The Impact of Teacher Training on Student Achievement
Quasi-Experimental Evidence from School Reform Efforts in Chicago

Brian A. Jacob
Lars Lefgren

A B S T R A C T

While there is a substantial literature on the relationship between general teacher characteristics and student learning, school districts and states often rely on in-service teacher training as a part of school reform efforts. Recent school reform efforts in Chicago provide an opportunity to examine in-service training using a quasi-experimental research design. In this paper, we use a regression discontinuity strategy to estimate the effect of teacher training on the math and reading performance of elementary students. We find that marginal increases in in-service training have no statistically or academically significant effect on either reading or math achievement, suggesting that modest investments in staff development may not be sufficient to increase the achievement of elementary school children in high-poverty schools.

I. Introduction

There is a substantial literature on the relationship between teacher characteristics and student learning. Most prior research on this topic has focused

on teachers' educational background, years of teaching experience and salaries. The results of this work are mixed. While it is clear that certain teachers are more effective than others at increasing student performance, there is considerably less consensus on whether specific, observable teacher characteristics such as education or experience produce higher performance.[1]

While most research has focused on general skills, school districts and states often rely on in-service staff development as a way to improve student learning. This on-the-job training seeks to instruct teachers in content as well as pedagogy. Professional development is an extremely widespread practice in U.S. public schools. Seventy-two percent of teachers report having engaged in training related to the subject area of their main teaching assignment during the previous 12 months (Parsad et al. 2000). A similar fraction reports having received training on how to implement new teaching methods. Despite the widespread nature of these activities, the intensity of training is typically fairly low, with more than half of the teachers engaging in eight hours or less of training in each of these areas per year. Unfortunately, most of the existing research on in-service training suffers from the fact that the training is endogenously determined by teachers and schools.

Recent school reforms in Chicago, however, provide an excellent opportunity to evaluate the causal impact of teacher training on student performance. In 1996, the Chicago Public School system (CPS) placed 71 of its 489 elementary schools on academic probation. These probation schools received special funding for staff development as well as technical assistance and enhanced monitoring. Eligibility for probation was determined on the basis of standardized reading scores—schools in which fewer than 15 percent of students scored at or above national norms in reading were subject to probation; those with 15 percent or more of students at national norms were not subject to probation. The existence of strict cutoffs created a highly nonlinear relationship between a school's reading achievement in 1996 and the likelihood that the school was on probation in subsequent years. We exploit this cutoff to identify the impact of teacher training on student achievement.

Note that this strategy does *not* identify the aggregate effect of the school probation policy since the accountability measures provided *all* low-achieving schools (both those who just missed and just made the cutoff) an incentive to increase student performance—low-achieving schools that did not demonstrate improvement were subject to further sanctions.[2] Rather, this strategy effectively identifies the impact of the *resources* provided to certain low-achieving schools under the probation policy.

---

1. There is still considerable disagreement regarding the causal effect of educational expenditures on academic achievement. Hanushek (1996) asserts that there is little evidence that increased educational expenditures can systematically increase academic achievement. Hedges and Greenwald (1996) offer a different interpretation of the evidence, claiming that although many individual studies find no significant effect, the average effect estimate is positive. More recent experimental evidence suggests that at least one form of expenditure— reduced class size—does have a substantial effect on student achievement (Krueger 1999). Using a quasi-experimental research design, Guryan (2000) also finds that increases in school funding may have increased the performance of elementary school students in Massachusetts.
2. Jacob (2002) finds evidence that the incentives provided by school probation, along with student-oriented accountability measures, led to a substantial increase in math and reading achievement.

Because the technical assistance and monitoring resources provided to probation schools were quite small (see discussion below) and were designed primarily to enhance teacher classroom performance (and thus might be considered a component of teacher training), our discussion in this paper will focus on the impact of teacher training with the understanding that it includes the effect of all of the resources provided to schools under the probation policy.

Utilizing exogenous variation in probation status caused by the cutoff described above, we find that moderate increases in teacher training have no statistically or academically significant effect on either reading or math achievement. These results do not vary across race, gender, socioeconomic background, or student ability, and are robust to a number of alternative specifications. Our results suggest that modest investments in staff development may not be sufficient to increase the achievement of elementary school children in high-poverty schools. The remainder of this paper is organized as follows. Section II reviews the literature on teacher training and provides background on the Chicago probation policy. Section III describes our data and Section IV explains our empirical strategy. Section V presents findings on the effectiveness of in-service training. Section VI explores the policy effects in more detail, examining the heterogeneity in effects across students and providing a series of robustness checks for our results. Section VII discusses some of the implications of these findings and concludes.

II. Background

A. *Prior Literature*

Despite the importance of teacher training in most school districts, there is surprisingly little evidence on the effect of teacher training on student achievement. Indeed, as Angrist and Lavy (2001) pointed out, there seems to have been more research on the impact of teacher training in developing countries than in developed countries. Early research on teacher training presents a rather pessimistic view of the effectiveness of staff development for increasing student performance. In a meta-analysis of 93 studies of the effect of teacher development on student performance, Kennedy (1998) reports that only 12 studies show positive effects of staff development. Consistent with this finding, Corcoran (1995) and Little (1993) claim that typically staff development is a low-intensity affair that lacks continuity and accountability. There are some notable exceptions to these findings however. Bressoux (1996), using a quasi-experimental research design, and Dildy (1982), examining the results of a randomized trial, find that teacher training increases student performance. Wiley and Yoon (1995) and Cohen and Hill (2000) are others who find teacher development programs to have at least small impacts on student performance.

One recent paper that finds particularly strong effects of teacher training is Angrist and Lavy (2001). While this paper presents strong evidence regarding the potential effectiveness of teacher training programs, this analysis has several limitations. In addition to funding teacher training, the intervention consisted of several other components that might have increased student achievement, including the establishment of a learning center to assist failing students after school and a project to support immigrant students and their families. Perhaps more importantly, the schools were

not randomly assigned to the treatment, forcing the authors to rely on matching and difference-in-difference strategies for the estimation. In the final section, we discuss several reasons why our findings may differ from those described above.

B. Background on School Reform in Chicago

The CPS is the nation's third-largest school district, serving over 430,000 largely low-income students. In the late 1980s, then Secretary of Education William Bennett described Chicago public schools as the worst in the nation. In 1996, the CPS introduced a highly publicized reform effort that emphasized holding students, teachers and administrators accountable for academic achievement.

Under the Chicago policy, schools in which fewer than 15 percent of students met national norms on standardized reading exams were placed on academic probation.[3] While several schools received waivers, 71 elementary schools serving over 45,000 students were placed on academic probation in the first year of the program.[4] To improve student achievement in these schools, the CPS provided probation schools additional resources to buy staff development services from an external organization of their choice. In 1998–99, probation schools were working with 17 different external partners, including universities, nonprofit organizations, and independent consultants. During the first year a school was on probation, the CPS paid 100 percent of the costs of the external partner (up to $90,000). In the second year, the reimbursement dropped to 50 percent. After two years, the Board paid one-third of the cost of external partners.

In addition to these direct resources, the CPS provided probation schools with technical assistance and monitored the progress of the school. The Office of Accountability (OA) assigned each probation school a probation manager, generally a high-level school administrator with experience as a principal, whose job was to help school staff to develop and implement a school improvement plan. Elementary schools on probation also were assigned a business manager intern to manage the operational and financial aspects of the school, freeing the principal to address educational issues and to assist the external partners in staff development.

Table 1 presents information regarding the effect of probation on teacher development (see Smylie et al. 2001 for a more detailed discussion of professional develop-

---

3. The Chicago reform also included a student accountability policy in which students in third, sixth and eighth grade were required to meet minimum achievement levels in reading and math in order to move to the next grade. For more details on the student accountability policy and its impact on student outcomes, see Jacob (2002) and Jacob and Lefgren (2001).

4. Probation schools that do not exhibit sufficient improvement may be reconstituted, which involves the dismissal or reassignment of teachers and school administrators. While no elementary schools were reconstituted during the initial years of the policy, teachers expressed concerns about job security in light of the policy. In the early years of the program, in order to move off of probation, at least 20 percent of students in the school had to meet national norms in reading. In 2000, the standard was raised so that schools with fewer than 20 percent of students at national norms in reading were subject to probation and all schools needed to meet a 25 percent standard to move off of probation. In 1997–98, eight elementary schools were removed from probation because of achievement gains, but 13 additional schools were placed on probation. By 1998–99, only 54 elementary schools were on probation.

Table 1
Average Monthly Participation in Professional Development Activities by 1997 Probation Status

Variable	1994		1997		1999	
	Probation Schools	Other Schools	Probation Schools	Other Schools	Probation Schools	Other Schools
All activities	3.119	2.976	3.420	2.644	3.445	2.877
	(2.167)	(2.206)	(2.432)	(2.110)	(2.440)	(2.205)
Activities with own school	1.174	1.129	1.069	0.821	1.114	0.881
	(0.645)	(0.679)	(0.708)	(0.650)	(0.761)	(0.682)
Activities with teacher network	0.443	0.470	0.723	0.596	0.799	0.644
	(0.520)	(0.594)	(0.676)	(0.655)	(0.706)	(0.669)
Activities with outside partner	0.496	0.469	0.645	0.494	0.642	0.498
	(0.547)	(0.558)	(0.644)	(0.590)	(0.647)	(0.601)
Activities with district	0.347	0.302	0.393	0.288	0.374	0.341
	(0.454)	(0.454)	(0.515)	(0.421)	(0.508)	(0.494)
Activities with union	0.255	0.201	0.195	0.152	0.226	0.174
	(0.407)	(0.373)	(0.408)	(0.352)	(0.442)	(0.393)
Activities with college or university	0.503	0.406	0.412	0.348	0.348	0.345
	(0.650)	(0.601)	(0.640)	(0.581)	(0.594)	(0.598)
Observations	215	936	365	2,068	145	1,375

Notes: Probation status refers to the probation status of the school in 1997. Average monthly participation rates for each activity are computed using the midpoints of response categories. These participation rates are added to compute the frequency of all professional development activities. Standard deviations are in parentheses. Data come from teacher surveys generously provided by the Consortium on Chicago School Research.

Table 2

Distribution of Monthly Participation in All Professional Development Activities by 1997 Probation Status

Variable	1994		1997		1999	
	Probation Schools	Other Schools	Probation Schools	Other Schools	Probation Schools	Other Schools
10th percentile	0.92	0.67	0.83	0.58	0.58	0.67
25th percentile	1.67	1.42	1.67	1.17	1.58	1.25
50th percentile	2.50	2.50	2.83	2.08	2.83	2.33
75th percentile	4.00	3.92	4.58	3.50	4.58	3.83
90th percentile	6.33	6.00	6.92	5.50	6.17	5.92
Observations	215	936	365	2,068	145	1,375

Notes: Probation status refers to the probation status of the school in 1997. Average monthly participation rates for each activity are computed using the midpoints of response categories. These participation rates are added to compute the frequency of all professional development activities. Data come from teacher surveys generously provided by the Consortium on Chicago School Research.

ment in the CPS).[5] The first two columns show that in 1994, teachers in schools that would be placed on probation in 1997 participated in school sponsored professional development at about the same rate as other teachers. In 1997 and 1999, teachers in probation schools were participating at substantially higher levels than their colleagues. In 1997, probation teachers attended an average of 3.4 professional development activities each month compared to only 2.6 activities for other teachers.[6] The increase is reflected in activities sponsored by the school, teacher networks, outside partners, and the CPS. This evidence suggests that probation increased the frequency of professional development activities by about 25 percent in the first year. Table 2 provides evidence regarding the distribution of professional development activity by time period and 1997 probation status. While probation increased the frequency of professional development activities across the distribution, the largest (absolute) gains came for those teachers in the top half of the distribution.

Moreover, there is some evidence that the quality of teacher-training activities in probation schools improved from 1997 to 1999. Using teacher survey data, Smylie

5. The data used for table come from surveys conducted by the Consortium on Chicago School Research. These surveys were administered in 1994, 1997, and 1999 to all CPS teachers and asked a number of questions regarding the teachers' work environment—including the extent and nature of professional development activities. We thank the Consortium for making these data available.

6. The data report participation during the past school year using ranges of values (for example, between three and five times in the last year). To calculate average participation we assume that teacher participation was in the midpoint of the range. We further assume teachers in the highest category attended 12 activities during the school year. We divide the number of reported activities by six (the number of months school had been in session at the time of the survey) to obtain monthly participation. To calculate the frequency of all professional development activities, we added the monthly participation rates of each type of development activity.

et al. (2001) constructed a measure of the quality of professional development that incorporates teacher perceptions of whether the professional development in their school was (among other things) related to student needs, sustained and focused, provided sufficient time to try and evaluate ideas, and included ample follow-up. The measure is reported on a ten-point scale with a 1997 mean and standard deviation of 6.0 and 1.5 respectively for elementary school teachers. From 1997 to 1999, teachers in probation schools reported an increased from 5.8 to 6.3 in comparison to colleagues in other schools who reported an average of 6.0 in both years. Consistent with the survey data, a qualitative case study of the external partners found that teachers were generally satisfied with the professional development offered by these organizations (Finnigan et al. 2001).

III. Data

This study utilizes administrative data from the Chicago Public School system (CPS). Student records provide detailed demographic and educational background data on individual students for each academic year, including prior achievement scores, previous school and residential mobility, birth date, race, gender, family composition, free lunch status, and special education and bilingual services received. School records provide average demographic data at the school level, including percent low-income, average daily attendance, and school mean test scores. The primary outcome measures we use are math and reading scores on the Iowa Test of Basic Skills (ITBS), a multiple-choice exam that CPS students take annually in grades two to eight. The ITBS is measured in terms of grade equivalents (GEs), which reflect the years and months of learning that a student has mastered. For example, a student at national norms in sixth grade will score 6.8 GEs, which means the student has mastered material up to the eighth month of sixth grade.

The baseline sample for this study consists of the cohort of third through sixth grade students who were enrolled in a Chicago elementary school in the fall of 1996 (n = 131,314). We limit the sample to students in these grades because we measure performance gains over three years and ITBS scores are not available for students beyond eighth grade. We delete 198 students who attended a special needs school in the fall of 1996, 3,981 students (3 percent) who are missing student or school demographic data, and 26,907 students (20 percent) who did not take the ITBS exam in the spring of 1996, which leaves us with a sample of 100,288 students in 461 different schools. The CPS does not require schools to test students below the third grade. The majority of students with missing test scores in spring 1996 were second graders at the time since testing is optional for students below third grade (determined by school staff). As a check, we have done all of the analysis including these students and obtained virtually identical results.[7]

Table 3 presents summary statistics on this sample. Roughly 20 percent of students attended a school on probation at some point between 1997 and 1999 and these students spent an average of 1.9 years in a school on probation. As one would expect,

---

7. In these checks, we set the missing test score to zero and included a dummy variable that indicated that this was the case.

Table 3
Summary Statistics

Means (standard deviations)	Total	Never in school on probation between 1997 and 1999	In school on probation for at least one year between 1997 and 1999
Treatment			
In school on probation in 1997	0.148	0.000	0.779
Years in school on probation from 1997 to 1999	0.358	0.000	1.924
Student outcomes			
1999 reading score	6.590 (1.984)	6.780 (1.973)	5.801 (1.832)
1999 math score	6.803 (1.809)	6.974 (1.793)	6.086 (1.694)
Not tested in 1999	0.143	0.150	0.113
Tested, but excluded from reporting in 1999	0.134	0.132	0.140
Enrolled in the CPS	0.875	0.866	0.913
Changed schools (left 1996 school)	0.237	0.204	0.371
Student demographics			
1996 reading score	3.591 (1.618)	3.721 (1.645)	3.021 (1.357)
1996 math score	3.884 (1.358)	3.994 (1.374)	3.402 (1.167)
Tested, but excluded from reporting in 1996	0.077	0.079	0.071
Black	0.606	0.545	0.871
Hispanic	0.254	0.284	0.121
Male	0.500	0.499	0.506
Black male	0.301	0.269	0.439
Hispanic male	0.254	0.144	0.062
Age in June 1996	10.264 (1.218)	10.278 (1.215)	10.202 (1.229)
Free lunch	0.794	0.756	0.959
Reduced price lunch	0.084	0.096	0.028
Currently in bilingual program	0.088	0.097	0.048
Formerly in bilingual program	0.180	0.207	0.062

Table 3 (*continued*)

Means (standard deviations)	Total	Never in school on probation between 1997 and 1999	In school on probation for at least one year between 1997 and 1999
Special education	0.111	0.113	0.106
Living with relatives	0.109	0.119	0.065
Living in foster care	0.050	0.043	0.076
Concentration of poverty	0.276	0.139	0.876
(block group)	(0.733)	(0.663)	(0.723)
Social status (block group)	−0.264	−0.176	−0.650
	(0.688)	(0.687)	(0.542)
Third grade	0.226	0.220	0.256
Fourth grade	0.240	0.237	0.252
Fifth grade	0.265	0.267	0.252
Sixth grade	0.269	0.276	0.239
School characteristics			
Enrollment	762	774	713
	(311)	(324)	(238)
Attendance rate	92.5	93.0	90.5
Mobility rate	29.0	27.5	35.6
Truancy rate	2.5	1.9	5.2
Percent Black	58.9	53.2	84.3
Percent Hispanic	27.1	29.9	14.8
Percent limited English proficient	14.3	15.8	7.8
Percent low income	84.9	82.6	95.1
Number of observations	100,228	81,554	18,674

Notes: The sample includes students who were in the third through sixth grades in Fall 1996. We exclude children who were missing demographic information and 1996 ITBS test scores. We also drop observations with missing school demographic variables as well as all students in special needs schools.

probation schools served the most disadvantaged students in the CPS. Students who spent at least one year in a probation school scored roughly six to seven months behind their peers in math and reading in 1996. Over 95 percent of students who attended a probation school received free lunch compared with 76 percent of students who did not attend a probation school and students in probation schools were nearly twice as likely to be living in a foster home in fall 1996. Hispanic students were substantially less likely to attend a probation school than Black students—on average 12 percent of students attending probation schools were Hispanic compared with 21 percent of the CPS whereas nearly 87 percent of students attending probation schools

were Black compared with only 61 percent in the CPS. Similarly, students in proba-tion schools experienced school level mobility, truancy, and low-income rates con-siderably higher than peers in nonprobation schools.

IV. Empirical Strategy

Teacher training is one of many factors that may influence student learning. The relationship between inputs such as teacher training and learning out-comes can be captured in the following education production function:

$$(1) \quad Y_{i,s,t+1} = \beta_1 (Training)_{s,t} + BX_{i,t} + \Gamma Z_{s,t} + u_s + v_i + \varepsilon_{i,s,t},$$

where i, s, and t are individual, school, and time subscripts respectively. Y is the outcome, *Training* indicates whether a student's teachers received in-service train-ing, X is a vector of student demographic and past performance variables, Z is a vector of other teacher and school characteristics, u represents the effect of unob-served school quality, v is time invariant unobserved student ability, and ε is an error term. This specification allows training to increase the efficacy of instruction and improve subsequent student academic outcomes.

The difficulty in estimating the causal impact of *Training* is that teachers and schools may select, or be selected, into training on the basis of characteristics that are unobservable to the researcher. In the case of teacher training, it is difficult to even sign the direction of the potential bias. On one hand, as Lavy (1995) and others have noted, there is often a negative correlation between school inputs and pupil achievement because measures of socioeconomic disadvantage are used to decide which schools get the most inputs. In this case, it is likely that Cov(*Training, u*) < 0, which will bias the estimate of β_1 downward. On the other hand, to the extent that teacher training is often a voluntary activity determined by the teachers and administra-tors in a particular school, it is possible that the most motivated teachers and schools seek training so that Cov(*Training, u*) > 0, which will tend to bias β_1 upward.

The recent school reform efforts in Chicago, however, provide a unique opportu-nity to identify the causal impact of teacher training on student achievement. The strict test score cutoff for probation generated a highly nonlinear relationship be-tween school reading performance in 1996 and the average number of years a student spent in a school on probation between 1997 and 1999. Figure 1 illustrates the rela-tionship between the percent of students meeting national norms in a student's 1996 school and the number of years between 1996 and 1999 that the student attended a school on probation. We can see that students enrolled in schools where 13 percent of students met national norms in 1996 attended schools on probation for an average of two years over this period. In contrast, students in schools where 15 percent of students met national norms in 1996 attended schools on probation for only 0.30 years on average over the same period.

If the assignment mechanism were followed perfectly, the resulting discontinuity would provide a way to estimate the effect of teacher training on student achieve-ment. Assuming that unobservable characteristics do not vary discontinuously around the cutoff, the probation decision rule essentially replicates random assign-ment of training to schools around the cutoff. One can thus identify the treatment
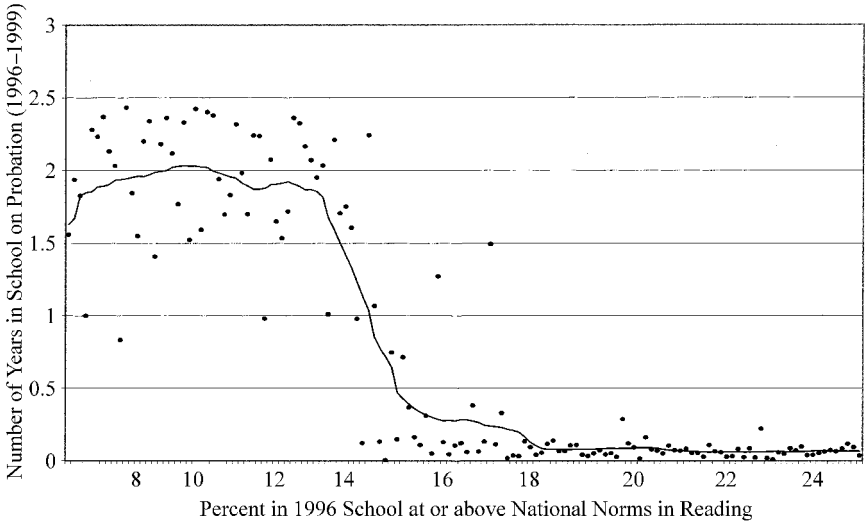
Figure 1

The Relationship between 1996 School Reading Achievement and Subsequent Probation Experience

Notes: The individual points are group averages of the raw data, averaged for every one-tenth of a percent at or above national norms. The continuous line is a regression lined that was smoothed using a lowess procedure with a bandwidth parameter of 0.1. The distances between the percentages on the x-axis are uneven since there were some values that did not have any data.

effect by simply comparing students in schools on either side of the cutoff. For example, if students in schools that just missed the cutoff (and were thus likely to be placed on probation and have access to the additional teacher-training resources) learned much more than students in schools that just made the cutoff (and thus avoided probation), then one might conclude that the staff development and technical assistance associated with probation has a positive impact. Even in the case in which some schools near the margin receive waivers, the highly nonlinear relationship between school performance and treatment provides exogenous variation in treatment allowing identification.

This strategy is often referred to as a regression discontinuity design.[8] If there is a perfect relationship[9] between 1996 school reading achievement and the number of

8. This type of regression discontinuity analysis was pioneered in educational evaluation research. In one of the first papers to introduce this design, Thistlethwaite and Campbell (1960) utilized the fact that National Merit Awards are given on the basis of whether a test score exceeds a threshold to estimate the effect of the award on a student's other scholarship receipt and college aspirations. This strategy was used widely in evaluations of compensatory education programs mandated under Title I (Trochim 1984) as well as other contexts. Other studies to use this design include Berk and Rauma (1983), Angrist and Lavy (1999), Black (1999), Hahn et al. (1999), Jacob and Lefgren (2001), and Guryon (2001).

9. By perfect relationship, we mean that the treatment is completely determined by observed performance. In this case, treatment is necessarily orthogonal to any unobserved characteristics. Thus, after controlling appropriately for performance, the OLS estimates should be unbiased because the treatment is orthogonal

years a student spends in a school on probation, then a properly specified OLS model that included a dummy variable indicating whether the student was in a school below the cutoff in 1996 would provide unbiased estimates of the training effect.

However, there are several reasons that the relationship between years in a probation school and 1996 school reading achievement is not perfect. First, several schools that scored below the probation cutoff were waived from the policy; for example, 15 of the 77 elementary schools that scored below the cutoff in 1996 received waivers.[10] Second, 25 schools that were placed on probation in 1996–97 raised achievement enough to be removed from probation in the next two years. Conversely, 16 schools that missed the probation cutoff in the first year were placed on probation in the following two years. Finally, there was substantial student mobility. Many students moved between probation and nonprobation schools during this period.

There are several approaches to estimating the treatment effect in the presence of this "fuzzy" discontinuity. We describe these approaches here, discussing the advantages and disadvantages of each and highlighting the assumptions underlying each strategy. In the next section, we show that each of these strategies produce comparable results.

In considering how to address the fuzzy discontinuity, it is useful to conceptualize the regression discontinuity approach in an instrumental variables framework. Here Equation 1 represents the second stage or "equation of interest." To obtain consistent estimates, we must predict the teacher training with some variable that is not directly correlated with student achievement. In our case, we can use information on the achievement level of a student's school in 1996 along with information about the decision rule to predict the number of years a student will spend in a school on probation. For example, the first stage equation may take the following form:

$$(2) \quad Training_{i,s,t} = \gamma_1 \, (Norms_{s,t-1}) + \Gamma_1 Z_{s,t} + B_1 X_{i,t} + \gamma_4 u_s + \gamma_5 v_i + \eta_{i,s,t}$$

where Z and X are defined as previously indicated and the variable $Norms$ captures the reading score of the student's 1996 school. In the simplest case, $Norms$ might be a single indicator variable that takes on the value of one if at least 15 percent of students scored at or above the national norm in reading. This would indicate whether the school had exceeded the official cutoff. Given the relationship we saw in Figure 1, this variable would clearly predict the teacher training.

Alternatively, we might specify a slightly more complicated first stage that takes the following form:

$$(3) \quad Training_{i,s,t} = \gamma_1 \, (Norms_{s,t-1}^{0-14}) + \gamma_2 \, (Norms_{s,t-1}^{14-15}) + \gamma_3 \, (Norms_{s,t-1}^{15+})$$
$$+ \, \Gamma_1 Z_{s,t} + B_1 X_{i,t} + \gamma_4 u_s + \gamma_5 v_i + \eta_{i,s,t}$$

where Z and X are defined as previously indicated. The variables labeled $Norms$ capture the nonlinear relationship observed in Figure 1—taking into account that

---

to the error term. Furthermore, IV and OLS estimates will be the same because the treatment is perfectly predicted in the first stage.

10. Waivers were granted by the superintendent or regional administrators on a case-by-case basis. There were a number of different reasons by waivers were given. In one case, the central office has just installed a new principal and wanted to give this person time to turn around the school without the added burden of probation. In another case, the local alderman and local residents intervened on behalf of the school, urging the CPS to give the school more time to improve.

time spent in a probation school falls dramatically (but *not* discontinuously) with school reading performance in a range just below the cutoff. The choice of the exact boundaries for the marginal regions in this specification was determined by visual inspection. The superscripts over the *Norms* variables indicate the use of a spline. For example, a student in a school with 16 percent of the students at or above national norms would have a value of 14 for $Norms^{0-14}$, a value of 1 for $Norms^{14-15}$, and a value of 1 for $Norms^{15+}$. Taking advantage either of a single indicator variable or a spline in school reading performance, we then estimate the 1999 achievement in a two-stage least squares framework using the predicted value of the years in a probation school. The results are essentially identical whether we specify the first stage using an indicator variable or a spline. The primary advantage of the spline specification is that it provides greater power in the first stage, thus producing slightly more precise estimates.

One might be concerned that the assignment of waivers in the marginal range just below the cutoff may be endogenous. This does not affect the consistency of our estimates since our strategy relies on variation in years in a probation school based on the *observed* distance below the cutoff. We do not rely on variation attributable to unobserved differences across schools or students.

Because we know the nature of the nonlinearity between school reading performance and years in a probation school ex ante, the functional form of the selection equation provides convincing exclusion restrictions necessary for IV estimation. However, our approach does rely on several assumptions. With sharp regression discontinuity designs, it is sufficient to assume that unobserved characteristics do not change discontinuously at the cutoff. With our fuzzy design, we must further assume that unobserved characteristics are not related to school performance in the same way as treatment assignment; for example, unobserved school characteristics cannot be changing dramatically in the marginal area below the cutoff. This may not be true in cases where participants have precise control over their performance, particularly near the margin of interest, or in cases in which failing to achieve a cutoff is associated with additional consequences not directly related to the treatment in question.

One such concern in our case is that teachers or school administrators may attempt to influence student scores on the margin. For example, a school that knows it is in danger of probation may attempt to influence testing to get on probation (and thus get the associated resources) or get off probation (to avoid potential sanctions). While Jacob and Levitt (2002) identify cases in which Chicago teachers may have improperly assisted students on exams, this behavior appears limited to a relatively small number of classrooms and is thus unlikely to affect our results. To ensure that teacher cheating does not bias our results, we reran all estimates omitting the small set of schools in which the authors identified a high degree of cheating in 1996. The results are virtual identical.

Another concern is student mobility. Because prior research indicates that student mobility rates are generally higher in lower-achieving schools (see Kerbow 1996; Hanushek, Kain, and Rivkin 2001), we expect to find higher mobility rates among probation schools in comparison to nonprobation schools. While high mobility in itself is not problematic, if probation causes high-achieving or motivated students to leave the CPS our estimates may be biased. Using the regression discontinuity

design above, we are able to examine whether probation status itself caused certain students to leave the school or the CPS. As we show in the next section, it appears that probation did not induce student mobility, which reinforces the validity of the achievement estimates.[11]

The other important assumption in our analysis involves the functional form of the relationship between current school achievement and future student achievement. Our instruments in Equation 2 are nonlinear functions of school-level achievement. If the true relationship between school mean achievement and future student performance is nonlinear for the range of values we examine, the estimated treatment effect could reflect underlying nonlinearity in the achievement relationship.

Although this concern is mitigated to some extent since we examine schools within a limited range around the probation cutoff, we nonetheless examine whether it is a serious concern in this study. First, we estimate models that allow for school mean achievement in 1996 to influence future student performance in a nonlinear fashion by including second and third order polynomials in Equations 1 and 2. We show that this does not change our general findings.

Second, we include a cohort of students who were enrolled in third to sixth grade in 1993, *prior* to the introduction of the Chicago school accountability reforms. If our probation estimate were due to a misspecified functional form, one would expect to see a similar finding in the years previous to the introduction of the probation policy. Taking advantage of students in the prior cohort, we are able to control for the nonlinear function of school achievement that we use as instruments in our baseline specification. For this specification, we obtain instruments by interacting the spline of school reading achievement with a dummy that takes a value of one if the probation policy was in effect. If the relationship between school and student achievement is stable over time, this procedure will guarantee that our findings are not driven by nonlinearity in the effect of school reading performance on student performance. In the next section, we show that this does not change our results.

V. Results

A. *Main Findings*

Under the assumptions described above, if teacher training has a substantial impact on academic achievement, we would expect to see a rapid change in the average achievement level around the probation cutoff. Figures 2a and 2b provide a way to visually identify the treatment effect. The heavy solid line shows the average number of years the student attended a school on probation between 1996–97 and 1998– 99. The other lines show the average 1999 reading and math achievement in Figures 2a and 2b respectively. If the teacher training associated with probation were beneficial, we would expect to see a drop in performance as school reading performance neared and surpassed the cutoff.

---

11. Because the probation policy was not commonly known until the beginning of the 1996–97 school year, it seems unlikely that students would have shifted schools before this point.
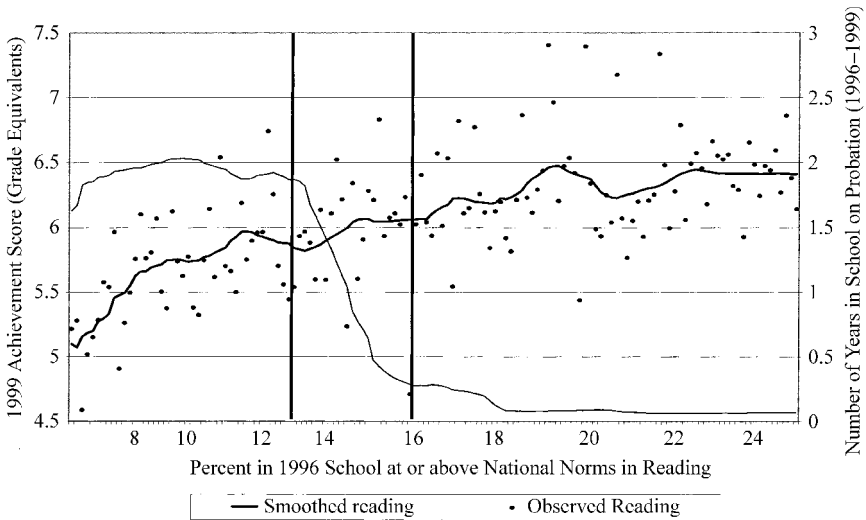
Figure 2a

The Relationship between 1996 School Reading Achievement and Subsequent Reading Performance

Notes: The figure displays data that were smoothed using lowess with a bandwidth parameter of 0.1. The distances between the percentages on the x-axis are uneven since there were some values that did not have any data.

As expected, we see that 1999 student achievement increases as a function of 1996 school mean achievement. However, the lines are relatively jagged, reflecting the fact that there are a limited number of schools at each level of school perfor-mance. The dark vertical lines at 13 and 16 percent bound the marginal area, where there is a sharp decline in the treatment. Average 1999 achievement increases stead-ily over this range, but does not appear to change discontinuously in reading or math, particularly in comparison to other jagged areas of the graph, for example, 18–20 percent and 11–13 percent). This suggests that the teacher training in Chicago did *not* have a substantial impact on student achievement.

Using the instrumental variables strategy described above, we can quantify our estimates of the treatment effect. In the baseline specifications, we limit our sample to students in low-performing schools where between 5 and 25 percent of students met national norms in 1996. We do so because the assumption of linearity between school reading achievement and student performance is most plausible in this nar-rower ranges of the data. In addition, by focusing on a narrow range around the cutoff, schools and students that receive treatment are likely to be comparable to their untreated counterparts. We later show that the results are robust to changes in the sample and model specification.

Table 4 presents the results of the first stage estimation. The dependent variable is the number of years a student attended a school on probation between the 1996–97 and 1998–99 school years (ranging from zero to three). Column 1 shows the
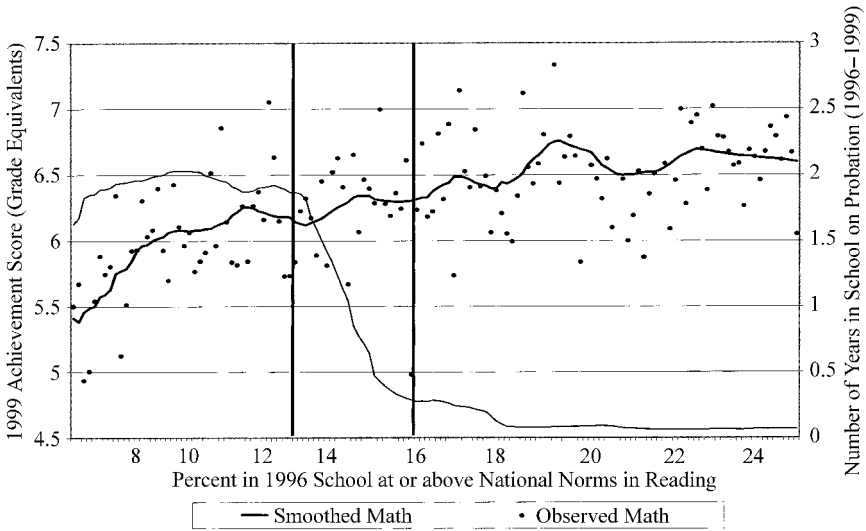
Figure 2b
The Relationship between 1996 School Reading Achievement and Subsequent
Math Performance

Notes: The figure displays data that were smoothed using lowess with a bandwidth parameter of 0.1.
The distances between the percentages on the x-axis are uneven since there were some values that did
not have any data.

results using a single dummy variable indicating whether the school was above or
below the 15 percent cutoff. Column 2 shows comparable results using a spline
specification. Note that all of the coefficients have the expected signs and the instru-
ments are highly predictive. The F-statistic that measures the joint significance of
the instruments in the first stage exceeds 100 regardless of whether we use an indica-
tor variable or a spline. The first stage R-squared is slightly higher, however, when
we specify the first stage relationship as a spline.

Table 5 presents the OLS and IV estimates. The OLS estimate in Column 1 of
-0.098 indicates that one additional year in a school on probation is associated with
a decrease of roughly one month of learning (0.10 GEs) in reading and math. How-
ever, we know from Table 3 that probation schools served a significantly more disad-
vantaged student population than other schools. When we control for a variety of
observable student and school characteristics, our estimates drop to one-fifth of this
size, although they remain negative and significant.

Columns 4 to 6 present the IV estimates in which we instrument for years in a
probation school using only a dummy variable that takes on a value of 1 if the
school surpassed the reading cutoff. Columns 7, 8, and 9 report the results from
specifications in which we instrument using a spline of school reading performance.
All of these estimates suggest that probation has no economically or statistically
important effect on reading and math achievement. Focusing on the results obtained
using the spline in the first stage, the coefficient with the largest absolute magnitude,

Table 4
The Effect of the 1996 School Reading Performance on the Average Years a Student Spends in a School on Probation between 1997 and 1999

Independent Variables	Dependent Variable Number of years a student spends in a school on probation from 1997 to 1999	
Discrete Cutoff Specification		
Above the probation cutoff (15 percent or more students at or above national norms in reading)	−1,400 (0.135)	—
Spline Specification		
Percent at or above national norms in reading (5–14 percent)	—	0.018 (0.026)
Percent at or above national norms in reading (14–15 percent)	—	−1.724 (0.124)
Percent at or above national norms in reading (15–25 percent)	—	−0.021 (0.010)
Number of schools	246	246
Number of observations	47,274	47,274
R-squared	0.608	0.642
F-statistic of Instruments	F = 106.5 [p = 0.000]	F = 105.7 [p = 0.000]

Notes: Sample includes students who were in schools in the fall of 1996 that had between 5 and 25 percent of students at or above national norms in reading. The first column shows the effect of being in a school at or above the probation cutoff on the number of years a student spends in a probation school controlling for percent at or above national norms in reading. In the second column we use a linear spline specification. In the second column, the *F*-statistic of the instruments takes into account that the instruments are jointly collinear with the second stage control variable "percent at or above national norms in reading." All test statistics are computed taking into account that observations within a school may not be independent. The regression includes the following variables that are not shown here: 1996 math and reading scores, student demographics including whether the student was included for test reporting purposes in 1996, age as of Fall 1996, race and gender indicators (Black, Hispanic, male, Black*male, Hispanic*male), free and reduced price lunch status, current and former bilingual status, special education, an indicator of whether the student was living with relatives or living in foster care, concentration of poverty and social status in the student's census block group, and school demographics including the attendance rate, percent Black, percent Hispanic, percent LEP, percent low income, mobility rate, truancy rate, and percent at or above national norms in math.

Table 5
IV and OLS Estimates of the Effect of Teacher Training on Student Achievement

Dependent Variables	OLS	OLS	OLS	IV Discrete	IV Discrete	IV Discrete	IV Spline	IV Spline	IV Spline
1999 reading scores	−0.098	−0.085	−0.020	−0.18	−0.017	−0.003	−0.026	−0.020	−0.007
[n = 47,274]	(0.019)	(0.018)	(0.010)	(0.33)	(0.030)	(0.024)	(0.029)	(0.027)	(0.022)
1999 math score	−0.093	−0.077	−0.021	−0.008	−0.005	0.010	−0.015	−0.006	0.005
[n = 47,118]	(0.018)	(0.017)	(0.010)	(0.035)	(0.031)	(0.028)	(0.032)	(0.028)	(0.026)
Student level covariates	No	No	Yes	No	No	Yes	No	No	Yes
School level covariates	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes

Notes: Sample includes students who were in schools in the fall of 1996 that had between 5 and 25 percent of students at or above national norms in reading. Student level covariates include 1996 math and reading scores, student demographics including whether the student was included for test reporting purposes in 1996, age as of Fall 1996, race and gender indicators (Black, Hispanic, male, Black*male, Hispanic*male), free and reduced price lunch status, current and former bilingual status, special education, an indicator of whether the student was living with relatives or living in foster care, concentration of poverty and social status in the student's census block group. School level covariates include the attendance rate, percent Black, percent Hispanic, percent LEP, percent low income, mobility rate, truancy rate, and percent at or above national norms in math.

−0.026, corresponds to roughly a 0.01 standard deviation effect. Note that the standard errors on the IV estimates in Column 6 are roughly 0.025, meaning that we could detect a positive significant result as small as 0.05 GEs. Considering that the average elementary student during this period gained roughly 0.90 GEs per year and the standard deviation of 1999 achievement scores in our sample was roughly 1.9 GEs, it does not appear that the teacher training and/or technical assistance provided to probation schools had any meaningful effect.

In addition, note that the results from Columns 4–9 show that our IV results are not sensitive to the inclusion of control variables. This suggests that after controlling for school reading performance, students in schools just above and below the cutoff have comparable observable characteristics. This lends more credence to the assumption that the unobserved characteristics of students in schools just above and below the cutoff are comparable as well. Finally, we see that there is no significant difference between the OLS and IV results presented in Columns 4 and 9. This suggests that, conditional on the set of student and school controls included in the models, probation waivers were not distributed on the basis of unobservable characteristics.

At this point, it is useful to consider once again what our estimated coefficient actually measures. The probation policy in Chicago was designed both to motivate low-achieving schools to improve student performance and to provide them with the support to do so. The incentive for teachers and administrators to improve was driven by the threat of reconstitution, which would have resulted in the reassignment and possible dismissal of all school staff. The support came in the form of the resources for professional development described earlier, along with a small degree of technical assistance, and monitoring designed to complement the training and enhance classroom instruction.

In assessing the overall impact of the probation policy, it is important to recognize that either, both, or neither aspects of the policy might have influenced student achievement. The probation incentives likely influenced a broad range of lower-performing schools, including those schools that initially scored above the cutoff, but perceived themselves to be in danger of scoring below the cutoff in subsequent years. Using an interrupted time-series design, Jacob (2002) finds evidence that the incentives generated by the probation policy did in fact lead to a substantial increase in performance among low-achieving schools in general.

The estimates in this paper measure the separate effect of the professional development resources provided by the probation policy. Note that contrary to the incentives, the resource provided under probation would only be expected to influence achievement in the schools that actually received the resources. By comparing students in schools that just missed the probation cutoff with those in schools that just made the cutoff, our strategy identifies resources, assistance, and monitoring provided to probation schools.

In interpreting the results, several points are important to keep in mind. First, the estimates capture the impact of teacher training within schools that faced significant incentives to improve student performance. Insofar as training in these circumstances is more or less effective than training provided without such incentives, our findings may differ from other evaluations of training programs. Second, our estimates incorporate the effect of any differential incentives faced by schools on either side of the cutoffs. Because schools that just missed being placed on probation in 1996 were

at risk of being placed on probation in subsequent years, they too had an incentive to increase student performance. However, to the extent that schools that were placed on probation in 1996 were one step closer to actual sanctions, they may have had even greater incentives to increase achievement. Importantly, because these factors are likely to operate in the same direction, if teacher training has any positive impact on student achievement, then students in schools placed in probation in 1996 should outperform students in schools that narrowly avoided probation that year. Given these circumstances, the fact that we consistently find zero effect reinforces the conclusion that the training simply did not influence student achievement.

B. Other Effects of Probation

As was mentioned previously, probation might influence student mobility and test-taking patterns. In particular, motivated families may want to remove their children from probation schools and probation schools may want to avoid testing the lowest ability children. Using the IV methodology described above, we examine the causal impact of being in a probation school in 1996–97 on the probability that a student changes schools, leaves the CPS, or fails to have an included test score. These results are found in Table 6.

The first row suggests that being in a probation school in 1997 has no significant effect on the probability of being enrolled in the CPS in 1999. Because there are few high achieving students in probation schools it is difficult to ascertain whether probation has a differential effect on the enrollment decisions of high-ability students. In the second row, we see that probation appears to increase the probability that a student changes schools by 1999. Furthermore, the point estimates are not trivial relative to the baseline mobility of 24 percent, particularly among the top ability quartile. Despite this, the coefficients are not significantly different from zero. The standard errors are particularly large for the high-ability students. Taken at face value, however, the point estimates suggest that probation may have induced high-ability students to change schools. Finally, it does not appear that being in a probation school is associated with changes in the probability that a previously tested student has test scores is included for evaluation. This holds even for students who are in the bottom of national reading distribution. This suggests that being put on probation does not cause administrators to discourage low-ability students from being tested or from having the test scores counted for school evaluation. Overall, probation may affect student decisions regarding school attendance within the CPS. However, there is no evidence that being in a probation school in 1997 causes students to leave the district, avoid testing, or have their scores excluded for evaluation purposes.

One might be concerned if probation causes high-ability students to change schools. It is unlikely, however, that this will bias our estimates. In particular, our unit of analysis is the student—not the school—and we control for a number of student level covariates. We also minimize problems associated with nonrandom school changes by utilizing variation in the number of years in a probation school attributable to the *observed* school level reading performance. Nevertheless, one might be concerned that high-ability students in schools placed on probation may have been induced to move to better schools far above the cutoff. In this case, we might attribute the benefits of being in a good school to probation. We would then

Table 6
IV Estimates of the Effect of Being in a Probation School in 1997

Dependent Variable	Sample Full Sample	1st Quartile of National Reading Distribution	2nd Quartile of National Reading Distribution	3rd Quartile of National Reading Distribution	4th Quartile of National Reading Distribution
Enrolled in CPS in 1999 [n = 53,767]	−0.002 (0.008)	−0.005 (0.009)	0.016 (0.011)	−0.022 (0.017)	−0.026 (0.040)
Changed schools by 1999 [n = 43,638]	0.034 (0.022)	0.039 (0.026)	0.031 (0.023)	0.014 (0.037)	0.069 (0.058)
Not tested in 1999 [n = 53,767]	0.004 (0.008)	0.013 (0.010)	−0.018 (0.012)	0.007 (0.018)	0.045 (0.041)
Excluded in 1999 [n = 47,075]	−0.015 (0.008)	−0.023 (0.012)	−0.002 (0.006)	0.002 (0.005)	−0.004 (0.005)
Not tested or excluded in 1999 [n = 53,767]	−0.009 (0.010)	−0.008 (0.014)	−0.018 (0.011)	0.010 (0.019)	0.038 (0.041)

Notes: Baseline sample includes students who were in schools in the fall of 1996 that had between 5 and 25 percent of students at or above national norms in reading. Also included (but not shown) in the regression specification are the controls described in Table 4. All estimates are computed using two stage least squares. The instruments and controls are as previously indicated.

expect probation to look better for high-ability students than for low-ability students. We show later that this is not the case. Furthermore, if probation caused students to move to better schools, we would expect the movers who started in probation schools to perform better than the movers from nonprobation schools. We also will show that this is not the case.

C. Heterogeneous Effects and Robustness Checks

Table 7 examines the heterogeneity of effects by student age, ability, and other demographic characteristics. Each row corresponds to a separate regression that includes only students in the subgroup listed. The cells contain IV estimates of the effect of the number of years in a probation school on 1999 achievement. The top panel shows that probation has no effect on student performance in any grade from third to sixth in either reading or math. The second panel shows separate effects for students at different points in the ability distribution in spring 1996. Because probation is determined by the percent of students who score above the 50th percentile, the policy creates an incentive for schools to focus attention on students near this point, since they are more likely to meet this standard with sufficient support. However, we see that probation does not appear to have any larger effect on students in the second and third quartiles than on students at the extremes of the ability distribution. The third panel shows no difference in impact across race, gender, or SES.

Table 8 displays results from a number of alternative specifications. The first row presents the original estimates from Table 4 as a basis for comparison. In Row 2, we include students who did not have initial test scores. For this specification, we set missing scores to zero and included a dummy variable indicating that the test score was missing. The results still suggest that being in a probation school had no effect on performance. In Row 3, we include students from all schools—regardless of the average school reading performance in 1996—which should increase the efficiency of our estimates. Once again we find probation to have no significant effect on reading and math achievement. When we include third order polynomials of school performance in Row 4, we find that our results do not significantly differ from the case in which we assume linearity. Row 5 shows that the inclusion of polynomials also does not change the result if we instrument using a dummy variable indicating position relative to the cutoff. In Row 6, concerned about the potential endogeneity of waivers in the marginal area, we drop schools in the marginal area, control for third order polynomials, and instrument using only an indicator variable. The results remain unchanged. In Row 7, we include a cohort of students from 1993, prior to the introduction of the school reforms. While no students or schools in this cohort received the treatment, we can use these data to make certain our findings are not driven by nonlinearity in the relationship between school reading performance and student achievement. The instruments in these models are the interaction between the splines of school reading performance and cohort. We see that the estimates do not change for either reading or mathematics.

Many of the schools that scored just above the probation cutoff, or were waived from probation, were placed on remediation. These schools did not receive the same close monitoring or financial support as probation schools, but they were subject to somewhat heightened oversight. To check whether this heightened oversight may

Table 7
The Effect of Probation on Student Achievement

	Probation Treatment Effect	
Subgroup	1999 Reading Score	1999 Math Score
Grade Level		
Third grade	0.053	0.038
	(0.036)	(0.040)
Fourth grade	−0.017	−0.001
	(0.033)	(0.032)
Fifth grade	−0.039	0.001
	(0.036)	(0.037)
Sixth grade	−0.014	0.000
	(0.037)	(0.041)
Prior Achievement		
1st quartile national reading distribution	−0.017	0.018
	(0.023)	(0.026)
2nd quartile national reading distribution	0.021	−0.000
	(0.031)	(0.031)
3rd quartile national reading distribution	−0.022	−0.022
	(0.036)	(0.041)
4th quartile national reading distribution	0.036	0.014
	(0.064)	(0.056)
Race, Gender, and SES		
Black	−0.007	−0.004
	(0.025)	(0.029)
Hispanic	0.025	0.054
	(0.045)	(0.047)
White/other	−0.287	−0.049
	(0.149)	(0.106)
Male	−0.026	0.009
	(0.023)	(0.026)
Female	0.009	0.000
	(0.027)	(0.029)
Free lunch	−0.009	0.008
	(0.022)	(0.026)
No free lunch	0.014	−0.011
	(0.055)	(0.046)

Notes: Sample includes students who were in schools in the fall of 1996 that had between 5 and 25 percent of students at or above national norms in reading. Also included (but not shown) in the regression specification are the controls described in Table 4.

have impacted achievement, row eight examines whether being in a school on proba-tion *or* remediation has any effect on academic achievement.[12] We find no effect.

Even after schools were taken off probation, they were required to maintain a relationship with their external partner for an additional year. Also, some low per-forming schools that were not placed on probation chose to hire an external partner even though it was not officially required. To test whether the presence of an external partner, rather than simply being on probation, influenced achievement levels in low-achieving elementary schools, Row 9 of Table 8 shows the estimated effect of being in a school with an external partner. This effect is not statistically different from zero.[13]

One might argue that the monitoring and staff development that probation schools receive should not have an observable impact on student achievement for several years. The next rows in Table 8 explore this possibility. In Row 10, we examine the effect of probation on students who remained in the same school between 1996 and 1999. Thus, in this sample, the students in probation schools received three full years of treatment. If one believes that probation has a greater impact for students who spend an extended period in the school, then these estimates should be larger than the original estimates. However, it appears that even these students received no significant benefit from being in a school on probation. It is possible that reforms instituted by the external partners and probation managers took a year or two to become effective, in which case one would not expect any impacts until the 1998–99 school year. By examining the three-year period, we will observe a small, diluted effect. To explore this possibility, Row 11 shows the effect of probation on 1998–99 gains, but still finds no effect. Row 12 shows the effect of probation on 1998–99 gains for the subsample of students who remained in the same school between 1996 and 1999. The point estimates are not statistically different than zero.

The final row of Table 8 shows the effect of being in a probation school on those who eventually *change* schools. We perform this check to ensure that those leaving probation schools were not better in unobserved ways than those leaving nonproba-tion schools. We are also concerned that probation may have induced students to attend high quality schools far from the probation cutoff. In Row 13, we see that probation does not appear related to the performance of those students who left.

A wide variety of nonprofit organizations and universities worked with probation schools in order to improve student achievement. These external partners varied considerably in their institutional affiliation (universities versus private organiza-tions), programmatic focus (school organization versus staff training versus curricu-lum development), and educational philosophy (whole language versus direct in-struction). It is possible that some external partners were more effective than others, which might explain the weak aggregate effects that we find. Note, however, that

---

12. For this specification, our instrument is a single dummy variable indicating whether 15 percent or more students in a school performed at or above national norms in reading in 1996, rather than the spline in 1996 reading achievement that is used in the baseline specification. We do so because the discontinuity between 1996 school mean reading achievement and the average years in a school on probation *or* remedia-tion is extremely sharp at the official probation cutoff of 15 percent, in contrast to the baseline specification that uses years on probation alone.

13. Once again, our instrument is a single dummy variable that indicates whether at least 15 percent of students in a school performed at or above national norms in reading in 1996.

Table 8
Robustness Checks

Row	Sample	Dependent Variable	Treatment	Reading	Math
Original Estimates					
1	Baseline sample (students in 1996 schools where 5–25 percent of students met national norms in reading)	1999 Achievement	Years in school on probation	−0.007 (0.022)	0.005 (0.026)
Alternative Samples & Specifications					
2	Including students with missing initial test scores	1999 Achievement	Years in school on probation	−0.001 (0.022)	0.009 (0.025)
3	All schools	1999 Achievement	Years in school on probation	−0.012 (0.013)	−0.003 (0.015)
4	Baseline sample (controlling for third order polynomials of school performance)	1999 Achievement	Years in school on probation	−0.016 (0.029)	0.027 (0.036)
5	Baseline sample (controlling for polynomials of school performance and instrumenting using a cutoff dummy)	1999 Achievement	Years in school on probation	−0.009 (0.035)	0.038 (0.044)
6	Dropping schools in marginal area (controlling for polynomials of school performance and instrumenting using a cutoff dummy)	1999 Achievement	Years in school on probation	−0.008 (0.023)	0.007 (0.026)

7	All schools and cohort of students in school in 1993 in schools where 5–25 percent of students met national norms in reading	1999 Achievement	Years in school on probation	0.024 (0.015)	0.018 (0.015)
8	Baseline sample	1999 Achievement	Years in school on probation or re-mediation	−0.004 (0.029)	0.011 (0.034)
9	Baseline sample	1999 Achievement	Years in school with external partner	−0.003 (0.027)	0.011 (0.032)
10	Students in same school from 1996 to 1999	1999 Achievement	Years in school on probation	−0.002 (0.025)	0.006 (0.028)
11	Baseline sample	1998–99 Achievement gain	Years in school on probation	−0.035 (0.018)	−0.027 (0.016)
12	Students in same school from 1996 to 1999	1998–99 Achievement gain	Years in school on probation	−0.026 (0.020)	−0.019 (0.017)
13	Students who change schools between 1996 and 1997	1999 Achievement	In probation school in Fall 1996	0.012 (0.049)	0.032 (0.042)

Notes: Baseline sample includes students who were in schools in the fall of 1996 that had between 5 and 25 percent of students at or above national norms in reading. Also included (but not shown) in the regression specification are the controls described in Table 4. All estimates are computed using two stage least squares. The instruments are as previously indicated except in the specifications in which the treatment is defined as years in school on probation or remediation and years in school with external partner. For these specifications, the instrument is a dummy variable indicating the school surpassed 15 percent at or above national norms in reading.

Table 9
OLS Estimates of the Effect of Specific External Partners on Student Achievement

	Treatment Effect of Year with External Partner on 1999 Achievement	
External Partner	Reading	Math
America's choice—NARE	0.022	0.020
	(0.031)	(0.031)
Malcolm X College	−0.001	0.017
	(0.017)	(0.019)
School achievement structure	−0.001	0.010
	(0.018)	(0.022)
DePaul University	0.021	−0.024
	(0.020)	(0.024)
Northeastern Illinois University	0.004	0.027
	(0.025)	(0.024)
North Central Regional	−0.019	−0.038
	(0.024)	(0.033)
Other external partner	0.023	0.027
	(0.021)	(0.023)
F-test of joint significance	F = 0.68	F = 1.77
(p-value)	[p = 0.6875]	[p = 0.0943]
Observations	47,274	47,118

Notes: The sample includes students who were in schools in the fall of 1996 that had between 5 and 25 percent of students at or above national norms in reading. We use all controls from the baseline (Table 4) specification as well as the total number of years a student was enrolled in a school on probation from 1997 to 1999.

the zero net effect implies that if some external partners increased student performance then others must have *decreased* student achievement levels. Table 9 examines the probation effects for several of the largest external partners. Because schools were largely free to select their external partner, these estimates cannot be interpreted as causal effects, although they still may provide some insight. Nonetheless, it does not appear that any of the major external partners had a significant impact on student achievement in the probation schools.

VI. Conclusions

In an effort to improve student achievement in Chicago in the mid-1990s, the CPS placed nearly 20 percent of the lowest-achieving elementary schools in the city on probation. The financial and technical support provided to probation

schools was dedicated specifically to improving classroom instruction, primarily through teacher training and staff development. Indeed, teachers in probation schools reported moderate increases in the frequency with which they attended professional development activities as well as more substantial increases in the quality of the professional development they received.

The preceding analysis, however, indicates that the training provided to teachers in probation schools had no discernable effect on student achievement. These results are robust to a variety of alternative specifications and do not differ across student ability, gender, race, or family income. While consistent with much of the earlier research on teacher training in the United States, these findings differ from recent work by Angrist and Lavy (2001).

Although it may not be surprising that different programs in different settings have different effects, it is useful to examine some of the possible explanations for the discrepancies in order to better understand how the results from each study might be generalized. Several differences stand out between the Chicago and Jerusalem programs. First, the Chicago program was implemented in a group of extremely high-poverty, low-achieving schools. In contrast, the program in Jerusalem took place in mostly middle- to lower-middle class neighborhoods, which included a combination of some upper-middle class schools attended by children of Hebrew University faculty as well as some poorer schools attended by immigrants (Angrist 2001). Second, the training provided in the Jerusalem schools was highly structured and closely aligned with the school curriculum whereas the training in Chicago was relatively unstructured and less well aligned. Finally, the training in Jerusalem was complemented by direct services to students in the form of after school learning centers and other programs for immigrant families.

It is also useful to put the magnitude of Chicago probation expenditures into perspective. Smylie et al. (2001) report that the CPS budgeted $75 million for professional development in the 1997–98 school year. This represented about 2.5 percent of the district's total expenditures. If teacher development expenditures were divided equally among grades (first to twelfth), then approximately $50 million would have been spent on elementary schools and average expenditures per elementary school would have been about $108,000. If we use this as a rough baseline for professional development expenditures, the additional financial resources that were available under the probation policy seem substantial.

While meaningful from the perspective of previous expenditures, the magnitude of new resources devoted to training still may have been insufficient to generate noticeable achievement gains. Finnigan et al. (2001) highlight the low-intensity level of the probation policy, arguing that it is one of the primary limitations of the support system. In the probation schools they studied, external partners spent on average one to two person days per week in each school. It is likely that this was simply inadequate to address the substantial needs in these extremely low-achieving schools.

In this light, one might interpret the findings of these two studies as showing that teacher training can have a significant, positive impact on student achievement under generally favorable conditions, but that such benefits depend on the context and quality of the program. Unfortunately, national data suggest that the frequency and nature of professional development activities in Chicago is comparable to other school districts in this country (Parsad et al. 2001). Thus, our findings suggest that moderate increases

in the intensity of the professional development efforts along the lines of the Chicago program will likely fail to improve the achievement of students in failing schools.

In conclusion, as a treatment designed to improve the performance of children in failing schools, the teacher training provided to probation schools in Chicago appears completely ineffective. While it is possible that the services offered to probation schools may have other positive outcomes that are not captured in student test scores, such as improved working relations among school staff, administrators would be well advised to rethink the current program. More generally, educators and school administrators should carefully examine the nature of teacher professional development in this country.

References

Angrist, Joshua D. 2001. Personal communication.

Angrist, Joshua D., and Victor Lavy. 1999. "Using Maimonides Rule to Estimate the Effect of Class Size on Scholastic Achievement." *Quarterly Journal of Economics* 114(2): 535–75.

———. 2001. "Does Teacher Training Affect Pupil Learning? Evidence from Matched Comparisons in Jerusalem Public Schools." *Journal of Labor Economics* 19(2):343–69.

Berk, Richard A., and David Rauma. 1983. "Capitalizing on Nonrandom Assignment to Treatments: A Regression-Discontinuity Evaluation of a Crime-Control Program." *Journal of the American Statistical Association* 78(381):21–28.

Black, Sandra. 1996. "Do Better Schools Matter? Parents Think So!" Working Paper. Cambridge, Mass.: Harvard University.

Bressoux, Pascal. 1996. "The Effect of Teachers' Training of Pupils' Achievement: The Case of Elementary Schools in France." *School Effectiveness and School Improvement* 7(3):252–79.

Cohen, David F., and Heather C. Hill. 2000. "Instructional Policy and Classroom Performance: The Mathematics Reform in California." *Teachers College Record* 102(2):294–343.

Corcoran, Thomas B. 1995. "Helping Teachers Teach Well: Transforming Professional Development." CPRE Policy Briefs. Consortium for Policy Research in Education. Philadelphia: University of Pennsylvania.

Dildy, Peggy. 1982. "Improving Student Achievement by Appropriate Teacher In-Service Training: Utilizing Program for Effective Teaching (PET)." *Education* 102(2):132–38.

Finnigan, Kara, Jennifer O'Day, and David Wakelyn. 2001. "Buddy, Can You Lend Us A Hand? The Provision of External Assistance to Chicago Elementary Schools on Probation." Unpublished. Madison: University of Wisconsin-Madison.

Guryan, Jonathan. 2000. "Does Money Matter? Regression Discontinuity Estimates from Education Finance Reform in Massachusetts." Working Paper. Chicago: University of Chicago.

Hanushek, Eric A. 1996. "School Resources and Student Performance." In *Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success,* ed. Gary Burtless, 43–73. Washington, D.C.: Brookings Institution Press.

Hanushek, Eric A., John F. Kain, and Steven G. Rivkin. 2001. "Disruption versus Tiebout Improvement: The Costs and Benefits of Switching Schools." Working Paper. Stanford: Stanford University.

Hedges, Larry V., and Rob Greenwald. 1996. "Have Times Changed? The Relation between School Resources and Student Performance." In *Does Money Matter? The Effect*

of School Resources on Student Achievement and Adult Success, ed. Gary Burtless, 74–92. Washington, D.C.: Brookings Institution Press.

Jacob, Brian A. 2002. "Accountability, Incentives and Behavior: The Impact of High-Stakes Testing in the Chicago Public Schools." National Bureau of Economic Research. Working Paper no. 8968. Cambridge, Mass.: Harvard University.

Jacob, Brian A., and Steven D. Levitt. 2002. "Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating." Working Paper. Cambridge, Mass.: Harvard University.

Jacob, Brian A., and Lars Lefgren. 2001. "Remedial Education and Student Achievement: A Regression-Discontinuity Analysis." National Bureau of Economic Research. Working Paper no. 8918. Cambridge, Mass.: Harvard University.

Hahn, Jinyong, Petra Todd, and Wilbert Van der Klaauw. 1999. "Evaluating the Effect of an Antidiscrimination Law Using a Regression-Discontinuity Design." Working Paper no. 7131. Cambridge, Mass.: National Bureau of Economic Research.

Kennedy, Mary M. 1998. "Form and Substance in In-service Teacher Education." Research Report from the National Institute for Science Education. Madison: University of Wisconsin.

Kerbow, David. 1996. "Patterns of Urban Student Mobility and Local School Reform." *Journal of Education for Students Placed at Risk* 1(2):147–69.

Ladd, Helen F. 1999. "The Dallas School Accountability and Incentive Program: An Evaluation of its Impacts on Student Outcomes." *Economics of Education Review* 18(1): 1–16.

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

Little, Judith W. 1993. "Teacher's Professional Development in a Climate of Educational Reform." *Educational Evaluation and Policy Analysis* 15(2):129–51.

Parsad, Basmat, Laurie Lewis, Elizabeth Farris, and Bernie Greene. 2001. "Teacher Preparation and Professional Development 2000." National Center for Educational Statistics, U.S. Department of Education. NCES 2001-088. Washington D.C.

Richards, Craig E., and Tian Ming Sheu. 1992. "The South Carolina School Incentive Reward Program: A Policy Analysis." *Economics of Education Review* 11(1):71–86.

Roderick, Melissa, Brian Jacob, and Anthony Bryk. 2000. "Evaluating Chicago's Efforts to End Social Promotion." In *Governance and Performance: New Perspectives,* ed. L. Lynn and C. Heinrich, 34–67. Washington, D.C.: Georgetown University Press.

Smylie, Mark A., Elaine Allensworth, Rebecca C. Greenberg, Rodney Harris, and Stuart Luppescu. 2001. "Teacher Professional Development in Chicago: Supporting Effective Practice." Consortium on Chicago School Research report. Chicago: University of Chicago.

Thistlethwaite, Donald and Donald Campbell. 1960. "Regression-Discontinuity Analysis: An Alternative to the Ex-Post Facto Experiment." *Journal of Educational Psychology* 51(6):309–17.

Trochim, William. 1984. *Research Design for Program Evaluation: The Regression-Discontinuity Approach*. Beverley Hills: Sage Publications.

Wiley, David E., and Bokhee Yoon. 1995. "Teacher Reports on Opportunity to Learn: Analyses of the 1993 California Learning Assessment System (CLAS)." *Educational Evaluation and Policy Analysis* 17(3):355–70.