

Does Raising the Principal's Wage Improve the School's Outcomes? Quasi-experimental Evidence from an Unusual Policy Experiment in Israel

*Victor Lavy**

Royal Holloway University of London
Hebrew University, Jerusalem 91905, Israel
msvictor@huji.ac.il

Abstract

This paper reports estimates of the causal effects of a 50 percent increase in the salary of headmasters of high schools in Israel. The results suggest that the program led to significant improvements in twelfth-grade students' academic achievements. However, the effect was relatively modest, comprising increases of about 5–10 percent in the school mean matriculation rate, average score and number of subjects and credit units taken in matriculation programs. Based on these results and the lack of evidence regarding the effect of increasing teachers' salary, it seems that priority should be given to paying higher wages to school principals.

Keywords: School principals; wages; cognitive achievements

JEL classification: I21; I28; J24

I. Introduction

Proposals to increase teachers' and school principals' salaries have recently attracted considerable attention and support among researchers and policy-makers. Many important reports¹ and proposals advocate substantial salary

* Special thanks go to the staff of the Ministry of Education for their assistance with the data as well as for providing the details and the participants in the Personal Compensation Contracts Program. I benefited from comments from the referees of this journal. Alex Levkov, Issi Romem, Orit Vaaknin and Eli Berglas provided expert research assistance. I also gratefully acknowledge funding from the Falk Research Institute. The views expressed in this paper are those of the author and are not endorsed by program sponsors or funding agencies.

¹ For example, a report by The Teaching Commission (2004), established and chaired by L. V. Gerstner, focuses on the relationship between teachers' quality and their compensation. The purpose of the report is to offer specific policy recommendations to ensure that student learning, rather than teacher protection, is a priority. The report calls for an intense, sustained and effective campaign to revamp our country's teaching force. The authors make recommendations regarding changes in how teachers are paid, how teacher education must be revamped and made a priority by college and university presidents, and how states need to improve—or overhaul—their licensing and certification requirements.

increases as a means of attracting and retaining more talented teachers and encouraging motivation and effort. Salary policies have also been cited as important to offset changes in demand outside of schools and to deal with the potentially unattractive working conditions often found in schools in poor neighborhoods. This paper provides evidence concerning the causal effect of paying school principals higher wages in a manner not explicitly conditional on their future productivity. The wage increase was a 50 percent raise above the implied collective bargaining agreement wage and it was guaranteed for only one year, but was renewed annually in practice for school principals once they were included.

The wage hike to some principals was part of a larger program that granted a similar wage increase to a small number of school teachers. The main objective of the program was to attract good teachers to schools in small and medium-size cities and in rural communities, mainly in the periphery, in an attempt to improve school quality. Within the high schools, the program's overall objective was to improve student outcomes in matriculation exams. In some schools only the principal was included in the program, in others only teachers, and in some both the school principal and some of the teachers. The program grew gradually from about 50 schools in 1997 to over 120 schools in 2005. Here I concentrate only on estimating the effect of paying high-school principals higher wages.

The paper focuses on the effect of the program during its first four years, 1997–2001, as expressed in test scores and related outcomes of matriculation exams administered at the conclusion of high school (twelfth grade). In the case of the principal's salary increase, the whole school is clearly the relevant unit; hence, all students attending that school should be viewed as treated. The measurement of the treatment effect therefore targets all high-school seniors and their matriculation outcomes. School principals play a major role in the preparation for matriculation exams, from motivating teachers and students through allocation of resources for additional instruction time, and special preparation activities (marathon study weekends outside school²) prior to the exams. Productivity and efficiency gains that result from paying school principals higher salaries should thus be observed through improved performance on matriculation exams. Using these outcomes as proxies for school productivity improvements is reasonable given that the reputation of high schools in Israel largely hinges on school

² Angrist and Lavy (forthcoming) have shown that providing financial incentives to students based on their achievements in the matriculation exams led to improved performance in such exams as well as in college enrollment. One mechanism of this effect was increased participation, mainly among girls, in marathon study sessions. The improvement in post-secondary schooling three to five years after the intervention suggests that these marathons achieve real gains in human capital and are not simply teaching-to-the-test activities.

success rate in these exams; a large proportion of school energy is therefore directed toward this aim.

Identification of the effect of interest here is based on the gradual enrollment of school principals into the program from one year to the next and on the use of panel data that permits before and after difference-in-differences estimation. The treatment group includes students of schools where the principal received the salary increase during the first two years of the program or just in its second year; the comparison group includes students from schools where the headmaster did not receive a salary increase during any of the first two years of the program but did receive this higher compensation in later years of the program. Based on observable pre-treatment students' characteristics and outcomes, the two groups are shown to be statistically similar. The school-fixed effects included in the difference-in-differences models account for any unobserved time-invariant heterogeneity that might confound the treatment effect of paying higher wages to school principals (or managers).

Another relevant identification issue is whether it is possible to estimate the unique effect of the wage augmentation of the school principal because, in some schools, this intervention was coupled with wage raises for a few teachers. I therefore control for the variation among treated schools in the estimation, i.e., whether some of the school teachers received higher pay. I also report estimates based on a sample among the treated schools that included only those schools where the principal alone received a salary increase. The results based on this sub-sample will be shown to be practically identical to the estimates obtained from the full sample, thus suggesting that controlling for whether some teachers also received higher wages is sufficient to net out any potential effect of this heterogeneity among treated schools. The fact that very few teachers in each school received this benefit and that not all of them taught the twelfth-grade students in our sample can explain this result.

Most of the available empirical evidence relates to teachers' pay and the effect of teachers' pay on students' outcomes,³ but there is no evidence about the link between school management compensation and school performance. Most estimates on the effect of teachers' compensation suggest that there is no causal link between teachers' wages and students' achievements. One explanation for this result is that the true connection between teachers' pay and their productivity is strong although methodological and data problems have impeded its identification; see e.g. Loeb and Page (2000). Others take a less sanguine position, arguing that the evidence accurately captures the weak performance incentives in public schools whereby principals make

³ See, for example, Cohen and Murnane (1985, 1986), Cohn (1996), Kenny (1980), Levinson (1988), Loeb and Page (2000), Murnane and Olsen (1989, 1990) and Rumberger (1987).

hiring and retention decisions that are not strongly linked to teacher quality; see Ballou and Podgursky (1995). The evidence, however, is quite strong on one point: teacher quality is an important determinant of achievement; see e.g. Hanushek, Kain and Rivkin (1999, 2004, 2005). A recent study by Jacob and Lefgren (2007) linked teachers' and principals' quality to schools' accountability and the measurement of teachers' performance. Their findings suggest that subjective principal assessments of teachers predict future student achievement significantly better than teacher experience, education or actual compensation, though not as well as value-added teacher quality measures. They also report that a principal's overall rating of a teacher is a substantially better predictor of future parent requests for that teacher than *either* the teacher's experience, education and current compensation *or* the teacher's value-added achievement measure. These findings outline an important mechanism through which principals can enhance their school productivity and speaks for the importance of motivating principals through better pay.

The evidence concerning the effect of the program suggests that increasing school principals' wages leads to improved students' outcomes of senior students in their matriculation exams. The estimated effects are marginally more pronounced and more precise when the analysis is based on schools that participated in the program for two consecutive years, although they are modest in size. The estimated treatment effects are very similar for students from low and high socio-economic background and they do not vary by school size. It should be noted, however, that the estimates are based on a relatively small sample of treated and control schools. They should therefore be interpreted with caution until additional evidence allows firmer conclusions to be drawn.

Despite the fact that the pay bonus to school principals was unconditional on improved performance, its positive effect on student outcomes may be explained by a few "hidden" motivational factors. For example, the explicit objective of the program was to improve achievements in the matriculation program and school principals may have worked towards achieving this objective. They could also have deduced that in the absence of such improvements, their pay bonus would be terminated. Efficiency wage behavior could also have been a motivational factor, as school principals could perceive their pay inclusive of the bonus to be fairer compensation. A related argument is that if it is common knowledge to teachers, community leaders and the superintendent of the school that the principal received a pay bonus, this may put pressure on recipients to make more effort to improve school matriculation outcomes.

The remainder of the paper is organized as follows. The program that increased school principals' wages is outlined in Section II. The data used in this paper are described in Section III. Section IV addresses the

identification and estimation strategies. Sections V and VI report the empirical results and their interpretation. Section VII concludes.

II. The Wage Increase through the Individual Compensation Contracts Program

In 1997, Israel's Ministry of Education and the Ministry of the Treasury announced a new experimental program that would allow paying teachers and school principals a 50 percent salary hike. The teachers' unions⁴ agreed to the program as long as the number of participants did not exceed 1,400 and as long as the salary increase was embedded in a special contract signed with each participant rather than within the framework of the collective bargaining agreement. The schools in the program were to be selected from localities included in the National Priority Areas (NPA) program that entitles its participants to additional government grants for social programs. The NPA primarily includes small and medium-size cities, mainly those located in the periphery of the country, and a few schools located in poor neighborhoods in larger cities.

The school district supervisors and Ministry officials who managed the program selected participants from eligible schools. In its first year (1997), the program included only 40 secondary schools and an equal number of primary schools. Many more schools were added gradually in later years. About 200 teachers and school principals participated in the program's first year; enrollment increased rapidly in each subsequent year and reached over 700 in 2005. Individuals were informed about their participation in the program at the beginning of each academic year.

III. Israel's Secondary Schooling System and the Data

Israel's education system consists of elementary school (grades 1–6), middle school (grades 7–9) and high school (grades 10–12). High-school students are enrolled in an academic track leading to a matriculation (*bagrut*) diploma or in a vocational track leading to a vocational diploma. High-school students in Israel are tested in a series of matriculation examinations (*bagruyot*). National exams are given in core and elective subjects beginning in the tenth grade, continuing in the eleventh grade and concluding in the twelfth grade, when most of the tests are taken. Pupils choose to be tested at various levels in each subject, with each test awarding one to five

⁴ Teachers in primary schools are organized in one union and teachers in junior and senior secondary schools are organized in another independent union.

credit units (hereafter: credits) per subject.⁵ The final matriculation score in a given subject is the mean of two intermediate scores. The first is based on the score in the national exams that are external to the school because they are written, administered, supervised and graded by an independent agency. The scoring process for these exams is anonymous; the external examiner is not told the student's name, school or teacher.⁶ The second intermediate score is based on a school-level (internal) exam that resembles the national exam in material and format but is scored by the student's own teacher.

Some subjects are mandatory; many of these must be taken at the level of at least three credits. Tests that award more credits are more difficult. A minimum of 20 credits is required to qualify for a matriculation diploma. About 52 percent of high-school seniors earned matriculation diplomas in 1999 and 2000, i.e., passed a sufficient number of exams to be awarded 20 credits by the time they graduated from high school or shortly thereafter; see Ministry of Education (2001).

The Data

The data used in this study came from records kept by Israel's Ministry of Education on the population of Israeli high-school students for the 1995–2002 school years. The data were based on reports sent by school authorities to the Ministry of Education at the beginning of the school year. The file contains an individual identifier, a school and class identifier, and detailed demographic information on all pupils in all grades. The current study focuses on twelfth-grade cohorts for the period in question (1995–2002).

The individual data on high-school enrollment and outcomes from matriculation exams⁷ taken in 1995–2002 were linked with information on school principals participating in the program. Five measures of success in

⁵ In Israel, a high-school matriculation certificate is a prerequisite for university admission and one of the most economically salient educational milestones. Many countries and some American states have similar high-school matriculation systems. Examples include the French Baccalaureate, the German Certificate of Maturity (*Reifezeugnis*), the Italian Diploma di Maturità, the New York State Regents' Examinations and the recently instituted Massachusetts Comprehensive Assessment System.

⁶ Exams are held in June and January, and all pupils are tested in a given subject at the same date. The national exams are graded centrally by two independent external examiners, with the final score an average of the two.

⁷ Towards the end of the last year of high school, students sit for exams in various subjects, some compulsory and some elective. Passing an exam awards matriculation credit units and a minimum of 20 such units is required to qualify for a matriculation certificate, which is a necessary though not sufficient requirement for admission to university. As noted in footnote 5 above, similar high-school matriculation exams are found in many countries and in some states in the United States.

matriculation exams were employed: an indicator of *bagrut* status (=1 if the student earned a matriculation diploma), the average score in all matriculation exams, the overall number of matriculation credits and the number of matriculation subjects studied at the honors level. The first two outcomes are particularly important in Israel because the first is required for university admission whereas the second is applied during the selection of candidates for high-demand courses of study such as medicine, engineering and computer science.

IV. Identification Strategy and Estimation

Selection of participants to the program involved selection of the schools followed by selection of the school principals. In most cases, the principals selected received the salary increase in the first year but continued to benefit from this increase as long as their school remained in the program. Selection of the schools required the discretionary judgment of the district supervisor and the program administrators. Therefore, because the principal's participation was conditional on the selection of his school for the program, it can be viewed as a non-random event. To estimate the effect of this non-randomness, I rely on the gradual implementation of the program as the basis for identification. If the early and later participating schools were selected through the same process, they might be very similar in observed and unobserved characteristics, particularly those characteristics that might confound the treatment effect of the program. Therefore, a natural comparison group may be taken from later participants in the program. Twenty high schools were enrolled into the program either in its first (1997) or second (1998) year and 20 others during the following four years (1999–2002). However, only 17 schools in each of these two groups followed the matriculation study program. I therefore used these 34 schools as the base sample in this study.⁸

The mean characteristics of the students at schools that enrolled early into the program closely resemble the observed characteristics of the students at schools that enrolled in later years, both for pre- and post-treatment twelfth-grade cohorts. These results can be seen in Tables 1 and 2, which present data based on cohorts of twelfth graders who graduated before the program was initiated (1995–1996) and cohorts that graduated during the program (1997–1998).

⁸ Most of the high schools in Israel follow the matriculation study program and their students take matriculation exams. However, some of the very religious (orthodox) high schools and most of the very low vocational high schools do not, although they participate in many Ministry of Education programs such as the one studied in this paper.

Table 1. *Balancing tests for pre- (1995–1996) and post- (1997–1998) treatment cohorts*

	Pre-treatment cohorts		Post-treatment cohorts	
	Control group mean	Treatment–control difference	Control group mean	Treatment–control difference
Demographic variables				
<i>Father's years of schooling</i>	11.0	0.615 (0.917)	11.5	–0.142 (0.788)
<i>Mother's years of schooling</i>	11.2	0.652 (0.847)	11.5	0.166 (0.762)
<i>Number of siblings</i>	3.04	0.068 (0.528)	3.19	0.129 (0.414)
<i>Gender (male = 1)</i>	0.446	–0.008 (0.061)	0.472	–0.022 (0.065)
<i>Immigration status</i>	0.068	–0.049 (0.022)	0.061	–0.030 (0.030)
<i>Country of origin</i>				
Israel	0.432	0.084 (0.065)	0.408	0.052 (0.073)
Asia/Africa	0.360	–0.038 (0.093)	0.344	0.007 (0.092)
Europe/USA	0.109	–0.024 (0.028)	0.123	–0.047 (0.028)
Ethiopia	0.006	–0.003 (0.005)	0.037	–0.034 (0.022)
USSR	0.093	–0.019 (0.031)	0.088	0.020 (0.034)
<i>Number of students (schools)</i>				
Control group	2,222	(17)	2,560	(17)
Treatment group	3,037	(17)	3,063	(17)

Note: Standard errors in parentheses are adjusted for school-level clustering.

Table 1 shows balancing test results for treatment and control schools for background characteristics: father's and mother's years of schooling, number of siblings, gender, immigration status and ethnic background defined by country of birth of the student's father (Israel, Asia/Africa, Europe/US, Ethiopia, former USSR). Each entry in the table represents an estimate from a bivariate regression of a characteristic on the treatment status. The first sample pools the two pre-treatment cohorts (1995 and 1996) and the second sample pools the two treated cohorts (1997 and 1998), but the results are not different when the balancing tests are carried out for each cohort separately.

Columns 1 and 2 report the results for the pre-treatment (1995 and 1996) cohorts. The first column shows the mean of the respective variable in the control group; the second column shows the treatment–control difference and its estimated standard error (estimated while allowing for school-level clustering). All the difference estimates are not statistically different from

zero with the exception of the immigration status indicator, and the majority of the values are very small. For example, the mean of father's years of schooling in the control schools is 11.0 years, with the treatment-control difference only 0.615 (SE = 0.92). The mean of mother's years of schooling is also 11.2 years in the control schools, with the treatment-control difference only 0.85 (SE = 0.53). Similarly, the control mean of number of siblings is 3.0, with the difference between the two groups again very small, 0.068 (SE = 0.53). Balancing tests for a series of dummy variables indicating the student's ethnic origin are also shown in the table. For example, 36 percent of the students in the control group were of Asian/African origin; the respective mean in the treatment group was 32 percent.

Columns 3 and 4 in Table 1 show similar balancing tests for the first two cohorts that participated in the program, having graduated in 1997 and 1998. The results are very similar to those of the full sample; there appear to be no observable differences in the characteristics of the students who graduated during the first two years of the program from those who graduated in the later two years. Although these two cohorts appear to come from more educated families, there are no differences in parental schooling between the two groups.

With regard to the comparison of the treatment and the control groups in these post-treatment cohorts, all the difference estimates are not statistically different from zero, with the majority of values even smaller than those observed for the two pre-treatment cohorts. For example, the mean of father's years of schooling in the control schools is 11.5 years, with the treatment-control difference only 0.15 (SE = 0.79). The mean of mother's years of schooling is also 11.5 years in the control schools, with the treatment-control difference only 0.17 (SE = 0.77). Similarly, the control mean of number of siblings is 3.2, with the difference between the two groups very small, 0.129 (SE = 0.41). Balancing tests for a series of dummy variables indicating the student's ethnic origin are also presented in the table. For example, 34 percent of the students in the control group were of Asian/African origin; the respective mean in the treatment group was also exactly 34 percent. The estimated difference that is closest to being significant refers to the proportion of students of European/US origin, -0.047 (SE = 0.028). However, the mean differences for the groups of other ethnic origins (Israel, former USSR and Ethiopia) are not significantly different from zero.

The four cross-sections used for the analysis in Table 1 can be applied to compare the change in school mean matriculation outcomes of seniors in the treatment and comparison schools. This approach amounts to a difference-in-differences estimation when school-fixed effects are included in the model. The main assumption underlying this strategy is that in the absence of any intervention, the average change in matriculation

outcomes would have been the same for treatment and comparison schools.⁹ Although the validity of this assumption cannot be verified, the similarity in characteristics of the graduating cohorts in schools that started the program earlier and later, as well as relatively similar mean outcomes of the two groups prior to the program's introduction, may provide strong supporting evidence of the small likelihood of omitted variable bias. The lack of any significant differences in the matriculation outcome trends between treated and comparison schools, particularly the lack of an observed Ashenfelter (1978) dip in outcomes in the years prior to the treatment, is likewise comforting in this regard.

As noted in Section II, the augmentation of school principals' salary was part of a program that also provided similar wage increases to some teachers. In about half of the schools in our treated sample (i.e., schools where the principal received a salary increase) a few of the teachers also received a wage increase. Therefore, an important issue is whether it is possible to identify uniquely the effect of the wage augmentation of the school principal. For this purpose I rely on several features of the program that allow me to distinguish between the effect of the school principal salary raise and the potential effect of teachers' higher compensation. First, very few teachers in each school received this increase (in some schools just one teacher and at most 5 percent of the teaching staff). In some cases these teachers did not teach a matriculation subject and therefore could not have a significant effect on *bagrut* outcomes. Another factor that very much diminishes the likelihood that the school principals' treatment effect estimated here is confounded with higher teachers' pay is that some of these teachers actually taught at middle-school grades (seventh to ninth), while I focus on the effect on outcomes of twelfth graders in the matriculation exam. The implication of these factors is that the intensity of the treatment of teachers' higher salary is too small to have a meaningful and measurable effect on the matriculation outcomes of twelfth graders. Nevertheless, in the estimation of the variation among treated schools, I control for whether some of the school teachers received higher pay. I also report estimates based on a sample that included schools where only the school principal received a salary increase.

Estimation Framework

Treatment effects can be estimated using both micro-data at the student level and aggregated data at the school level. In a model without additional covariates, analyses using micro-data or school averages weighted according to the number of twelfth graders will provide identical results.

⁹ The model also assumes constant and additive treatment effects.

However, when school and student characteristics are added to the model, the two estimates may differ because adding these covariates may improve the estimates' precision. Although the difference-in-differences models include school-fixed effects, it is important to control for student and school characteristics, particularly when analyzing data from low-achieving schools that often experience severe year-to-year volatility in student characteristics; see Koretz (1996). The following model was first estimated:

$$y_{ist} = \beta' X_{ist} + \omega' S_{st} + \alpha T_{st} + \nu_{st} + \varepsilon_{ist}, \quad (1)$$

where i indexes students, s indexes schools, and t indexes years. y_{ist} is the matriculation outcome, X_{ist} a vector of student characteristics that includes parents' schooling, gender, number of siblings and immigration status, S_{st} a vector of school characteristics that includes the school's religious status, the number of students enrolled in the school, the square of enrolled students and the number of students in twelfth grade and whether some teachers in school received higher pay due to the program (this indicator is set to zero for all schools in the pre-treatment period); T_{st} is a dummy variable that indicates program participation status. For $t=1995$ or 1996 (pre-program years) all schools have $T_s=0$. For $t=1997$ all schools that enrolled in this year get $T_s=1$ and for others $T_s=0$. For $t=1998$ all schools that participated in the program in this year (those already enrolled in 1997 and the new enrollers in 1998) get $T_s=1$ and others have $T_s=0$. $\nu_{st} + \varepsilon_{ist}$ is the error term, comprised of a school year-specific random element and an individual random element.

To control for pre-existing differences between treated and comparison schools as well as to improve the efficiency of the estimates, the data for pre-treatment and post-treatment years were stacked. The difference-in-differences model with school- and year-fixed effects was estimated as follows:

$$Y_{ist} = \beta' X_{ist} + \omega' S_{st} + \varphi_s + \delta_t + \alpha D_{st} + \nu_{st} + \varepsilon_{ist}, \quad (2)$$

where φ_s is a school-fixed effect, δ_t is a time effect, and D_{st} equals the interaction between the dummy for treatment years and treatment status for school s .

The variation in the number of years that a school principal participated in the program (one or two years) may facilitate more precise estimation of the treatment effect. Exploration of this direction requires assessing whether the intensity of treatment affects the program's impact, i.e., if the number of years a principal is in the program leads to a differential effect on school matriculation outcomes. The main potential problem of this strategy is the possible endogeneity of treatment intensity. The correlation between school treatment intensity and potential achievements may be positive or negative. For example, successful school principals (e.g. those whose schools have

high matriculation rates) may have been enrolled in the program earlier than others. On the other hand, program administrators may have targeted principal's salary increases to weaker schools. To assess the likelihood of these possibilities, I estimated alternative models that evaluate whether pre-program achievements pertaining to the school predict school treatment intensity. The evidence suggests that this was not the case.

V. Results

This section opens with a discussion of the results based on the full sample. Table 2 shows the results of comparing the outcomes of four cohorts: the two pre-treatment graduating cohorts (1995–1996) pooled together and the two treated cohorts (those that graduated high school in 1997–1998), also pooled together. The table shows the results for the six outcomes described in the data section: matriculation certificate eligibility, average matriculation score, average score including the bonuses for honors-level subjects, number of matriculation credits, number of matriculation subjects and number of subjects at honors level. For each outcome for the pre- and post-treatment periods, the table shows the mean outcome in the control group, the simple mean differences between the treatment and the control group and the respective conditional mean differences.

The estimates in the first column of Table 2 show that the program schools had lower pre-treatment school outcomes than the comparison schools but that the differences were not significantly different from zero. For example, the unconditional mean of the *bagrut* eligibility rate in 1995–1996 was 48.6 percent in the comparison schools but 3.1 percent lower in the treatment schools. However, the *t*-value for this difference was only 0.72. The mean *bagrut* score was 65.9 in the comparison schools but lower by 1.75 points in the program schools; the standard error of this difference was 2.68. Similarly negligible and non-significant differences are observed for the other outcomes shown in the table. However, controlling for students' characteristics slightly increased these baseline differences; in some cases the differences were sufficient—relative to the estimated standard error—to make them marginally statistically significant. This evidence weakens somewhat the pre-treatment balance between the treatment and control groups. However, given that these differences are relatively constant in both pre-treatment years, conditional on the school-fixed effects, the comparison group can be viewed as reliable for the identification of counterfactual results for the program.

Columns 4–6 in Table 2 present the results for the first two cohorts that could potentially be affected by the program. First, comparing the simple treatment–control mean differences of the 1997–1998 cohorts with those of the two pre-treatment cohorts (1995–1996) reveals improvements in the

Table 2. Pre- and post-treatment outcome means

	1995-1996		1997-1998	
	Control group mean	Treatment-control difference	Control group mean	Treatment-control difference
Outcome				
<i>Eligible for bagnet</i>	0.486	-0.031 (-0.043)	0.508	-0.021 (-0.047)
<i>Mean score</i>	65.9	-1.75 (-2.68)	66.0	-0.091 (-2.74)
<i>Mean score (weighted)</i>	71.4	-2.24 (-3.02)	72.5	-1.29 (-3.20)
<i>Credit units</i>	20.3	-1.66 (-1.10)	20.5	-1.26 (-1.14)
<i>Subjects</i>	7.17	-0.541 (-0.310)	7.09	-0.222 (-0.308)
<i>Honors-level subjects</i>	2.16	-0.132 (-0.208)	2.27	-0.214 (-0.217)
<i>Number of students (schools)</i>				
Control group	2,222	(17)	2,560	(17)
Treatment group	3,037	(17)	3,063	(17)

Note: Standard errors in parentheses are adjusted for school-level clustering.

outcomes of students from schools that were enrolled in the program. For example, the matriculation rate unconditional mean difference was reduced by a third of its pre-treatment level and the test-score gap declined from 1.8 points to zero.

Comparing the post-treatment *controlled* differences to the respective pre-treatment differences revealed further evidence of an improvement in the educational outcomes of the treated population. For example, the differences in the average score in the matriculation exams declined from -3.4 to -1.9 . The gap in the number of matriculation subjects declined from -0.724 to -0.408 .

Table 3 shows the results of estimating the difference-in-differences model based on the panel of the four-cohort stacked data with school-fixed effects. The first column reports the control-group means of all six outcomes, while column 2 shows the difference-in-differences estimates of the program effect without any control included except year dummies and school-fixed effects. These treatment estimates are all positive (except for the estimate for honors-level subjects) and they resemble the simple mean difference-in-differences derived from comparison of the before and after cross-sectional results presented in Table 2. However, these estimates are not very precise and none is statistically different from zero. In the third column I present the estimates when students' demographic variables are added as controls to the simple difference-in-differences equation. The estimates are again positive (excluding the effect on the number of subjects studied at honors level) and there is a sharp increase in their precision. The estimates are significant at the 5 or 10 percent level of significance, with t -values ranging from 1.73 (mean-weighted score) to 2.3 (number of matriculation subjects). They are also sensitive to whether the regressions include student-level individual controls. The size of all the effects is relatively modest, about 6–12 percent of the control mean for all outcomes. The largest estimated effect is on the matriculation rate, a 5.8 percent increase relative to a 49.3 matriculation rate in the control group. The effect on credit units is 1.39 relative to 19.7 in the control group; on matriculation subjects it is 0.564 relative to 6.98 in the control group.

The estimated effect of the indicator as to whether some teachers in the school also received higher pay (not reported in the table as only the estimated treatment effects are shown) varied in sign across outcomes and was insignificant in all cases.

The last column in Table 3 reports estimates based on a sample of treated and control schools where none of the teachers received pay increases as part of the program. These estimates are very similar to those in column 3 of the table. For example, the effect on the matriculation rate in column 4 is 0.062 and in column 3 it is 0.058. For the mean score, the two respective estimates are 4.699 and 4.428. However, the estimated standard errors are

Table 3. DID estimates of effect of school principals' wages on students' outcomes

Outcomes	Sample: all schools		Sample: schools where only the principal received a wage increase	
	Control group mean	Treatment-control difference	Treatment-control difference	Treatment-control difference
<i>Eligible for bagrut</i>	0.493	0.015 (-0.032)	0.058 (-0.028)	0.062 (-0.041)
<i>Mean score</i>	65.7	2.00 (-2.26)	4.70 (-2.23)	4.43 (-2.41)
<i>Mean score (weighted)</i>	71.4	1.14 (-2.41)	4.10 (-2.37)	3.74 (-2.43)
<i>Credit units</i>	19.7	0.463 (-0.638)	1.39 (-0.779)	1.88 (-0.731)
<i>Subjects</i>	6.98	0.316 (-0.265)	0.564 (-0.245)	0.673 (-0.232)
<i>Honors-level subjects</i>	2.14	-0.084 (-0.106)	-0.098 (-0.183)	-0.005 (-0.169)
<i>Number of students (schools)</i>				
Control group	7,945 (34)			3,517 (16)
Treatment group	2,318 (17)			845 (6)

Note: Standard errors in parentheses are adjusted for (primary) school-level clustering.

somewhat larger in column 4, due most likely to the much smaller sample size. However, the similarity between the estimates in columns 3 and 4 raises the reliability that they reflect the effect of the school principals' higher compensation and none of it can be attributed to the fact that in some of the treated schools, some teachers also enjoyed a wage increase.

Additional Results: The Effect of Two Years of Treatment

The effect of the program may be larger and estimated more precisely if the sample could be limited to schools that received the treatment longer. This can be examined by limiting the sample to those schools that received the treatment in 1997 as well as in 1998. Here, I replicate the above analysis while limiting the sample to schools that received the treatment in both years. Table 4 shows balancing tests on treatment–control differences in characteristics for this sub-sample. Focusing first on the pre-program

Table 4. *Balancing tests for pre- (1995–1996) and post- (1997–1998) treatment cohorts, two years of treatment*

	Pre-treatment cohorts		Post-treatment cohorts	
	Control group mean	Treatment–control difference	Control group mean	Treatment–control difference
Demographic variables				
<i>Father's years of schooling</i>	11.1	0.218 (–1.23)	11.7	–0.012 (–0.812)
<i>Mother's years of schooling</i>	11.3	0.160 (–1.14)	11.7	0.295 (–0.782)
<i>Number of siblings</i>	3.02	–0.436 (–0.545)	3.13	–0.047 (–0.439)
<i>Gender (male = 1)</i>	0.464	0.013 (–0.057)	0.466	–0.013 (–0.056)
<i>Immigration status</i>	0.097	–0.041 (–0.028)	0.053	–0.033 (–0.033)
<i>Country of origin</i>				
Israel	0.359	0.058 (–0.077)	0.436	0.005 (–0.094)
Asia/Africa	0.394	–0.037 (–0.108)	0.322	0.034 (–0.117)
Europe/USA	0.110	–0.023 (–0.033)	0.124	–0.040 (–0.037)
Ethiopia	0.010	–0.002 (–0.006)	0.035	–0.033 (–0.022)
USSR	0.127	0.004 (–0.043)	0.084	0.033 (–0.046)
<i>Number of students (schools)</i>				
Control group	2,222	–17.0	2,560	–17.0
Treatment group	1,615	–9.00	1,626	–9.00

Note: Standard errors in parentheses are adjusted for school-level clustering.

years, the results suggest that all the difference estimates are not statistically different from zero and that they tend to be very small. For example, the mean of father's years of schooling in the control schools is 11.1 years, with the treatment–control difference only 0.218 (SE = 1.227). The mean of mother's years of schooling is 11.27 years in the control schools, with the treatment–control difference 0.160 (SE = 1.142). Similarly, the control mean of number of siblings is 3.015, with the difference between the two groups very small, -0.436 (SE = 0.545). The balancing tests for the series of dummy variables, indicating the student's ethnic origin, also suggest a very close resemblance between the two groups. Thus, the two pre-treatment cohorts in the treatment and control groups are not significantly different in terms of their observable characteristics.

Columns 3 and 4 in Table 4 show similar balancing tests for the first two cohorts that participated in the program, i.e., students who graduated in 1997 and 1998. The results are very similar to those of the full sample, namely, there seem to be no observable differences in the characteristics of students who graduated during the first two years of the program.

Columns 4 to 6 in Table 5 report the results from cross-section regression of the outcome variables for the sample of schools that received two years of treatment. Comparing first the simple treatment–control mean differences of the 1997–1998 cohorts with those of the two pre-program cohorts (1995–1996) reveals larger improvements. For example, the matriculation rate unconditional mean difference was reduced from -3.2 percent to -0.7 percent, and the test-score gap changed from -2.710 points to 0.232.

Comparing the post-treatment *controlled* differences to the respective pre-treatment differences reveals further evidence of improvements in the educational outcomes of the treated population. For example, the difference in the average score in the matriculation exams declined from -4.096 to -1.816 , while the gap in the number of matriculation subjects declined from -0.873 to -0.505 .

Table 6 shows the results of estimating the difference-in-differences model based on the panel of four-cohort stacked data with school-fixed effects in the sample of schools that received the treatment for two years. The second column presents the difference-in-differences estimates from a specification that did not include the students' demographic variables as controls. Similar to the pattern seen in Table 3, these treatment estimates resemble the simple mean difference-in-differences derived from comparison of the pre- and post-treatment cross-sectional results in Table 5. They are all positive, excluding the effect on the number of subjects studied at honors level, but are not precise enough to be significantly different from zero.

In the third column of Table 6, I present the estimates from a specification that includes the students' demographics as controls. Three of the

Table 5. *Pre- and post-treatment outcome means, two years of treatment*

	1995–1996		1997–1998	
	Control group mean	Treatment–control difference	Control group mean	Treatment–control difference
Outcome				
<i>Eligible for bagnet</i>	0.498	-0.032 (-0.055)	0.551	-0.007 (-0.057)
<i>Mean score</i>	66.5	-2.71 (-3.11)	69.7	0.232 (-3.41)
<i>Mean score (weighted)</i>	72.2	-3.04 (-3.46)	76.9	-1.02 (-3.90)
<i>Credit units</i>	20.3	-2.25 (-1.38)	21.7	-1.52 (-1.42)
<i>Subjects</i>	7.21	-0.730 (-0.405)	7.47	-0.294 (-0.403)
<i>Honors-level subjects</i>	2.15	-0.133 (-0.226)	2.43	-0.299 (-0.247)
<i>Number of students (schools)</i>				
Control group	2,222	(17)	2,560	(17)
Treatment group	1,615	(9)	1,626	(9)

Note: Standard errors in parentheses are adjusted for school-level clustering.

Table 6. DID estimates of effect of school principals' wages on students' outcomes

	Sample: schools with two years of treatment		
	Control group mean	Treatment-control difference	Treatment-control controlled difference
Outcomes			
<i>Eligible for bagrut</i>	0.498	0.055 (-0.032)	0.070 (-0.028)
<i>Mean score</i>	65.7	4.26 (-2.70)	4.99 (-2.37)
<i>Mean score (weighted)</i>	71.6	3.65 (-2.95)	4.52 (-2.55)
<i>Credit units</i>	19.8	1.29 (-0.940)	1.40 (-0.859)
<i>Subjects</i>	6.99	0.511 (-0.309)	0.555 (-0.255)
<i>Honors-level subjects</i>	2.16	-0.113 (-0.208)	-0.104 (-0.206)
<i>Number of students (schools)</i>			
Control group	6,523	(17)	
Treatment group	1,626	(9)	

Note: Standard errors in parentheses are adjusted for school-level clustering.

estimates are significant with *t*-values higher than 2.0. Overall, these estimates are only marginally larger than those in Table 3, column 3. The estimates suggest that the program increased the matriculation eligibility rate by 7.0 percent, the average test score by 4.99 points, the number of matriculation credits by 1.4 and the number of matriculation subjects by 0.55. The largest estimated effect is on the matriculation rate, which showed a 14 percent increase.

VI. Allowing for Treatment Heterogeneity

Heterogeneity of Treatment Effect by School Size

The effect of paying higher wages to headmasters may vary by school size. The larger the school in terms of teachers and students, the more difficult it may be for a school principal to initiate change, monitor implementation of new programs and so on. Table 7 shows evidence of treatment heterogeneity by school size. Columns 1 and 2 show the results when the full sample is used; columns 3 and 4 present the results obtained when the sample includes only the schools that participated in the program for two years. The results do not produce a clear pattern with reference to whether salary increases to school principals are more effective in small versus larger

Table 7. *DID estimates of effect of school principals' wages on students' outcomes with interaction of treatment and school size^a*

	Full sample		Sample after two years of treatment	
	Treatment– control difference	Treatment– size interaction	Treatment– control difference	Treatment– size interaction
Outcome				
<i>Eligible for bagrut</i>	–0.012 (0.046)	0.370 (0.220)	0.012 (0.044)	0.301 (0.204)
<i>Mean score</i>	2.21 (2.42)	8.33 (7.98)	3.05 (2.53)	6.76 (10.5)
<i>Mean score (weighted)</i>	1.98 (2.59)	4.69 (9.23)	3.06 (2.77)	2.06 (12.1)
<i>Credit units</i>	1.60 (0.816)	–5.78 (3.59)	2.00 (0.978)	–6.74 (4.95)
<i>Subjects</i>	0.577 (0.306)	–0.965 (1.29)	0.622 (0.335)	–0.910 (1.48)
<i>Honors-level subjects</i>	0.158 (0.145)	–1.75 (0.965)	0.221 (0.189)	–2.25 (1.12)
<i>Number of students (schools)</i>				
Control group	7,945	(34)	6,523	(26)
Treatment group	2,318	(17)	1,626	(9)

Note: Standard errors in parentheses are adjusted for school-level clustering.

^aSchool size = no. of students in school's twelfth grade/1,000.

schools. For three outcomes, the interaction term between the treatment variable and school size is negative, whereas for three other outcomes it is positive. However, for all six outcomes, the interaction term estimates are not measured precisely enough to draw meaningful conclusions about this issue.

Heterogeneity of Treatment Effect by Students' Socio-economic Background

Another interesting and policy-relevant question is whether heterogeneity in the treatment effect varied by the students' ability or by their socio-economic background. The available data did not allow me to examine treatment heterogeneity by students' ability, but I used the student's father's years of schooling as a proxy measure for the student's socio-economic background. Father's and mother's years of schooling are highly correlated so it does not make much sense if mother's years of schooling are used instead to measure the socio-economic background of students. A natural division of the sample by parental schooling is the cut-off of high-school completion. I therefore stratified the sample by whether father's years of schooling were below 12 even though the two resulting samples were not exactly identical in size.

Table 8 presents the results based on the two samples derived by the father's years of schooling. The first three columns report the results for the lower socio-economic group and the last three columns those for the higher socio-economic group. A comparison of the means of outcomes of students in the two groups clearly reveals that achievements are negatively correlated with socio-economic background. For example, the mean matriculation rate in the lower socio-economic group (29.2) is 18 percentage points lower than that of the higher group (56.7).

A comparison of the treatment-effect estimates does not reveal any clear pattern of differences between the two groups. The effect on the matriculation rate, credit units and number of subjects is higher among students from lower socio-economic background, but it is the opposite with respect to the mean score and number of subjects studied at honors level. However, in both cases it is not possible to reject the hypothesis that the respective estimates for the two groups are not different from each other.

VII. Conclusions

The evidence in this paper suggests that increasing school principals' wages leads to improved students' outcomes. The estimated effects are marginally more pronounced when the analysis is based on schools that participated in the program for two consecutive years. The empirical analysis is based on an Israeli program that raised the salaries of principals in some schools by 50 percent. The raise was unconditional on any performance measures and was extended from year to year almost automatically. This evidence is different from evidence which suggests that raising teachers' salaries unconditionally is not an effective policy in terms of improving students' and schools' outcomes.

My estimates are based on a relatively small sample of treated and control schools. This limitation suggests caution in interpreting the findings, so that it would be wise to seek more evidence before drawing firmer conclusions. Bearing this caveat in mind, the results in this study have some policy implications. While teachers' wages in Israel have slipped relative to other opportunities available to college graduates, simply raising all salaries would be both very expensive and very inefficient. The largest potential impacts of overall increases in salaries for teachers would be felt by attracting a new group of teachers into the profession and retaining practicing teachers who would otherwise leave the profession. But as Ballou and Podgursky (1995) point out, there is no reason to believe that this would necessarily increase the quality of teachers in the short term. Retaining teachers would be beneficial only if high-quality teachers were those retained—but the existing data on teacher labor markets do not indicate that this would be the case. In fact, recent work by Hanushek *et al.* (2004) and Hanushek and

Table 8. *DID estimates of the effect of school principals' wages on students' outcomes in sub-samples by high and low parental schooling*

	Sample: students whose fathers' years of schooling ≤ 11			Sample: students whose fathers' years of schooling ≥ 12		
	Control group mean	Treatment-control difference	Treatment-control controlled difference	Control group mean	Treatment-control difference	Treatment-control controlled difference
Outcomes						
<i>Eligible for bagrut</i>	0.392	0.042 (0.035)	0.071 (0.030)	0.567	0.008 (0.040)	0.051 (0.036)
<i>Mean score</i>	61.0	3.33 (3.54)	4.02 (3.82)	69.1	1.97 (2.18)	5.64 (1.67)
<i>Mean score (weighted)</i>	65.3	2.87 (3.80)	3.83 (4.09)	75.9	1.07 (2.42)	4.77 (1.72)
<i>Credit units</i>	17.2	1.83 (1.00)	2.16 (1.13)	21.5	-0.211 (0.682)	0.272 (0.612)
<i>Subjects</i>	6.44	0.699 (0.380)	0.729 (0.392)	7.37	0.104 (0.272)	0.324 (0.171)
<i>Honors-level subjects</i>	1.64	0.120 (0.184)	0.028 (0.200)	2.50	-0.180 (0.117)	-0.289 (0.199)
<i>Number of students (schools)</i>						
Control group	3,351	(33)		4,594	(34)	
Treatment group	975	(17)		1,343	(17)	

Note: Standard errors in parentheses are adjusted for school-level clustering.

Rivkin (2004) has found that the average effectiveness of those who exit a large urban district is below the average of those who remain. This raises doubts that reducing turnover, given the current composition of entering teachers, should receive high priority. Higher salaries would certainly tend to increase the pool of potential teachers, but the impact of that change on overall teacher quality depends on the ability of principals to choose the best teachers. Existing research demonstrates that principals do in fact know who the better teachers are.¹⁰ While the evidence is not as complete as one might wish, the ability of principals to identify teachers at the top and bottom of the quality distribution almost certainly goes beyond this, particularly if effective tests of student achievement are administered regularly. Furthermore, based on the results presented above, it may be concluded that a first priority should be to pay school principals higher wages because this will not only lead to better school performance, but also motivate more stringent use of their ability to identify good teachers. Therefore, the obvious implication is that principals must be held accountable for the impact of their hiring, retention and other management decisions on students' achievement. Of course, such quality-motivating structures are rare in education, so there is little to build on in the actual structuring of such a notion. Experimenting scientifically with incentive systems and structures, as well as relying on lessons learned from other economic sectors, are necessary to inform this policy context.¹¹

References

- Angrist, J. and Lavy, V. (forthcoming), The Effect of High-stakes High School Achievement Awards: Evidence from Group-randomized Trial, *American Economic Review*.
- Ashenfelter, O. A. (1978), Estimating the Effect of Training Programs on Earnings, *Review of Economics and Statistics* 60, 47–57.
- Ballou, D. and Podgursky, M. (1995), Recruiting Smarter Teachers, *Journal of Human Resources* 30, 326–338.
- Cohen, D. K. and Murnane, R. J. (1985), The Merits of Merit Pay, *Public Interest* 80, 3–30.
- Cohen, D. K. and Murnane, R. J. (1986), Merit Pay and the Evaluation Problem: Understanding Why Most Merit Pay Plans Fail and a Few Survive, *Harvard Educational Review* 56, 1–17.
- Cohn, E. (1996), Methods of Teacher Remuneration: Merit Pay and Career Ladders, in W. E. Becker and W. J. Baumol (eds.), *Assessing Educational Practices: The Contribution of Economics*, MIT Press, Cambridge, MA, 209–238.
- Hanushek, E. A. and Rivkin, S. G. (2004), How to Improve the Supply of High Quality Teachers, in D. Ravitch (ed.), *Brookings Papers on Education Policy*, Brookings Institution, Washington, DC.

¹⁰ See Jacob and Lefgren (2005) who identify total teacher effects as discussed above and relate them to principals' evaluations.

¹¹ See Lavy (2002, forthcoming) for evidence on the effect of group and individual teachers' incentives in high school in Israel.

- Hanushek, E. A., Kain, J. F. and Rivkin, S. G. (1999), Do Higher Salaries Buy Better Teachers?, NBER Working Paper no. 7082, Cambridge, MA.
- Hanushek, E. A., Kain, J. F. and Rivkin, S. G. (2004), Why Public Schools Lose Teachers, *Journal of Human Resources* 39, 326–354.
- Hanushek, E. A., Kain, J. F. and Rivkin, S. G. (2005), Teachers, Schools and Academic Achievement, *Econometrica* 73, 417–458.
- Jacob, B. and Lefgren, L. (2007), Principals as Agents: Subjective Performance Assessment in Education, manuscript, Ford School of Public Policy, University of Michigan.
- Kenny, L. W. (1980), Compensating Differentials in Teachers' Salaries, *Journal of Urban Economics* 7, 198–207.
- Koretz, D. (1996), Using Student Assessments for Educational Accountability, in E. A. Hanushek and D. W. Jorgenson (eds.), *Improving America's Schools: The Role of Incentives*, National Academy Press, Washington, DC, 171–196.
- Lavy, V. (2002), Evaluating the Effect of Teachers' Performance Incentives on Pupils' Achievements, *Journal of Political Economy* 110, 1286–1317.
- Lavy, V. (forthcoming), Performance Pay and Teachers' Effort, Productivity and Grading Ethics, *American Economic Review*.
- Levinson, A. M. (1988), Reexamining Teacher Preferences and Compensating Wages, *Economics of Education Review* 7, 357–364.
- Loeb, S. and Page, M. E. (2000), Examining the Link between Teacher Wages and Student Outcomes: The Importance of Alternative Labor Market Opportunities and Non-pecuniary Variation, *Review of Economics and Statistics* 82, 393–408.
- Ministry of Education (2001), *Bagrut Test Data 2000*, Chief Scientist's Office, Ministry of Education, Jerusalem.
- Murnane, R. J. and Olsen, R. (1989), The Effects of Salaries and Opportunity Costs on Length of Stay in Teaching: Evidence from Michigan, *Review of Economics and Statistics* 71, 347–352.
- Murnane, R. J. and Olsen, R. (1990), The Effects of Salaries and Opportunity Costs on Length of Stay in Teaching: Evidence from North Carolina, *Journal of Human Resources* 25, 106–124.
- Rumberger, R. W. (1987), The Impact of Salary Differentials on Teacher Shortages and Turnover: The Case of Mathematics and Science Teachers, *Economics of Education Review* 6, 389–399.
- The Teaching Commission (2004), *Teaching at Risk: A Call to Action*, <http://www.theteachingcommission.org>.