Does Teacher Training Affect Pupil Learning? Evidence from Matched Comparisons in Jerusalem Public Schools

Joshua D. Angrist, *Massachusetts Institute of Technology and National Bureau of Economic Research*

Victor Lavy, *Hebrew University of Jerusalem*

Most research on the relationship between teacher characteristics and pupil achievement focuses on salaries, experience, and education. The effect of in-service training has received less attention. We estimate the effect of in-service teacher training on achievement in Jerusalem elementary schools using a matched-comparison design. Differences-in-differences, regression, and matching estimates suggest training in secular schools led to an improvement in test scores. The estimates for religious schools are not clear cut, perhaps because training in religious schools started later and was implemented on a smaller scale. Estimates for secular schools suggest teacher training provided a cost-effective means of increasing test scores.

The Jerusalem Schools Authority provided the data and partial funding for this project. Thanks go to Jon Guryan, Xuanhui Ng, and Analia Schlosser for outstanding research assistance, and to the teachers, principals, and supervisors at the schools involved for their cooperation. Special thanks also go to Avi Sela, the project manager in Jerusalem, and his staff for their assistance and support. Alan Krueger and seminar participants at the Massachusetts Institute of Technology labor lunch provided helpful comments.

[*Journal of Labor Economics*, 2001, vol. 19, no. 2]

0734-306X/2001/1902-0004$2.50

I. Introduction

The question of how teacher characteristics affect pupil learning has long been of concern to economists, educators, and parents. The interest in this question is more than academic since a recent study by Black (1999) suggests many parents would be willing to pay for better teacher qualifications if they knew that this would cause their children to learn more. The widespread interest in teacher characteristics has led to numerous studies that attempt to estimate the effect of teachers' characteristics on their pupils' test scores. The characteristics that have been most studied are teachers' educational background, years of teaching experience, and salaries. This research, while plentiful, has failed to produce a consensus regarding the causal impact of teacher characteristics, and the question of whether ostensibly more-qualified teachers produce better students remains open (see, e.g., Hanushek [1986] and Hedges, Laine, and Greenwald [1994] for contrasting assessments of the literature).[1]

While many econometric and statistical studies have focused on the consequences of teachers' general skills as captured by education or experience, a second strand of research has looked at the question of how specific episodes of on-the-job or in-service training affect pupil performance. On-the-job and in-service training may be no less important than the more widely studied education or general experience measures. For example, one survey of research on teacher training in developing countries concludes that "pre-service training is essential to teach subject matter. In-service training is essential to teach teaching skills" (Farrell and Oliveira 1993, p. 15). In spite of the potential importance of in-service training, the literature on the effects of this sort of training is sparse. In a recent study that is similar in spirit to ours, Bressoux (1996) looks at the impact of in-service training on novice teachers in France using a quasi-experimental research design. Bressoux observes that most of the work on this topic to date falls into the category of "process evaluation"

[1] The best-known study of teacher characteristics and other school resources is probably the Coleman et al. report (1966). Other early examples include Hanushek (1971) and Murnane (1975). Recent studies of the effect of teachers' education, experience, and subject-matter knowledge are Ehrenberg and Brewer (1994), Monk (1994), and Goldhaber and Brewer (1997). Behrman, Khan, Ross, and Sabot (1997) looked at teacher qualifications in Pakistan. Hanushek, Rivkin, and Kain (1998) show that teachers have an important effect on student outcomes but suggest that the definition of an effective teacher is complex and context-dependent. Researchers have also studied the relationship between teachers' salaries and their pupils' earnings as adults. See Welch (1966) and, more recently, Card and Krueger (1992).

that ignores outcomes or involves comparisons of different training programs with no untrained control group.[2]

This article contributes to the literature on effects of in-service training by presenting an evaluation of teacher training in Jerusalem elementary schools, with the purpose of estimating the causal effect of the program on pupils' test scores. The training program we study was designed to improve the teaching of language skills and mathematics, and involves pedagogical methods developed in U.S. schools. Most of the existing research on teacher training, like studies of the effects of other school resources, is complicated by the fact that school inputs are not exogenously determined.[3] In an attempt to overcome the methodological difficulties inherent in an evaluation study of this type, our research design exploits the fact that in 1995 a handful of public schools in Jerusalem received a special infusion of funds that were primarily earmarked for teachers' on-the-job training. This program presents an unusual research opportunity because, even though the intervention was not allocated using experimental random assignment, the Jerusalem intervention can be studied with the aid of a matched group of students from schools not subject to the intervention. Considerable information is available on the students enrolled at the affected schools both before and after the intervention began. Similar information is also available for a group of comparison schools in adjacent neighborhoods and elsewhere in the city, so these schools can play the role of a control group.

Our research strategy uses a variety of statistical methods to estimate the causal effect of teachers' on-the-job training on their students' test scores. All of the methods suggest that the training received by teachers who work in the secular branch of the Jerusalem school system led to an improvement in their pupils' test scores. In contrast with the estimates for secular schools, however, the estimates for religious schools are sensitive to the details of model specification. The absence of a clear effect

[2] Perhaps surprisingly, there seems to have been more research on the impact of teacher training in developing countries than in developed countries. It should also be noted that economists have estimated the effects of teacher training in their own discipline. See, e.g., Highsmith (1974) and Schober (1984). In-service training for graduate teaching assistants has been studied as well (see, e.g., Bray and Howard 1980 or Carroll 1980). Recent research on the productivity consequences of on-the-job and in-service training for workers other than teachers includes Bartel (1995) and Krueger and Rouse (1998).

[3] An exception is Dildy (1982), who reports the results of a randomized trial to evaluate the effects of teacher training in Arkansas elementary schools. The potential for bias in naive comparisons was highlighted by Lavy (1995), who noted that the apparent negative correlation between school inputs and pupil achievement in Israel is at least partly due to the fact that measures of socioeconomic disadvantage are used to decide which schools get the most inputs.

in religious schools may be because the intervention there started later and was implemented on a smaller scale.

The most plausible estimates for secular schools suggest effect sizes in the range of .2–.4 standard deviations, which is noteworthy given the modest cost of the intervention. In light of recent critical comments regarding the level of teacher training in subject matter (see, e.g., *Boston Globe* 1998; *New York Times* 1998), it may also be noteworthy that the Jerusalem training emphasized pedagogy rather than subject content (of course, the value of this emphasis might differ at upper grade levels). In an attempt to assess the economic value of the training program studied here, we compare the treatment effect and costs of training to the effect size and costs of alternative school improvement strategies involving reductions in class size and lengthening the school day. This analysis suggests that teacher training may provide a less expensive strategy for raising test scores than reducing class size or adding school hours.

II. The Intervention and the Data

A. Teacher Training in Jerusalem

Beginning in 1995, the 30 Towns intervention led to a major increase in resources for schools in two neighborhoods in North Jerusalem (Neve Yaakov and Pisgat Zeev), as well as in other towns and cities in Israel. These neighborhoods were selected for the program because they form a distinct geographic bloc, constituting a self-contained school district with a high proportion of immigrant pupils and lower achievement levels than the city average. The intervention essentially came in the form of budget increases, though in some cities the money was earmarked for certain uses. In Jerusalem, the 30 Towns money was spent mainly on teacher training, though some resources were used for reorganization and for programs for children considered to have special needs, including immigrants.[4] The extra training was provided in the school on a weekly basis by outside instructors who focused on improving instruction techniques for Hebrew language skills (which we call "reading"), mathematics, and English. The training was given jointly to all the relevant subject teachers in a grade and all teachers had to (and did) participate. We confirmed that these same teachers taught the children in our data set, except for one teacher who retired at the end of the first year of the project. The goals of the intervention were to increase the pass rate from grade to grade, to increase scores on achievement tests, and (more vaguely) to improve the school climate. Our study is limited to the effect of the

[4] The total amount involved was about $960,000 per year for elementary schools. For a more detailed description of the intervention, see the appendix.

program on reading and math achievement since there is no test-score data for English.[5]

Table 1 summarizes the program impact on school inputs for our sample of treatment and control schools using data from the Jerusalem Schools Authority (JSA) (described in greater detail below). Public schools in Jerusalem and elsewhere in Israel are separated into religious and secular systems, so we discuss the effect of the program on the two types of schools separately. In Jerusalem, a total of 10 elementary schools were affected, seven secular and three religious, though we drop one of the religious schools from the analysis since it opened after the intervention began. Panel A shows the data for secular schools (seven treated, six control), and panel B shows the corresponding data for religious schools (two treated, five control).[6] The information is shown for three dates: 1994, which is before the training program started; 1995, which is the year training began; and 1996, which is a year in which training inputs remained at the new higher level in treated schools. Training is measured as the total hours of instruction received by all teachers in each school per week, on average over the course of the relevant school year, and is recorded separately for reading and math.[7]

Panel A of the table shows a marked increase in hours of training provided to teachers in the secular treated schools between 1994 and 1995 and again between 1995 and 1996. For example, in 1994, treatment schools received an average of 1.2 training hours per week in math. This increased to 6.7 hours of training in 1995 and to 12.5 hours of training in 1996. In contrast, training hours changed little in the control schools. Thus, by 1996, the treatment schools received 10.5 more math training hours per week than the control schools. Similarly, the treatment schools actually received fewer hours of training in reading in 1994, but by 1996 the gap between treatment and control schools was 7.7 hours in favor of treated schools.

[5] The training interventions varied somewhat from school to school, but all involved a mixture of similar counseling and feedback sessions for teachers, changes in the organization of class time, and the use of instructional aids. There was generally an emphasis on developing strategies for helping children who are doing poorly, including immigrants and those with learning disabilities. The math teachers received training based on a modern "humanistic mathematics" philosophy of teaching (see White 1987, 1993). The reading teachers received training based on the "individualized instruction" approach to schooling. See, e.g., Bishop (1971) for a description, and Bangert, Kulik, and Kulik (1983) and Romberg (1985) for research summaries.

[6] The selection of control schools is discussed in the next section.

[7] The total school allocation of training hours was divided among the grades in the school. On average, each school had about 450 pupils and 12 classes (including all grades). All teachers in a given grade received the same amount of training.

Table 1
Inputs and School Resources

Inputs	1994 (Before Training) Treatment (1)	Control (2)	Difference (3)	1995 (Training Begins) Treatment (4)	Control (5)	Difference (6)	1996 (Training Continues) Treatment (7)	Control (8)	Difference (9)
A. Secular schools:*									
Teacher training, math (hours per week)	1.17	.50	.67	6.67	1.17	5.50	12.50	2.00	10.50
	(.75)	(.34)	(.83)	(2.11)	(.83)	(2.27)	(3.10)	(1.00)	(3.37)
Teacher training, reading (hours per week)	.00	5.67	−5.67	11.00	4.17	6.83	16.00	8.33	7.67
	(.00)	(1.91)	(1.89)	(2.29)	(1.60)	(2.78)	(2.45)	(3.78)	(2.91)
Expenditure per pupil (current NIS)	515.6	580.5	−64.9	759.6	568.4	191.2	780.5	663.9	116.6
	(49.0)	(47.7)	(68.5)	(50.3)	(40.2)	(63.7)	(55.5)	(35.8)	(64.8)
Class size	30.6	32.3	−1.70	30.5	32.5	−2.00	28.62	31.69	−3.07
	(2.59)	(2.68)	(3.79)	(2.89)	(2.57)	(3.85)	(2.78)	(2.96)	(4.06)
Total instruction hours per pupil	1.44	1.31	.13	1.48	1.44	.04	1.54	1.48	.06
	(.12)	(.10)	(.14)	(.10)	(.10)	(.13)	(.10)	(.10)	(.14)
Instruction hours per pupil for special projects	.37	.33	.04	.23	.24	−.01	.44	.45	−.01
	(.04)	(.06)	(.07)	(.05)	(.05)	(.09)	(.05)	(.07)	(.09)

B. Religious schools:[†]									
Teacher training, math (hours per week)	3.5	...	...	.0	...	...	7.0	...	...
	(3.5)			(.0)			(.0)		
Teacher training, reading (hours per week)	3.5	...	...	.0	...	...	7.0	...	...
	(3.5)			(.0)			(.0)		
Expenditure per pupil (current NIS)	568.5	718.9	−150.4	715.9	710.8	5.1	903.5	984.7	−81.2
	(134.1)	(87.3)	(160.0)	(291.1)	(206.7)	(364.3)	(350.6)	(128.4)	(406.0)
Class size	33.7	26.4	7.3	30.3	29.6	.7	28.3	27.2	1.1
	(4.8)	(.9)	(4.8)	(2.8)	(1.3)	(3.3)	(3.6)	(1.1)	(4.5)
Total instruction hours per pupil	1.44	1.61	−.17	1.72	1.65	.07	1.83	1.80	.03
	(.20)	(.07)	(.21)	(.15)	(.10)	(.18)	(.38)	(.05)	(.33)
Instruction hours per pupil for special projects	.41	.44	−.03	.36	.32	.04	.60	.56	.04
	(.05)	(.04)	(.08)	(.09)	(.12)	(.10)	(.20)	(.06)	(.20)

NOTE.—Standard errors are reported in parentheses. Years refer to school years; for example, 1994 is the school year beginning in September 1993. NIS = new Israeli shekels.

* There are seven treated and six control schools. Training in secular schools began in January 1995, during the 1995 school year.

[†] There are two treated and five control schools. Training in religious schools began in September 1995, at the beginning of the 1996 school year.

Table 2
Information Availability

	June 1994	January 1995	June 1995	September 1995	June 1996
A. Pupil status:					
Secular	Baseline	Treatment begins	First follow-up	...	Second follow-up
Religious	Baseline	...	Pretreatment	Treatment begins	First follow-up
B. Information:					
Test scores	JSA feedback tests in math and Hebrew	...	JSA feedback tests in math and Hebrew	...	Special tests in math; JSA middle school assessment tests in Hebrew
Control variables	Pupil and school characteristics	...	...	...	Pupil and school characteristics
C. Grade level	Fourth graders	...	Fifth graders	...	Sixth graders

NOTE.—JSA = Jerusalem Schools Authority.

After the data on training, the next row of the table shows expenditure on special projects per pupil by type of school (in nominal shekels), including the additional 30 Towns money but excluding regular teachers' salaries. The changes in expenditure reflect the additional expenditure on training. In contrast with the data on training and expenditure, instruction hours per pupil (which reflects the number of teachers and their work hours) and class size were essentially unchanged in both treated and control schools.

Panel B reports the data on inputs for religious schools, though this analysis is limited by the fact that we were unable to obtain data on teacher training in the religious control schools. The data on training inputs for religious schools actually show a decline in training hours between 1994 and 1995, with a modest increase in 1996. This accords with the fact that in religious schools the 30 Towns money was not used until September of 1995, that is, in the 1996 school year (see Angrist and Lavy 1998). Thus, for religious schools, the 1995 school year is a pretreatment year. In addition to starting later, the intervention was also less intense in the religious schools than in the secular schools. This can be seen in the expenditure data, which show no differential increase in the religious treated schools.

B. Test Score Data and the Choice of Control Schools

Table 2 summarizes information on test scores and other variables available in each of three years: in 1994, before the intervention began; in 1995, when the intervention first started; and in 1996. Although the intervention was school-wide, the only pupils for whom we have data on test scores both before and after the intervention began is the cohort of fourth graders enrolled in the 1993–94 school year. Test score data are available for these children when they were finishing fourth grade in June 1994, as fifth graders in June 1995, and as sixth graders in June 1996. The 1994 and 1995 data were collected in a routine JSA testing program in which all schools in the treated area, as well as 20 other schools, participated. The tests were based on the core material taught during the year in each grade. All students, including special education and immigrant pupils, were required to take the exams, which were for general assessment purposes only and did not determine grade advancement. The study is limited to the 1994 fourth graders in these schools because test-score data for 1996 are available only for sixth graders.

The training effort in secular schools began in January 1995, so the June 1995 test scores provide an early indicator of program effectiveness in these schools. But training in religious schools did not begin until September 1995, so the June 1995 data provide an additional pretreatment observation for religious schools. To obtain information on test scores in 1996, which provides follow-up information 1.5 years after training began in secular schools and 1 year after training began in religious schools, we used data from two sources. The Hebrew test scores come from a middle-school exam given for planning purposes to all sixth graders in Jerusalem. Since the placement test does not cover math, at our request, special math tests were developed and administered in treatment and control schools for the purposes of evaluating 30 Towns. These tests were designed to be similar in style to the 1996 test in reading. To minimize differences between the tests across years, we analyze standardized scores with mean zero and unit standard deviation (except in the presentation of descriptive statistics in table 3).

The evaluation strategy compares the population of fourth-grade pupils who were enrolled in treated schools in 1994 with the population of fourth graders who attended a sample of control schools. The control schools were selected for comparability and for practical reasons related to data collection. First, the control schools had to be in the group of 20 non-treated schools that were tested by the JSA in 1994 and 1995. Second, to maximize comparability and to encourage the cooperation of school principals with our 1996 testing effort, we chose control schools that report to the same district supervisors as the treatment schools. A total of 11 out of the 20 candidate control schools met the criteria for inclusion in the study—six secular schools and five religious schools.[8]

[8] In addition to test-score data, the JSA also provided information on the characteristics of pupils and schools in our sample in 1994. School characteristics were

Table 3
Basic Test Score Data

	Reading				Math			
	Treated Schools		Control Schools		Treated Schools		Control Schools	
	Mean (1)	N (2)	Mean (3)	N (4)	Mean (5)	N (6)	Mean (7)	N (8)
A. Secular schools:								
1994:								
All	60.48	406	73.18	405	65.18	428	71.20	420
	(17.22)		(15.68)		(17.27)		(16.52)	
Tested in 1995	60.33	364	75.82	284	65.24	387	70.33	307
	(17.08)		(15.08)		(17.14)		(16.02)	
Tested in 1996	59.38	301	73.19	290	64.79	372	71.15	262
	(17.17)		(15.10)		(17.07)		(16.74)	
1995:								
All	76.49	449	79.16	347	76.25	456	77.48	354
	(15.35)		(12.87)		(19.32)		(16.72)	
Tested in 1994	77.02	364	80.44	284	75.95	387	78.20	307
	(14.83)		(11.64)		(19.07)		(16.40)	
1996:								
All	66.39	395	68.68	357	75.88	452	73.10	307
	(17.72)		(16.29)		(18.36)		(19.37)	
Tested in 1994	66.69	301	68.87	290	75.51	372	73.64	262
	(17.05)		(16.43)		(18.37)		(19.40)	
Number of schools	7		6		7		6	
B. Religious schools:								
1994:								
All	73.59	85	85.20	184	74.29	114	73.78	183
	(21.94)		(12.41)		(17.04)		(16.36)	
Tested in 1995	78.24	67	86.59	167	75.52	90	74.12	165
	(19.35)		(11.12)		(16.79)		(16.10)	
Tested in 1996	...	...	...	...	76.64	86	73.98	110
					(15.29)		(15.42)	
1995:								
All	71.77	111	83.33	200	67.07	103	82.01	187
	(17.38)		(13.18)		(16.81)		(15.59)	
Tested in 1994	71.94	67	85.84	167	68.11	90	82.34	165
	(16.51)		(9.74)		(68.58)		(14.75)	
1996:								
All	...	...	...	...	62.14	101	73.39	127
					(26.89)		(23.18)	
Tested in 1994	...	...	...	...	62.34	86	74.15	110
					(25.97)		(22.71)	
Number of schools	2		5		2		5	

Note.—Standard deviations are reported in parentheses. The statistics presented in the table are based on the distribution of raw scores.

In June 1994, 406 fourth-grade pupils took the reading test in the secular treatment schools, and 405 pupils took the reading test in the secular control schools. The corresponding figures for the math test are 428 treated test takers and 420 control test takers. This can be seen in table

assigned on the basis of school identifiers at baseline. Two schools later merged, another later split, and a new school opened, but there was no crossover between treatment and control groups.

3, which reports average test scores and sample sizes in reading and math by treatment and control status for 1994–96. About 90% of all pupils who were enrolled in the treated and control schools in 1994 took the tests. Of the original 406 pupils tested in reading in 1994 in treated schools, 364 were tested in 1995 and 301 were tested in 1996. Of the original 405 tested in reading in 1994 in control schools, 284 were tested in 1995 and 290 were tested in 1996. The sample sizes for math tests show a similar pattern of attrition. Note also that some pupils who were not tested in 1994 were tested in 1995 and 1996 since the tests covered whoever was enrolled at the time. The number of those tested in 1995 and 1996 who were also tested in earlier years is reported in the table, along with average scores for subsamples of pupils where we have repeated observations. A similar analysis for religious schools appears in panel B of table 3.

The proportion of pupils tested in 1994 who were also tested in 1996 ranges from .62 for math tests in control schools to .87 for math tests in treatment schools. The higher attrition rates in the control group are potentially a cause for concern. It should be noted, however, that most of the nonparticipation in follow-up waves resulted from entire classes not taking the test rather than from individual pupils missing the test. Attrition at the class level probably generates less bias than attrition by pupils. This claim is supported by the fact that in most cases, average scores in the entire 1994 cross section are similar to those in the longitudinal samples with information on scores in 1995 and 1996. Similarly, average scores in the 1996 cross section are similar to those in the longitudinal samples with scores in 1995 and 1994. The scores for different years and samples are shown in columns 1, 3, 5, and 7 of table 3.

C. Pupil Characteristics in Treatment and Control Schools

Pupils in treated and control schools are similar along many dimensions, but pupils in treated schools have less-educated parents. This can be seen in table 4, which compares the characteristics of pupils in the treatment and control schools in 1994. The difference in parents' education is 1.7 years in religious schools and almost 1.5 years for fathers' education in secular schools. This suggests that it may be important to control for pupils' family background when estimating the effect of teacher training. We use a variety of strategies for doing this: differencing test scores to eliminate time-invariant characteristics, controlling for pupils' characteristics using regression, and using regression and matching methods to control for 1994 scores.

III. Evaluation Strategies

The ideal evaluation strategy for the 30 Towns intervention (or any similar intervention) would involve the random assignment of pupils to

Table 4
Pupil Characteristics in 1994

Characteristics	Secular Schools Treatment (1)	Secular Schools Control (2)	Secular Schools Differences (3)	Religious Schools Treatment (4)	Religious Schools Control (5)	Religious Schools Differences (6)
Percent male	.520	.533	−.013	.53	.52	.01
	(.024)	(.024)	(.026)	(.024)	(.024)	(.02)
Percent immigrants	.133	.095	.038	.15	.14	.01
	(.016)	(.014)	(.022)	(.03)	(.03)	(.06)
Family size	4.89	4.85	.04	5.44	6.34	−.90
	(.057)	(.055)	(.08)	(.13)	(.15)	(.20)
Percent married	.84	.84	−.00	.87	.87	−.00
	(.018)	(.018)	(.01)	(.03)	(.03)	(.04)
Father's education	12.36	13.83	−1.47	12.12	13.82	−1.70
	(.146)	(.179)	(.23)	(.28)	(.32)	(.43)
Mother's education	12.47	13.95	−.148	12.26	13.77	−1.51
	(.129)	(.154)	(.20)	(.25)	(.25)	(.36)
Child's age	10.16	10.10	.06	10.14	10.14	.01
	(.019)	(0.15)	(.02)	(.04)	(.03)	(.05)
Number of observations	428	420		114	183	

Note.— Standard errors are reported in parentheses. The sample in this table includes all the pupils who were tested in 1994.

treatment and control groups, with those in the treatment group being taught by teachers who had received training. Random assignment would insure that pupils in the control group are indeed comparable to pupils in the treatment group, so that any difference between pupils in the two groups could be confidently attributed to the intervention. In the absence of random assignment, statistical methods must be used to control for differences between pupils in treated and nontreated schools. We use a variety of models for this purpose.

One simple model that has been used in numerous evaluation studies (see, e.g., Ashenfelter 1978; Ashenfelter and Card 1985) is based on the assumption that any differences between pupils in the treatment and control groups are fixed over time. In this case, repeated observations on the same pupils can be used to make the treatment and control groups comparable. Let D_{it} be a dummy variable that indicates treatment status and let $Y_{it}(0)$ denote the potential test score of any pupil if he or she were not exposed to the treatment. The fixed-effects model says that in the absence of the intervention, the potential test score of pupil i at time t can be written as follows:

$$Y_{it}(0) = \phi_i + \delta_t + \epsilon_{it}, \tag{1a}$$

where D_{it} is assumed to be independent of the time-varying component, ϵ_{it}, although it may be correlated with the pupil-specific intercept, ϕ_i. The term δ_t is a period effect common to all pupils. Thus, while pupils in the treatment group may have lower scores than pupils in the control group, this difference is assumed to be due to differences in family background or neighborhood characteristics that can be viewed as permanent and that have time-invariant effects on test scores.

We begin with a model that also has a constant treatment effect, so treated pupils are assumed to have test scores equal to $Y_{it}(0)$ plus the treatment effect:

$$Y_{it}(1) = Y_{it}(0) + \alpha. \tag{1b}$$

Equations (1a) and (1b) can therefore be used to write the observed test score of pupil i at time t as

$$Y_{it} = Y_{it}(0)(1 - D_{it}) + Y_{it}(1)D_{it} = \phi_i + \delta_t + \alpha D_{it} + \epsilon_{it}, \tag{2}$$

where ϵ_{it} is the error term in (1a). In this model, posttreatment comparisons of test scores by treatment status do not estimate the causal effect of treatment because the time-invariant characteristics of pupils in the treatment and control groups differ. Formally, this means that $E(\phi_i \mid D_{it} = 1) \neq E(\phi_i \mid D_{it} = 0)$. However, the assumptions of the model imply that the change in test scores in the treatment group can be compared to the change in test scores in the control group. Let $t = a$ denote posttreatment scores ("after") and let $t = b$ denote pretreatment test scores ("before"). Note that $D_{ib} = 0$ for both treatment and control pupils. Then we have

$$E(Y_{ia} - Y_{ib} \mid D_{ia} = 1) - E(Y_{ia} - Y_{ib} \mid D_{ia} = 0) = \alpha. \tag{3}$$

The sample analog of equation (3) is a differences-in-differences estimate of the program effect because it contrasts the change in test scores between treatment and control pupils.

The most important difference between pupils in treatment and control schools is that pupils in the treatment schools have lower pretreatment test scores, on average, than pupils in the control schools. The differences-in-differences method controls for these pretreatment differences by subtracting them from the posttreatment difference in scores. To see this, note that equation (3) can be rearranged to read

$$(E[Y_{ia}|D_{ia} = 1] - E[Y_{ia}|D_{ia} = 0])$$
$$-(E[Y_{ib}|D_{ia} = 1] - E[Y_{ib}|D_{ia} = 0]) = \alpha.$$

The term $(E[Y_{ib}|D_{ia} = 1] - E[Y_{ib}|D_{ia} = 0])$ is the pretreatment difference in scores by treatment status.

We also explore alternative strategies that control directly for pretreatment score differences between the treatment and control groups using regression or matching. The regression approach to controlling for pretreatment score differences (and other covariates) is based on a model where lagged test scores determine future test scores and treatment is independent of potential scores conditional on lagged scores and other covariates, X_i. Formally, the model is

$$Y_{ia}(0) = X_i'\beta + \gamma Y_{ib} + \delta_a + \epsilon_{ia}, \tag{4}$$

where ϵ_{ia} is assumed to be independent of D_{it}. Maintaining the constant treatment effects assumption, we have

$$Y_{ia} = X_i'\beta + \gamma Y_{ib} + \delta_a + \alpha D_{ia} + \epsilon_{ia}, \tag{5}$$

This amounts to replacing ϕ_i in (2) with $X_i'\beta + \gamma Y_{ib}$. Estimates of α are constructed by regressing posttreatment test scores on a constant, pretreatment scores, X_i, and D_{ia}. It is important to emphasize that the causal interpretation of α in (5) turns on the assumption that once X_i and pretreatment scores are held constant, the only reason for a posttreatment difference between treated and control pupils is, in fact, the treatment.

Equation (5) relies on a linear model to control for the covariates X_i and Y_{ib}. In addition to regression estimates, we also use a matching strategy to nonparametrically control for pretreatment score differences. This is accomplished by dividing the pretreatment (1994) test score distribution into four equal-sized groups (quartiles). We then compare the posttreatment (1996) scores of treatment and control pupils in each group. The final step in this method is to produce a single treatment-control contrast from the four groups by averaging across groups using the number of treated pupils in each group as weights. Assuming that the treatment is independent of potential outcomes conditional on past test scores, this matching strategy produces an estimate of the average "effect of treatment on the treated" in a model with heterogeneous treatment effects (see, e.g., Rubin 1977; Card and Sullivan 1988; Angrist 1998).

Finally, note that the evaluation strategies discussed in this section involve matching treatment and control pupils and not treatment and control schools. This is because pupils are the unit of observation; we would like pupils in the control group to be comparable to pupils in the treatment group. For some of the regression estimates, however, we limit our sample to treatment and control schools where school-level average test scores are similar in the two groups. This is useful if there is something special about the treatment schools ("random school effects") that cannot be captured by matching pupils on personal characteristics and previous test scores. The cost of school matching is that information on many pupils is discarded.

IV. Empirical Results

A. Differences-in-Differences Estimates

Table 5 presents differences-in-differences estimates of the effect of the intervention. The row labeled "Control differences" is an estimate of the

Table 5
Difference-in-Differences Estimates of Training Effects

	Reading			Math		
Estimate	1995–94 (1)	1996–95 (2)	1996–94 (3)	1995–94 (4)	1996–95 (5)	1996–94 (6)
A. Secular schools:						
Control differences	−.128	.115	−.061	.045	−.110	−.052
	(.055)	(.055)	(.057)	(.055)	(.057)	(.058)
	[.196]	[.119]	[.163]	[.184]	[.141]	[.147]
Differences-in-differences	.602	−.021	.616	.173	.249	.460
	(.073)	(.073)	(.080)	(.073)	(.072)	(.075)
	[.264]	[.162]	[.230]	[.251]	[.186]	[.203]
N	648	624	591	694	549	634
B. Religious schools:						
Control differences	−.343	...	...	.054	−.241	−.192
	(.071)			(.066)	(.088)	(.097)
	[.173]			[.133]	[.116]	[.208]
Differences-in-differences	−.490	...	...	−.867	.246	−.719
	(.133)			(.110)	(.137)	(.147)
	[.312]			[.236]	[.193]	[.370]
N	234	...	...	255	187	196

Note.—Control differences are the change in test scores for pupils in control schools. Differences - in-differences are the difference in changes between treatment and control schools. Conventional standard errors reported in parenthesis. Standard errors corrected for within-school correlation are reported in brackets. Standardized scores were used for this analysis.

change in test scores for pupils in control schools, that is, the sample analog of $E(Y_{ia} - Y_{ib} \mid D_{ia} = 0)$. The row labeled "Difference-in-differences" is the sample analog of $E(Y_{ia} - Y_{ib} \mid D_{ia} = 1) - E(Y_{ia} - Y_{ib} \mid D_{ia} = 0)$. Test score growth is measured between 1995 and 1994, between 1996 and 1995, and between 1996 and 1994. The estimates in each column of the table were computed by regressing the change in scores on a constant and a dummy for pupils in the treatment group. Two sets of standard errors are presented: those based on the usual regression assumptions and standard errors that allow for within-school correlation in test scores using the formula in Moulton (1986).

The results for secular schools show no significant change in test scores in the control schools between 1994 and 1996. For example, reading scores fell by .06 and math scores fell by .05, but neither of these changes is significantly different from zero. In contrast, test scores increased considerably and significantly in the treatment schools, both in absolute terms and relative to the control schools. Thus, the differences-in-differences estimate based on 1994-to-1996 changes is .62 for reading scores and .46 for math scores. This implies the program increased test scores by roughly half a standard deviation for both reading and math. The estimates are smaller when 1994-to-1995 changes are used, especially for math scores. However, math scores in treatment schools increased between 1994 and 1995 and between 1995 and 1996, while reading scores increased only between 1994 and 1995.

The differences-in-differences estimates for secular schools are significant whether conventional standard errors or Moulton standard errors are used to measure precision. It is important to note, however, that the Moulton standard errors are 2–3 times larger than the conventional standard errors. This reflects the fact that, even though the within-school correlation between pupils' test scores is not very large (on the order of .1), the model being estimated has features that cause a small amount of within-group correlation to have a large effect on regression standard errors. First, the regressor (treatment status) is fixed within groups. Second, there are not many groups (schools), so the group size is large relative to the sample. Moulton (1986) shows that when the regressor of interest is fixed within groups and the groups are large, conventional standard errors may seriously overestimate the precision of regression estimates.

Estimates for religious schools are reported in panel B of table 5. The data for religious schools do not include reading scores for 1996.[9] Between 1994 and 1995, test scores in the religious treated schools fell sharply, both in absolute terms and relative to control schools. This can be seen in columns 1 and 4 of the table. As noted earlier, the treatment in the religious schools started only in September of 1995 (i.e., the beginning of the 1996 school year), so the 1994–95 score decline in treated schools is probably unrelated to the intervention. A consequence of the decline in pretreatment years is that the differences-in-differences estimates of effects on math scores are positive when 1995 is taken as the base year and negative when 1994 is taken as the base year (see cols. 5 and 6). Sensitivity to the choice of base year implies that the assumptions underlying the differences-in-differences strategy are not satisfied (Ashenfelter and Card 1985). In fact, the results for religious schools seem generally less reliable and harder to interpret than those for secular schools. The treated religious sample includes fewer than 100 pupils, from only two schools. Moreover, as noted earlier, we do not have data on inputs in religious control schools,

[9] Since most religious schools combine primary and middle grades, they do not use the reading test that is used in secular schools for placing students in the transition from primary to middle school.

while the data on treated schools actually show a decline in teacher training between 1994 and 1995.

B. Regression Estimates

Comparisons of means by treatment status show that pupils in treated schools had lower test scores than pupils in control schools in 1994, before treatment began. Thus, pupils in the two sets of schools are not directly comparable. The differences-in-differences strategy is motivated by a model where the difference between pupils in treatment and control schools is captured by the fixed effect in equation (1a), ϕ_i. Differences-in-differences estimates do not have a causal interpretation, however, if the lower pretreatment scores in treatment schools are due to temporarily low scores as opposed to permanent differences. In such cases, test-score growth in the treatment schools will tend to be larger than test-score growth in the control schools regardless of the program effect, generating spurious positive estimates of treatment effects.[10]

To avoid this sort of bias, the regression approach discussed in Section III replaces the fixed-effects assumption with the assumption that pupils in treatment and control schools are comparable conditional on past test scores and other observed covariates. The resulting estimates for secular schools are reported in table 6. The first column shows results from a specification that includes only the treatment dummy as an explanatory variable. The estimates reported in column 2 are from a model where a limited set of student characteristics (year of birth, sex, immigrant status) were added as controls. The estimates in column 3 are from a model where additional controls (parental schooling, family status, and dummy variables for continent of origin) were added to the list of regressors. Column 4 shows the results of adding lagged test scores, and column 5 shows the results from a model where lagged test scores are the only control variable. Columns 1–5 are for reading scores, and columns 6–10 repeat this sequence for math scores. As before, conventional standard errors are reported in parentheses, and standard errors corrected for within-school correlation are reported in brackets.

The regression results are reported in separate rows for test scores in 1994, 1995, and 1996. The 1994 results are shown because it is of interest to know whether the pretreatment differences between pupils in treatment and control schools are explained by differences in observed covariates. The raw difference in 1994 reading scores by treatment status is $-.74$, and the regression-adjusted difference in column 3 is $-.563$. The raw difference in 1994 math scores is $-.37$, and the regression-adjusted difference in column 8 is $-.28$. Thus, most of the difference in 1994 scores

[10] See, e.g., Ashenfelter and Card (1985).

Table 6
Regression Estimates of Training Effects

Dependent Variable	Reading: No Control (1)	Reading: Basic Control (2)	Reading: Extended Control (3)	Reading: Extended and 1994 Score (4)	Reading: 1994 Score Only (5)	Math: No Control (6)	Math: Basic Control (7)	Math: Extended Control (8)	Math: Extended and 1994 Score (9)	Math: 1994 Score Only (10)
A. All secular schools:										
1994 score	−.744	−.709	−.563	...	...	−.370	−.362	−.284	...	...
	(.072)	(.072)	(.074)			(.080)	(.079)	(.086)		
	[.248]	[.242]	[.204]			[.150]	[.150]	[.162]		
1995 score	−.231	−.195	−.049	.221	.173	−.124	−.091	−.002	.057	.034
	(.072)	(.071)	(.076)	(.076)	(.071)	(.076)	(.075)	(.088)	(.076)	(.065)
	[.220]	[.197]	[.201]	[.204]	[.217]	[.243]	[.222]	[.256]	[.243]	[.238]
1996 score	−.128	−.116	.069	.315	.251	.089	.082	.138	.260	.262
	(.081)	(.082)	(.084)	(.082)	(.078)	(.072)	(.072)	(.076)	(.067)	(.063)
	[.144]	[.139]	[.095]	[.130]	[.152]	[.158]	[.151]	[.145]	[.146]	[.165]
B. Secular schools, matched sample:										
1994 score	−.468	−.399	−.373	...	...	−.086	−.083	.035	...	...
	(.087)	(.088)	(.092)			(.097)	(.096)	(.105)		
	[.321]	[.299]	[.275]			[.122]	[.122]	[.153]		
1995 score	−.411	−.341	−.284	−.045	−.102	−.016	−.012	.053	.030	.025
	(.096)	(.096)	(.103)	(.096)	(.087)	(.084)	(.083)	(.096)	(.082)	(.072)
	[.215]	[.178]	[.152]	[.139]	[.175]	[.284]	[.298]	[.310]	[.290]	[.297]
1996 score	−.063	−.063	.008	.178	.173	.147	.148	.241	.228	.182
	(.098)	(.100)	(.102)	(.096)	(.091)	(.085)	(.084)	(.091)	(.082)	(.075)
	[.192]	[.183]	[.111]	[.162]	[.206]	[.195]	[.189]	[.188]	[.200]	[.219]

Note.—Conventional standard errors are reported in parentheses. Standard errors corrected for within-school correlation are reported in square brackets. Standardized scores were used in this analysis.

between treatment and control pupils remains even after accounting for treatment-control differences in the observed covariates.

Both 1995 and 1996 are posttreatment years for secular schools. The estimated treated effects on 1995 and 1996 scores in models that do not control for pretreatment test scores, reported in columns 1–3, are negative. These negative estimates are not surprising since pupils in treated schools had lower test scores before the treatment began, and since this difference is not accounted for by the included covariates. Controlling for 1994 scores, however, the treatment effects become positive in both 1995 and 1996, and at least marginally significant in 1996, whether or not covariates other than the 1994 scores are included. For example, the effect of the treatment on 1996 reading scores is .32 with a (corrected) standard error of .13 in a model with lagged scores and the extended set of control variables (reported in col. 4). Dropping covariates reduces the estimate to .25. The estimate for math scores controlling for lagged test scores and the full set of covariates is .26 with a standard error of .15. For both math and reading scores, this is about half as large as the corresponding differences-in-differences estimate, suggesting that the differences-in-differences estimates are biased upward.

In addition to results for the full sample of secular schools, we also computed regression results for a matched sample with similar 1994 school-level average scores in treatment and control schools. As noted earlier, this is a nonparametric procedure for making the treatment and control schools as comparable as possible. An attractive feature of this approach is that it reduces the likelihood of bias from school effects that are correlated with treatment status. The matched sample was constructed by selecting groups or pairs of treatment and control schools where the difference in average math scores in 1994 is less than .1 (in standard deviation units). Treatment schools for which no control school could be found with a comparable average score were discarded and vice versa. The matched subsample includes five treatment and three control schools. The results of analyzing this subsample are reported in panel B of table 6. Columns 6–8 of the table show that matching effectively eliminates treatment-control differences in 1994 math scores, though not in reading scores (we were not able to match schools very closely on 1994 reading scores and therefore use the same matched sample for analysis of both math and reading scores).

Although they are less precise, the results for math scores in the matched subsample are generally similar to those in the full sample. An important difference, however, is that since the schools have already been chosen to have similar math scores in 1994, the estimated effects on 1996 math scores are positive even without controlling for 1994 scores. For example, the estimate in column 8, from a model with the extended set of covariates but no lagged test score, is .241 with a corrected standard error of .19.

Table 7
Regression Estimates for Religious Schools

	Math					
Dependent Variable	No Control (1)	Basic Control (2)	Extended Control (3)	Extended and 1994 Score (4)	Extended and 1995 Score (5)	Extended and 1994 + 1995 Scores (6)
1994 score	.171	.170	.357	...	...	...
	(.131)	(.134)	(.152)			
	[.271]	[.261]	[.178]			
1995 score	−.732	−.729	−.542	...	...	...
	(.122)	(.124)	(.145)			
	[.181]	[.166]	[.110]			
1996 score	−.478	−.492	−.304	−.696	.190	−.059
	(.167)	(.169)	(.213)	(.187)	(.177)	(.182)
	[.206]	[.212]	[.300]	[.324]	[.238]	[.263]

NOTE.—Conventional standard errors are reported in parentheses. Standard errors corrected for within-school correlation are reported in square brackets. Standardized scores were used in this analysis.

For reading scores, the results in the matched sample show smaller treatment gaps in 1994 than in the full sample, but the difference by treatment status for 1994 reading scores is not completely eliminated. Controlling for 1994 reading scores, the estimated effect on 1996 reading scores is positive as before, though not significantly different from zero and smaller than in the full sample.

The last set of regression estimates, for religious schools, is reported in table 7. These results are for math scores only, since there are no reading scores in 1996 (the only posttreatment year in religious schools). Columns 1–3 show that test scores in treated schools were higher in 1994 but much lower in 1995, regardless of which covariates are included in the regression. In contrast with the estimates for secular schools, adding the 1994 score to the equation for 1996 scores actually makes the estimated treatment effect more negative. However, models that include the 1995 score as a control variable, the results of which are reported in column 5, lead to a positive treatment effect (though not significantly different from zero). This sensitivity to the choice of lagged control variable is similar to the sensitivity to the choice of base year observed in the differences-in-differences estimates for religious schools. Finally, column 6 shows that adding both pretreatment scores to the regression for religious schools leads to a very small negative estimate that is not significantly different from zero. Since religious treatment and control schools clearly differed in both pretreatment years, this last estimate seems most credible. The absence of an effect in religious schools may be due to the fact that the training program began later than in the secular schools and because the scale of the intervention in religious schools was smaller.

C. Additional Matching Estimates

As a final check on the estimates for secular schools, we matched individual pupils on the basis of their 1994 test scores, instead of school averages as in panel B of table 6. An advantage of pupil matching over school matching is that for both reading and math it is possible to find pupils with similar 1994 scores in all treatment and control schools even though their school averages may differ. The pupil-matching strategy was implemented by dividing the distribution of 1994 test scores into quartiles, comparing treatment and control scores in each quartile, and then summing the quartile-by-quartile treatment effects into a single weighted average with weights given by the distribution of treated observations across quartiles. Assuming the only difference between treatment and control pupils besides the treatment is their 1994 scores, this procedure produces a nonparametric estimate of the effect of treatment on the treated. We chose to match pupils based on quartiles of the 1994 score distribution instead of a finer breakdown because it turns out that division into four groups is enough to eliminate almost all treatment-control differences in 1994 test scores.

The only significant within-quartile treatment-control difference in 1994 scores is for math in the first quartile. This can be seen in column 3 of table 8, which reports the within-quartile difference in 1994 scores for math and reading. The overall average effect of treatment status on 1994 scores (i.e., the average across quartiles weighted by the number treated in each quartile) is .031 with a standard error of .031 for math and −.052 with a standard error of .04 for reading. This suggests that the quartile-matching strategy does indeed balance the treatment and control groups.[11]

The matching estimates of treatment effects on 1996 scores are reported in column 6 of table 8. Except for the upper quartile of the reading score distribution, the within-quartile contrasts for 1996 are all positive and, in some cases, individually significant. The overall matching estimate is .25 with a standard error of .16 for math scores and .4 with a standard error of .16 for reading scores. The matching estimates for math are virtually identical to the corresponding regression estimates (reported in col. 10 of table 6). The matching estimates for reading are somewhat larger than the corresponding regression estimates (reported in col. 5 of table 6). Thus, these results reinforce the finding that pupils with same 1994 test scores did better in 1996 if they were in schools where teachers received additional training.

[11] The standard errors reported in this table were corrected for within-school correlation in test scores.

Table 8
Matching on 1994 Scores—Estimates for Secular Schools

	1994 Scores			1996 Scores		
Subject	Treatment (1)	Control (2)	Differences (3)	Treatment (4)	Control (5)	Differences (6)
Math, 1994 quartile:						
I	−1.483	−1.611	.128 (.062)	−.484	−.636	.152 (.201)
II	−.350	−.364	.014 (.055)	−.092	−.149	.241 (.188)
III	.339	.370	−.032 (.055)	.482	.131	.350 (.192)
IV	1.045	1.070	−.025 (.057)	.737	.441	.296 (.210)
Average effect	.031 (.031)			.250 (.158)		
Reading, 1994 quartile:						
I	−1.433	−1.355	−.078 (.069)	−.402	−1.108	.706 (.235)
II	−.589	−.538	−.051 (.058)	−.020	−.193	.173 (.201)
III	.164	.200	−.036 (.055)	.395	.062	.333 (.189)
IV	.969	.957	.012 (.063)	.402	.601	−.199 (.214)
Average effect	−.052 (.040)			.399 (.157)		

NOTE.— Standard errors are reported in parentheses. Standardized test scores were used in this analysis. The standard errors are corrected for within-school correlation in test scores.

V. The Cost-Effectiveness of Training Inputs

The estimates in tables 6 and 8 suggest that a conservative estimate of the effect of the training program in nonreligious schools is $.25\sigma$, where σ denotes the standard deviation of pupils' test scores.[12] The overall cost of the program in Jerusalem was about $12,000 per class. To determine whether this expenditure was worthwhile, we would need to estimate the value of an increase in children's test scores, something that is hard to quantify. However, we can compare the cost of attaining a given increase in test scores obtained by increasing teacher training with estimates of the cost of achieving a similar increase in test scores obtained by reducing class size or lengthening the school day. This tells us which of these three important inputs is most likely to be worth increasing.

Angrist and Lavy (1999) estimated that reducing the maximum class size in Israel from 40 to 30 would raise test scores by $.15\sigma$ and require a 28% increase in the number of classes. An unpublished memorandum

[12] This is broadly consistent with estimates of effect size for other inputs reported in the education research literature (see, e.g., Hedges et al. 1994).

from the Ministry of Education suggests that the annual operating cost of each new class would be about \$75,000, so the proposed reduction in class size would cost about \$21,500 per existing class. Assuming the relationship between class size and performance is linear, we estimate that it would cost \$35,000 per existing class to achieve a $.25\sigma$ test-score gain. Using the same data, Lavy (1998) estimated that lengthening the school week by 3.8 hours would raise test scores by $.15\sigma$. Since the annual cost of an extra weekly hour of instruction is \$2,000 for each class, this would cost \$7,600 per class. Again, assuming a linear relationship between inputs and performance, the cost of a $.25\sigma$ increase in scores is estimated to be about \$12,600. These calculations suggest that if the objective is improving pupil achievement, teacher training may be at least as cost-effective a strategy as lengthening the school day and considerably cheaper than reducing class size.

VI. Summary and Conclusions

The relationship between teachers' characteristics and their pupils' achievement has been the subject of many studies, but few have looked at the impact of in-service training. Our estimates suggest that an in-service training program run in Jerusalem's secular elementary schools raised children's achievement in reading and mathematics. These findings appear using a variety of statistical methods, including differences-in-differences, regression, and matching. The estimates for religious schools are not clear cut, but this is possibly because the training program in religious schools started later and was implemented on a smaller scale. The estimates for secular schools suggest that teacher training may provide a less costly means of increasing test scores than reducing class size or adding school hours.

Of course, the question remains whether the particular training program studied here is similar to training programs that might be used in other settings. Discussions with school officials lead us to believe that the approach taken in Jerusalem is not unusual in the Israeli context. Moreover, the type of training given to reading and math teachers using 30 Towns money was based on widely used pedagogical strategies ("humanistic mathematics" and "individualized instruction") originally developed in U.S. schools. However, one possibly unique feature of the 30 Towns intervention is the fact that it was closely linked to a weekly curriculum. Another important and possibly unique feature is the immediate feedback that participating teachers received from instructors. Since the basic philosophies behind the training strategies studied here are not especially foreign, however, it seems reasonable to imagine that the Israeli variant on training delivery could also be adapted for non-Israeli schools. At a minimum, the results here suggests that highly focused but relatively inexpensive teacher training of this type warrants further study.

Appendix

The 30 Towns Project

Teachers' In-Service Training

From the program's inception in September 1994, math and Hebrew teachers in program schools received in-house training on teaching methods. The training was carried out by instructors who met weekly with the math and Hebrew teachers for each grade, separately, and reviewed teaching methods for the material to be covered the following week. The math teachers received training based on the "humanistic mathematics" philosophy of teaching, developed by White (1987, 1993) for U.S. schools.

Hebrew training was carried out by two organizations: the Ministry of Education (six schools) and the Center for Educational Technology (four schools), which develops educational software marketed under well-known brand names worldwide. Curriculum for Hebrew teachers was based on the "individualized instruction" approach to education, which uses a systematic processes of diagnostics, individualized instruction, follow-up, and evaluation. This approach is based on a school of thought developed in the United States (see, e.g., Bishop 1971; Gibbons 1971; Glazer 1977). More recent applications of this approach are described in Corno and Snow (1986), Fullan and Miles (1992), and Fullan (1993). Adaptation and development for Israeli schools was carried out by a group of researchers at the Center for Educational Technology.

Beginning in January 1995, math and Hebrew teachers in program schools also received a course aimed at refreshing and deepening knowledge of subject matter. This training also covered a number of other teaching issues, for example, identifying and working with nonreaders and failing students.

Other Elements of the Intervention

In addition to in-house training, the project established a learning center to assist failing students in the treatment area. Targeted students from all treated schools were sent to the center for additional instruction after school. The project also funded a center for students with learning disabilities. A learning disability diagnostic expert and two teachers staffed the center. Only about 50 elementary school students received additional instruction in the centers, which also served middle schools and high schools in the treatment area. Finally, the project funded special activities with immigrant pupils intended to enhance learning and social integration and to get parents more involved with school and community activities.

References

Angrist, Joshua. "Using Social Security Data on Military Applicants to Estimate the Effect of Military Service Earnings." *Econometrica* 66, no. 2 (March 1998): 249–88.

Angrist, Joshua, and Lavy, Victor. *Evaluation of the 30 Towns Project in Jerusalem: Pisgat Zeev-Neve Yaakov* (in Hebrew). Jerusalem: Jerusalem Schools Authority, January 1998.

———. "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement." *Quarterly Journal of Economics* 114 (May 1999): 533–76.

Ashenfelter, Orley A. "Estimating the Effect of Training Programs on Earnings." *Review of Economics and Statistics* 60, no. 1 (February 1978): 47–57.

Ashenfelter, Orley A., and Card, David. "Using the Longitudinal Structure of Earnings to Estimate the Effect of Training Programs on Earnings." *Review of Economics and Statistics* 67, no. 4 (November 1985): 648–60.

Bangert, R.; Kulik, J.; and Kulik, C. "Individualized Systems of Instruction in Secondary Schools." *Review of Educational Research* 53, no. 2 (1983): 143–58.

Bartel, A. P. "Training, Wage Growth and Job Performance: Evidence from a Company Database."*Journal of Labor Economics* 13, no. 3 (July 1995): 401–25.

Behrman, J. R.; Khan, S.; Ross D.; and Sabot, R. "School Quality and Cognitive Achievement Production: A Case Study for Rural Pakistan." *Economics of Education Review* 16, no. 2 (April 1997): 127–42.

Bishop, Lloyd K. *Individualizing Educational Systems: The Elementary and Secondary School: Implications for Curriculum, Professional Staff and Students*. New York: Harper & Row, 1971.

Black, Sandra E. "Do Better Schools Matter? Parental Valuation of Elementary Education." *Quarterly Journal of Economics* 114 (May 1999): 577–99.

Boston Globe. "Towards Better Teachers" (July 23, 1998), p. A14.

Bray, J. H., and Howard, G. S. "Methodological Considerations in the Evaluation of a Teacher-Training Program." *Journal of Educational Psychology* 72, no. 1 (1980): 62–70.

Bressoux, P. "The Effect of Teachers' Training on Pupils' Achievement: The Case of Elementary Schools in France." *School Effectiveness and School Improvement* 7, no. 3 (1996): 252–79.

Card, David, and Krueger, Alan. "Does School Quality Matter? Returns to Education and the Characteristics of Public Schooling in the United States." *Journal of Political Economy* 100, no. 1 (February 1992): 1–40.

Card, D., and Sullivan, D. "Measuring the Effect of Subsidized Training Programs on Movements In and Out of Employment." *Econometrica* 56, no. 3 (May 1988): 497–530.

Carroll, J. G. "Effects of Training Programs for University Teaching

Assistants." *Journal of Higher Education* 51, no. 2 (March–April 1980): 167–83.

Coleman, J. S., et al. *Equality of Educational Opportunity*. Washington, DC: U.S. Government Printing Office, 1966.

Corno, L., and Snow, R. E. "Adapting Teaching to Individual Differences among Learners." In *Handbook of Research on Teaching*, edited by M. C. Wittrock. New York: Macmillan, 1986.

Dildy, P. "Improving Student Achievement by Appropriate Teacher In-Service Training: Utilizing Program for Effective Teaching (PET)." *Education* 103, no. 2 (Winter 1982): 132–38.

Ehrenberg, R. G,. and Brewer, D. J. "Do School and Teacher Characteristics Matter? Evidence from *High School and Beyond.*" *Economics of Education Review* 13, no. 1 (March 1994): 1–17.

Farrell, J. P., and Oliveira, J. "Teacher Costs and Teacher Effectiveness in Developing Countries." In *Teachers in Developing Countries: Improving Effectiveness and Managing Costs,* edited by J. P. Farrell and J. B. Oliveira, pp. 175–86. Economic Development Institute Seminar Series. Washington, DC: World Bank, 1993.

Fullan, M. G. "Change Forces: Probing the Depths of Educational Reform." London: Falmer, 1993.

Fullan, M .G., and Miles, M. B. "Getting Reform Right: What Works and What Doesn't." *Phi Delta Kappan* 73, no. 10 (1992): 744–52.

Gibbons, M. *Individualized Instruction*. New York: Teachers College Press, 1971.

Glazer, R. *Adaptive Education: Individual Diversity and Learning*. Psychological Series. New York: Holt, Rinehart & Winston , 1977.

Goldhaber, Dan D., and Brewer, D. J. "Why Don't Schools and Teachers Seem to Matter: Assessing the Impact of Unobservables on Educational Productivity." *Journal of Human Resources* 32, no. 3 (Summer 1997): 505–23.

Hanushek, Eric A. "Teacher Characteristics and Gains in Student Achievement: Estimation Using Micro Data." *American Economic Review* 61, no. 2 (May 1971): 280–88.

———. "The Economics of Schooling: Production and Efficiency in Public Schools." *Journal of Economic Literature* 24, no. 3 (September 1986): 1141–57.

Hanushek, Eric A.; Rivkin, Stephen G.; and Kain, John F. "Teachers, Schools, and Academic Achievements." Working Paper no. 6691. Cambridge, MA: National Bureau of Economic Research, 1998.

Hedges, L. V.; Laine, R. D.; and Greenwald, R. "Does Money Matter? A Meta-analysis of Studies of the Effects of Differential School Inputs on Student Outcomes." *Educational Researcher* 23, no. 3 (1994): 5–14.

Highsmith, R. "A Study to Measure the Impact of In-Service Institutes on the Students of Teachers Who Have Participated." *Journal of Economic Education* 5, no. 2 (Spring 1974): 77–81.

Krueger, Alan B., and Rouse, C. "The Effect of Workplace Education on Earnings, Turnover, and Job Performance." *Journal of Labor Economics* 16, no. 1 (January 1998): 61–94.

Lavy, Victor. "Endogenous School Resources and Cognitive Achievement in Primary Schools in Israel." Discussion Paper no. 95.03. Jerusalem: Falk Institute for Economic Research in Israel, 1995.

———. "Using Quasi-Experimental Designs to Evaluate the Effect of School Hours and Class Size on Student Achievement." Photocopied. Jerusalem: Hebrew University of Jerusalem, Department of Economics, 1998.

Monk, David. H. "Subject Area Preparation of Secondary Mathematics and Science Teachers and Student Achievement." *Economics of Education Review* 13, no. 2 (June 1994): 125–45.

Moulton, Brent R. "Random Group Effects and the Precision of Regression Estimates." *Journal of Econometrics* 32 (1986): 385–97.

Murnane, Richard J. *The Impact of School Resources on the Learning of Inner City Children.* Cambridge, MA: Ballinger, 1975.

New York Times. "The Flaw in Student-Centered Learning" (July 20, 1998), p. A15.

Romberg, Thomas A. *Toward Effective Schooling—the IGE Experience.* Lanham, MD: University Press of America, 1985.

Rubin, D. B. "Assignment to a Treatment Group on the Basis of a Covariate." *Journal of Educational Statistics* 2, no. 1 (1977): 1–26.

Schober, H. M. "The Effects of In-Service Training on Participating Teachers and Students in Their Economics Classes." *Journal of Economic Education* 15, no. 4 (Fall 1984): 282–95.

Welch, F. "Measurement of the Quality of Schooling." *American Economic Review* 56, no. 1 (March 1966): 379–92.

White, Alvin. "Mathematics as a Humanistic Discipline." In *The International Encyclopedia of Education.* Oxford: Pergamon, 1987.

———. *Essays in Humanistic Mathematics.* Washington, DC: Mathematical Association of America, 1993.