample)
Treatment Effect Estimate
Mean for 2001 Treated Group
New Israeli Shekels (NIS)
Years Since High-School Graduation
*Shaded area indicates two sided confidence intervals, 10% significance level.