PUBLIC POLICY INSTITUTE OF CALIFORNIA

Technical Appendix

Does Diagnostic Math Testing Improve Student Learning?

Julian R. Betts, Youjin Hahn, Andrew C. Zau

Supported with funding from the Donald Bren Foundation

Introduction

This technical appendix provides details on the data sources and statistical methods used to estimate the effects of MDTP testing on student outcomes. The appendix includes many robustness tests that are mentioned briefly in the main text.

Readers who want only to find the underlying source for the estimated effects of mandated MDTP (as shown in Figures 3 and 4 in the report) can find the underlying estimates in Table A.4, column 3, and Table A.6, column 3, respectively. The text describing those tables also explains how we convert the estimated effects into the percentile changes presented in the report.

The sections below mirror the sections in the report, and provide a complete and stand-alone summary of the research undertaken. There is thus some overlap with the text and tables in the report, which we reproduce here for the sake of the reader.

Data

Student Longitudinal Database

Our longitudinal data consist of complete student academic records, including test scores, academic grades, courses taken and absences, from fall 2001 through spring 2007. The data include indicators of MDTP tests taken in a given year, as well as a rich set of covariates related to the student's class size and teacher qualifications (overall in elementary school, and in math class in middle and high school).

Teacher qualifications in the San Diego Unified School District (SDUSD) fall under three main areas: education, credentialing, and years of experience. For most teachers, our dataset includes the degree attained (bachelor's, master's, or doctorate) and subject major. Credentials can be divided into two types: overall credential, and subject authorization. The credential indicates that the teacher has completed requirements thought to help a person operate a classroom. A math subject authorization, in contrast, indicates that a teacher has taken a requisite set of college math courses. For elementary teachers, subject authorization is not necessary; a multiple subject credential is sufficient to teach within a self-contained classroom. Secondary school teachers usually obtain an authorization to teach specific subjects, which requires that teachers complete prescribed college courses in the given subject. Authorizations fall under four levels. In declining order, these are full, supplementary, board resolution, and limited assignment emergency.[1] We control for the level of overall credential and the math subject authorization of the student's homeroom teacher in elementary grades and the student's math teacher in middle and high school.

California has administered various standardized tests at different times. It mandated the Stanford 9 test from spring 1998 through spring 2002, and the California Standards Test (CST) in spring 2002 and later

---

[1] For further details on the credentialing system, see Julian Betts, Andrew Zau, and Lorien Rice, *Determinants of Student Achievement: New Evidence from San Diego* (San Francisco: Public Policy Institute of California, 2003).

years. Our outcome of interest is a gain in standardized math CST scores, and we use CST data from spring 2002 through 2007. We convert CST scores into Z-scores by subtracting the district-wide mean and dividing by the district-wide standard deviation for a given grade and subject.

Further Information on Math Diagnostic Testing in San Diego

Beginning in 1999-2000, the district mandated the use of at least one MDTP test at the end of the school year. This mandate was phased out during our sample period in higher grades and implemented for the first time about half way through our sample period in lower grades. In addition, students are subject to testing in only a few grades. Thus we are in a position to compare individual student trends in achievement during years and grades with and without testing. Table A.1 shows the way in which SDUSD mandated testing, while Table A.2 shows the actual proportions of students in each grade and school year that took an MDTP diagnostic test. Table A.2 suggests fairly high, but not complete, compliance with the district mandate. We return to this point later.

TABLE A.1
SDUSD mandated use of MDTP tests by school year, test, and grade level

Year	MDTP readiness test	Grade levels
1999	Geometry	8 and 9
2000	Geometry	8 and 9 in the spring, and grade 8 students enrolled in Algebra during summer school
2001	Geometry	8 and 9
2002	Geometry	8 and 9
2003	Algebra	7
2004	Prealgebra	6 at selected schools
2004	Algebra	7
2005	Prealgebra	6
2005	Algebra	7
2006	Prealgebra	6
2006	Algebra	7

SOURCE: Authors' calculations.

NOTE: Year refers to the start of the school year (e.g., 2006 refers to the 2006–2007 school year). This information was provided by Bruce Arnold, the Coordinator of MDTP for the San Diego area.

TABLE A.2
Proportion of students who took MDTP in a given grade and year

	Grade				
Year	6	7	8	9	10
2001–2002	0	0.03	0.81	0.51	0.04
2002–2003	0	0.03	0.6	0.43	0.03
2003–2004	0	0.88	0.02	0	0
2004–2005	0.23*	0.91	0	0	0
2005–2006	0.95	0.88	0	0	0
2006–2007	0.63	0.63	0.01	0	0

NOTE: Calculated using the San Diego Unified School District data. Numbers in bold indicate the mandated years.
* In school year 2004–2005, MDTP was mandated for 6th-year students in selected schools.

In addition to the mandatory testing, individual math teachers have voluntarily adopted MDTP testing early in the school year. Beginning in 2000, the district mandated the use of at least one MDTP test at the end of the school year. Since then, some teachers have continued to provide testing at the start of the school year. Those results are included in this study. The dates range from 2003 to 2007. A total of 165,863 tests were scored during this time period. Attempts were made to match those tests with actual student data based on student identifier, birthday, name, and teacher. The nature of scanned, voluntary data leads to some students who cannot be identified because of insufficient data on identifiers. Approximately 80 percent of the voluntary tests could be matched to student data.

Constructing the Outcome Variable

We convert the CST score into Z-scores. Up until grade 7, one type of CST is offered in each grade. For those students in grade 7 and below, we standardize the CST score by subtracting the district-wide mean and dividing by the district-wide standard deviation for a given grade, calculated using scores available over the years we study.

Starting in grade 8, students take different versions of the math CST, depending on the subject matter they study. Thus, for the students in grade 8 and above, we standardize the CST score by grade and type of test. Table A.3 presents the seven most common sequences of course-taking for cohorts who were in the 8th grade in 2001–2002. Up until the 7th grade, students take similar courses. However, in the 8th and especially the 9th grade, students start diverging in terms of their courses-taking behavior. For instance, some students in 9th grade repeat Algebra 1, while others take Geometry or Algebra 2. This diversity in course-taking behavior implies that standardizing the test score by type of test and grade is required for upper-grade students.

TABLE A.3
Seven popular course sequences

Grade						
8	**9**	**10**	**11**	**Freq.**	**Percentage**	**Cum.**
AL1	Gmtry	AL2	HsMath	1269	30.71	30.71
AL1	AL1	Gmtry	AL2	956	23.14	53.85
AL1	AL1	Gmtry	I. Math	397	9.61	63.46
AL1	AL1	Gmtry	Gmtry	250	6.05	69.51
AL1	Gmtry	AL2	AL2	177	4.28	73.79
Gmtry	AL2	HsMath	HsMath	153	3.7	77.49
AL1	AL1	AL1	Gmtry	110	2.66	80.15

NOTE: We followed the cohort that was in the 8th grade in the 2001-2002 academic year up to when it was in the 11th grade in 2004–2005. Subject abbreviations: AL1 = Algebra 1; AL2 = Algebra 2; Gmtry = Geometry; HsMath = High school math; I. Math = Integrated Math.

A second approach to standardizing is to standardize not just by grade and CST subject, but also by year. Thus, growth in a student's test score refers to his or her change in relative test scores compared to other students in the same grade, taking the same subject-matter CST test, and in the same year. This is a severe approach to take because it completely removes any trend over time in absolute levels of achievement (the test scores have mean zero each year). We prefer the models that study changes in Z-scores when standardized by grade and subject, but not by year, because if the mandated MDTP testing had systemic effects on achievement in a given grade, these trends will be completely removed and therefore undetectable in models of such scores.[2]

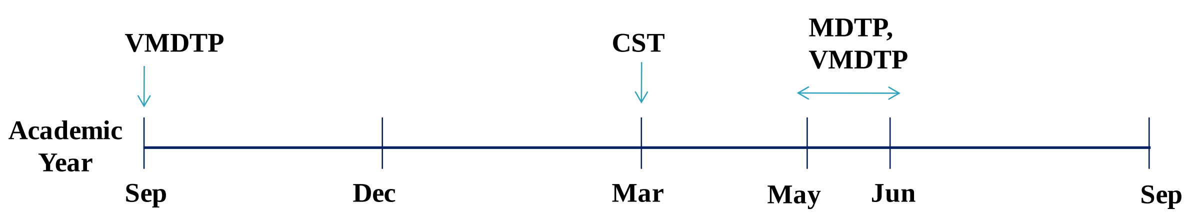
Empirical Strategy

Fixed Effect Approach

We model gains in test scores as a function of student's and family characteristics, teacher quality, school effects, and an error term.

As Figure A.1 shows, the mandated use of MDTP took place in May or June of the academic year, while the CST is offered in March. Thus, we relate the outcome variable, the change in standardized CST score between year t-1 and t, to the indicator of whether a student took MDTP in year t-1. We later discuss the voluntary use of the MDTP by individual teachers ("VMDTP"). The voluntary tests were clustered in September, May, and June, as shown in the figure.

FIGURE A.1
Timing of the MDTP and CST



NOTE: This figure shows the most typical months for statewide CST testing, district-mandated testing using MDTP tests ("MDTP" above), and voluntary testing by individual math teachers ("VMDTP" above).

The basic model is the following. Below, MDTP refers to district-mandated MDTP testing.

$$Y_{icgst} - Y_{icgs,t-1} = \alpha_i + \beta_s + MDTP_{i,t-1}\Omega + Family_{it}\Phi_0 + Class_{icst}\Phi_1 + School_{ist}\Phi_2 + Teacher_{icst}\Phi_3 + Grade_i\Phi_4 + Year_t\Lambda + Test_{it}\Gamma + \varepsilon_{it} \quad (1)$$

Y_{icgst} is the outcome of interest, a standardized test score in the California Standards Test (CST) for student i in (math) classroom c, grade g, school s, in year t. The coefficient α_i is a student fixed effect and β_s is a school fixed effect. As using the student fixed effect (α_i) removes variations that are constant within each student, from both observed and unobserved sources, the MDTP effect is identified by within-student variation over time. The coefficient of lagged MDTP, Ω, is of primary interest in this study. When student fixed effects are included, the coefficient of MDTP indicates the level by which district-mandated MDTP testing shifts individual students' test scores from their average growth trajectory.

The other control variables are as follows. A vector $Family$ contains parental education and $Class$ has an annual average math class size for each student. The vector $School$ contains the percentage of students on

[2] In models for which we standardize by grade, subject, and school year, our results on mandated MDTP testing are similar in terms of statistical significance, but slightly more muted in terms of coefficients. Smaller coefficients are to be expected given that standardization in this way removes almost the entire trend in achievement to which systemic testing might have contributed. Results are available upon request from the authors.

free lunch; percentage of students who are Asian, White, Hispanic, and other race; and percentage of students who are English learners. These are constructed by averaging variables for each student in a given year across semesters.

We also include various teacher characteristics. The vector *Teacher* includes the proportion of teachers for each student in a given year across semesters who have credentials (intern, emergency credential, full credential); math subject authorization (full, supplemental, board resolution, limited assignment emergency); years of teaching experience; Master's degree; Bachelor's in math; the Cross-cultural Language and Academic Development credential (CLAD); and indicators for teachers who are black, Asian, Hispanic, other non-whites, and female. Grade consists of dummy variables for each grade. *Year* dummies are included to account for any changes in the educational system affecting all students in the district. Lastly, we control for the type of tests taken by adding Test dummies for the version of the math CST test that the student took in the given year; Algebra 1, Algebra 2, Math 8/9, Geometry, High School Math, and Integrated Math.

Results

Baseline results

Table A.4 shows the main results. In addition to the regressors listed earlier, we include several variants of the indicator for whether the student took a district-mandated MDTP test. (Table A.a, in the Supplementary Tables at the end of this appendix, shows the means and other summary statistics for all of the regressors.) In the first model, we condition on $MDTP_{t-1}$, thus testing whether, for the individual student, having taken a mandated MDTP test in May or June of the prior year is associated with higher test score gains on the math CST between March of that year and March of the current year. The MDTP coefficient is significant at the 1 percent level and fairly large at 0.22, indicating that a student moves up in the distribution by about 0.22 of a standard deviation after having taken the MDTP.

In the next column we add $MDTP_{t-2}$, an indicator for having taken the MDTP as part of the district mandate two years earlier. This specification suggests that there may be some dropoff from having taken the MDTP test two years ago: Approximately 30 percent of the one-year gain dissipates in year two. There are of course many explanations for such a pattern, but one that seems logical is that math teachers pay particular attention to information from the prior spring's MDTP administration to learn about a student's specific strengths and weaknesses as they relate to the upcoming course. (Recall that the MDTP provides "readiness" tests that examine a student's preparation for a specific upcoming course, for instance through the Algebra Readiness test.) Whatever gains accrue to the student from this additional attention partly wear off the next year, perhaps in part because it is unusual for math teachers to look at MDTP test results from earlier than the prior spring, and in part because whatever remediation the teacher in year t provides after an MDTP test in year t-1 is not fully permanent.

TABLE A.4
Estimated effect of mandated MDTP on gains in Math CST scores

	(1)	(2)	(3)	(4)
MDTP t-1	0.219***	0.196***	0.184***	
	(0.019)	(0.017)	(0.017)	
MDTP t-2		-0.059**	-0.076***	
		(0.025)	(0.027)	
MDTP t-1 * MDTP t-2			0.097**	
			(0.042)	
Pre-algebra t-1				0.148***
				(0.030)
Algebra t-1				0.190***
				(0.034)
Geometry t-1				0.240***
				(0.037)
Observations	342571	311415	311415	342571

NOTE: Estimates are based on equation (1) in the text. Dependent variable is Gain in CST test score, standardized by Grade and Test. All regressions include student, grade, year, CST subject type, and school fixed effects. Standard errors are clustered by school and shown in parentheses.

Significance indicators: * p<0.10, ** p<0.05, *** p<0.01.

In the third column of Table A.4, we add an interaction between the indicators for whether the student took the MDTP one and two years ago. Our hypothesis was that continued diagnostic testing would be more likely to lead to sustained gains. As Table A.2 shows, this interaction effect is identified by two specific cohorts of students: those who were in grade 6 in 2004-2005 and those who were in grade 8 in 2001-2002.[3] The results suggest some complementarity between testing one and two years ago, with the effect about half as large as the estimated effect of being tested last year only.

We use the results in column 3 of Table A.4 as our main specification for the report. Note that it implies a higher rate of depreciation two years after the MDTP test, about 40 percent compared to 30 percent in the simpler model in column 2. However, a student who is tested two years in a row has a higher predicted score than a student who is given an MDTP test only once.

For the report, we converted the predicted effect from column 3, which is in terms of standard deviations, into the predicted change in a student's percentile ranking, under the assumption that the student starts at the 50th percentile and would remain there in future years if not for the effect of MDTP testing. To give one example, the effect of being tested once in the prior year on current test scores is 0.184 in the model in column 3, Table A.4. Thus the predicted percentile for this student the next year is $100*\Phi(0.184)$, where Φ is the cumulative distribution function for a standardized normal distribution. As another example, the predicted percentile for a student after two consecutive years of MDTP testing is $100*\Phi(0.205)$, where 0.205 is

[3] For a cohort to contribute to the identification of the interaction term, students must have been in grades and years mandated for testing two years in a row. Moving diagonally to the "southeast" in Table A.2 shows that the two cohorts listed above were mandated for testing two years in a row. A third cohort, those in grade 6 in 2005-2006, were also mandated for testing in two consecutive years, but we lack the spring 2008 test-score data needed to explore the effects of having been tested two years in a row in 2006 and 2007.

the sum of the coefficients on MDTP lagged once, lagged twice and the interaction term between MDTP lagged once and twice.

To discover whether only one or two of the specific MDTP readiness tests is responsible for the positive coefficient on $MDTP_{t-1}$, we show our model that replaces $MDTP_{t-1}$ with variables indicating whether students in the prior academic year had taken the Pre-Algebra, Algebra, or Geometry readiness tests. As shown in the last column of Table A.4, all three coefficients are positive and significant at the 1 percent level. Geometry readiness test has the greatest effect, followed by Algebra and Pre-Algebra readiness tests.

Although not reported, the estimates are virtually unaffected after dropping the set of explanatory variables, teacher characteristics, and class size over concerns that they might be endogenous.

Identification Issues: Incomplete Take-up

Table A.2 above shows the proportion of students who took the MDTP for a given year and grade. The average compliance rate for the cohorts who are affected by the mandate is about 0.7, reasonably high but not complete.

Even though the MDTP in the SDUSD region became mandatory for certain cohorts so that there was less likely to be a selection issue than if students were volunteered for testing by their schools, we are concerned that students who took the MDTP might be different in certain unobserved ways, thus biasing the effects. The reason that we use the student fixed-effect approach instead of directly comparing those students who took and did not take the MDTP is to address this possible selection issue. Identification comes from comparison of individual students' test score changes in years in which they were tested using the MDTP the prior spring to years in which they were not tested. To guard against effects specific to given grades or CST subject tests, we add dummies for each of these. Thus, the identification derives from a difference-in-differences approach in which we search for unusual differences in achievement for students when in grades and years they are tested using the MDTP, implicitly using a similar difference for students who are not subject to the MDTP testing so as to control for grade and year effects.

We also investigate whether there is any systematic relationship between observable characteristics and non-compliance, using the following linear probability model.

$$\Pr(MDTP_t = 1 \mid Mandate_{it} = 1) = \alpha_i + \beta_s + Y_{it}\tau + Grade_{i,t}\Omega + Year_t\Lambda + Test_{it}\Gamma + \varepsilon_{it}$$

Thus, we regress the outcome dummy variable of MDTP-taking for the subsample of students mandated to take one of the MDTP tests, on student fixed effects, a set of school dummies, the standardized CST score of the current academic year (which is administered before MDTP testing), dummies for each grade and year, and test dummies for each CST subject. The result suggests that schools are different in compliance rate.[4] However, to the extent that unobserved differences across schools are constant over the four-year pass of our panel, we have fully controlled for variations in compliance rates by including school fixed effects in our test-score models.

We also tested for dynamic selectivity bias in the form of Ashenfelter's Dip by testing whether performance in the CST that spring or the prior spring predicted whether the student took the MDTP that May or June.

[4] Joint hypothesis testing that coefficients of all the school dummies are zero is rejected at 1%.

The CST score in March of the current academic year does not predict the take-up in May or June, regardless of whether we include student fixed effects. Similarly, last year's CST score does not predict whether the student takes the MDTP in May or June. This finding suggests that schools are not endogenously choosing the students to whom they give the MDTP tests.

Other factors may also explain the incomplete compliance. For instance, total enrollment data might include those students who stayed in the district only temporarily, as little as 90 days. If these students were in the district at the time of the CST (in March) but not at the time of the MDTP (in May or June), they would seem like non-compliers in our dataset. The failure to take-up in this regard causes less concern once we account for student fixed effects, as the timing of the MDTP is likely to be exogenous to a student's decision on when to move.

An Instrumental Variable Approach: Impact of Treatment on the Treated and Intention to Treat Effect

Although we presented evidence above that schools did not discriminate by prior achievement in deciding who to test, there was less than 100 percent compliance with the district testing mandate, and this variation, in part, occurs across schools. Because there could be selectivity bias in terms of who was tested, we elect to reduce this potential endogeneity by instrumenting $MDTP_{t-1}$ with a dummy variable set to 1 if, in the student's current grade and year, the district had mandated that students should be administered the MDTP. The first-stage fit is very strong, and the IV readily passes weak instrument tests. For instance, the coefficient on our instrument, $Mandate_{t-1}$, was 0.69 with a t-stat of 447. The partial R-squared was 0.48.

Table A.5 replicates our baseline estimate from Table A.4, and then in the second column shows the IV estimate. The IV estimate is actually higher than the baseline estimate, suggesting that, if anything, the schools that complied most strongly, and the students within schools who were most likely to be given the test, had weaker test-score gains quite independent of MDTP testing. In the final column we provide intent to treat estimates, in which we include the instrument itself. We obtain results close to the baseline estimates, which again are highly statistically significant.

TABLE A.5
IV and intention to treat estimates

	Baseline	IV	ITE
MDTP t-1	0.219***	0.317***	
	(0.019)	(0.009)	
Mandate t-1			0.218***
			(0.02)
Observations	342571	342571	342567

NOTE: IV estimates are obtained by using Mandate t-1 as an instrument for MDTP t-1. First-stage t-statistics is 447.18 and corresponding p-value on exclusion of Mandate t-1 is 0.0000. Dependent variable is Gain in CST test score, standardized by Grade and Test. All regressions include student, grade, year, CST subject type, and school fixed effects. Standard errors are clustered by school and shown in parentheses An exception is the IV model for which it was not possible to cluster the IV model at the school level given that we include student fixed effects but students are not entirely clustered within schools.

Significance indicators: * p<0.10, ** p<0.05, *** p<0.01.

Robustness Tests and Extensions

As described above, in grades 8 through 11 the math CST taken depends on the math material the student has actually studied. We have controlled for the subject matter of the test taken this year, but it might be more appropriate to control for both the type of CST test taken in the current year and the type of test taken in the previous year. This did not change the results meaningfully. In the model in the first column of Table A.4, the coefficient on $MDTP_{t-1}$ remained significant at the 1 percent level, but the coefficient fell from 0.219 to 0.184. In a separate model, instead of separately adding grade dummies and dummies for the CST test taken, we interacted the two. The coefficient on lagged MDTP was virtually unchanged at 0.218, as was its level of significance.

As an alternative approach to the fixed effect model that models gains in test scores, a more general approach is to model the *level* of test scores as a function of the lagged test score and other explanatory variables. This specification is helpful because the gains model we have used thus far implicitly imposes a coefficient of 1 on a lagged test score in a model of the *level* of test scores. If, for instance, we are concerned that test scores regress to the mean, the coefficient on lagged test score might in truth be far smaller than 1 or even negative. Thus, we adopt the fixed-effect approach of Anderson and Hsiao that allows for a lagged dependent variable.[5] We start with the following specification.

$$Y_{icgst} = \alpha_i + \beta_s + Y_{icgs,t-1}\delta + MDTP_{i,t-1}\Omega + Family_{it}\Phi_0 + Class_{ist}\Phi_1 + School_{ist}\Phi_2 + Teacher_{ist}\Phi_3 + Year_t\Lambda + Test_{it}\Gamma + \varepsilon_{it}$$

This model cannot be estimated in the standard way because in short panels (short T), results will be inconsistent. After first-differencing, we get:

$$\Delta Y_{icgst} = \Delta\beta_s + \Delta Y_{icgs,t-1}\delta + \Delta MDTP_{i,t-1}\Omega + \Delta Family_{it}\Phi_0 + \Delta Class_{ist}\Phi_1 + \Delta School_{ist}\Phi_2 + \Delta Teacher_{ist}\Phi_3 + \Delta Year_t\Lambda + \Delta Test_{it}\Gamma + \Delta\varepsilon_{it} \qquad (2)$$

As the terms $\Delta Y_{icgs,t-1}$ and $\Delta\varepsilon_{it}$ are correlated, we instrument $\Delta Y_{icgs,t-1}$ using $Y_{icgs,t-2}$.

The results based on equation (2) are presented in Table A.6, which replicates our models in Table A.4. The coefficient on the indicator for whether the student took a district-mandated MDTP test the previous year is significant at the 1 percent level, and positive, but the coefficient is about half as big as in the main model, at 0.098. This is still a sizeable coefficient. Notably, the coefficient on lagged achievement is far less than 1, and in fact is slightly negative, implying slightly more than 100 percent regression to the mean. Columns (2) and (3) show that MDTP t-2 has a positive effect on student test score, which is in fact quite consistent with the models shown in Table A.4. Taking MDTP for two consecutive years has a positive effect on student test scores, as shown in the coefficient on the interaction term in Column (3). Column (3) of Table A.6 is the source of our alternative specification shown in Figure 4 of the report. Note that because the lagged test score has a coefficient of -0.03, the predicted change in a student's Z-score after being tested for two years in a row would be -0.03(0.116) + 0.048 + 0.149 = 0.193. This translates to a student who had been at the 50th percentile rising to almost the 58th percentile after two years, as 100*Φ(0.193)=57.7.

---

[5] See T. W. Anderson and Cheng Hsiao, "Estimation of Dynamic Models with Error Components," *Journal of the American Statistical Association* 76 (375): 598–606.

TABLE A.6
Estimated effect of MDTP on test scores: Anderson-Hsiao models of overall effect and effect by type of MDTP test

	(1)	(2)	(3)	(4)
MDTP t-1	0.098***	0.135***	0.116***	
	(0.014)	(0.013)	(0.013)	
MDTP t-2		0.075***	0.048***	
		(0.015)	(0.016)	
MDTP t-1 * MDTP t-2			0.149***	
			(0.023)	
Pre-algebra t-1				0.048**
				(0.023)
Algebra t-1				0.058*
				(0.030)
Geometry t-1				0.156***
				(0.024)
Lagged test gains	-0.019	-0.030	-0.032	-0.020
	(0.022)	(0.022)	(0.022)	(0.022)
Observations	214223	196634	196634	214223

NOTE: Estimates are based on equation (2) in the text. Dependent variable is Level in CST Test score, standardized by Grade and Test. All regressions include grade, year, and CST subject type fixed effects. First-differenced school fixed effects are partialled out. Standard errors are clustered by school and shown in parentheses.

Significance indicators: * p<0.10, ** p<0.05, *** p<0.01.

Column (4) in Table A.6 repeats our model that replaces the MDTP variable with indicators for which MDTP test was taken. All three test coefficients are positive and significant, indicating that all three are contributing to the positive MDTP coefficient in column (1). The relative effects of the three types of MDTP tests are different though, with the Geometry readiness test having a coefficient roughly triple that of the Pre-Algebra and Algebra tests. In contrast, in the gains model the coefficients were more closely aligned, with the coefficients on these three variables in a relatively narrow range of 0.15 to 0.24, with the Geometry test having a slightly higher coefficient than the Algebra test.

Table A.7 interacts lagged mandated MDTP testing with indicators for whether the student's lagged math achievement was medium or high. We use two different definitions of medium and high achievement. Model (1) uses the prior year's math score on the CST, and model (2) instead uses the earliest math CST available for each student. The former is a more recent measure but the latter has the advantage of observing math achievement in grades that are typically before MDTP testing begins.

TABLE A.7
Variation in relation between mandated MDTP testing and initial math achievement

	(1)	(2)
MDTP t-1	0.091***	0.101***
	(0.015)	(0.022)
MDTP t-1*Medium Achievement	0.095***	
	(0.023)	
MDTP t-1*High Achievement	0.088***	
	(0.028)	
MDTP t-1* Medium Achievement		0.010
		(0.022)
MDTP t-1*High Achievement		0.129***
		(0.045)
Constant	1.727***	-1.357***
	(0.141)	(0.113)
Observations	342571	342571

NOTES: Indicators for the middle or the highest third of initial distribution for Model (1) are based on last year's math test score, while for model (2) they are based on the first available math test score. Standard errors in parentheses.

Significance indicators: * p<0.10, ** p<0.05, *** p<0.01.

In both cases we find evidence that students in the top two-thirds of the past math achievement gain more from MDTP testing, although all three groups benefit. There is some disagreement between the two models as to whether it is students in the middle or at the top of the distribution who gain the most.

Voluntary Versus Mandated MDTP Testing

We also obtained data on voluntary use of MDTP tests by individual math teachers. To distinguish between teachers' voluntary use of MDTP tests and district-mandated MDTP testing, we will refer to the former, voluntary testing, as VMDTP and reserve the term MDTP for district-mandated testing. We obtained information on which students were tested from 2003 forward.

There are several issues with the VMDTP testing that make it relatively more difficult to infer its effects on subsequent student achievement. First, the testing is voluntary, and there is no obvious way to take into account why some teachers used the MDTP and others did not. Second, we were able to match the VMDTP data to specific students in 80 percent of the cases, so there will be some undercounting. Third, we found a bimodal distribution of VMDTP testing over the school year. As shown in Figure A.1, many teachers gave the MDTP test to their students in early fall, ostensibly to learn about the specific areas of mathematical understanding in which their students needed the most help. But many teachers used the MDTP in spring, after the statewide tests that we use as our outcome measure were administered. Testing so late could only have a causal effect on the student's CST scores during late winter of the next grade. For this reason, we distinguish two types of VMDTP testing: testing that occurred between September and March of the current school year, and therefore before the statewide testing occurs in late March, and VMDTP testing between April and August of the prior school year.

The first column in Table A.8 repeats our specification from column (1) of Table A.4, but adds an indicator for whether either type of prior voluntary testing occurred (during the current school year or between April and August of the prior school year). The coefficient of MDTP remains significant and positive but the voluntary testing coefficient (VMDTP) is significant and has a small negative coefficient. In the ensuing two columns, we enter separately voluntary testing between September and March of the current year and then VMDTP during April to August of the prior school year (or ensuing summer). In column (2) we find that, if anything, there is a small negative correlation between voluntary testing during the current year and gains in the statewide test; and in column (3) we see no link between testing between April and August of the prior year and gains in the statewide CST test during the current academic year. As a test of collinearity, columns (4) through (6) repeat models (1) through (3) but drop MDTPt-1. These results suggest that the negative coefficient on VMDTP between September and March of the current year does not depend on whether we control for mandatory testing. However, the same cannot be said for voluntary testing between April and August of the prior school year, which now becomes positive and significant. This pattern almost surely reflects a high degree of collinearity between district-mandated and voluntary testing in spring of the prior academic year. Teachers show a surge in the voluntary use of MDTP in the spring in years and grades in which the district imposes the same test. It is likely that the positive coefficient on lagged VMDTP is simply capturing the positive effect of mandated MDTP.

TABLE A.8
Estimated effect of MDTP and VMDTP (mandatory and voluntary testing, respectively) on test scores

	(1)	(2)	(3)	(4)	(5)	(6)
MDTP t-1	0.232***	0.218***	0.232***			
	(0.019)	(0.019)	(0.021)			
VMDTP Sep–Mar this year or Apr–Aug last year	-0.032**			0.055***		
	(0.014)			(0.015)		
VMDTP Sep–Mar this year		-0.034**			-0.039**	
		(0.016)			(0.017)	
VMDTP Apr–Aug last year			-0.027			0.134***
			(0.021)			(0.020)
Observations	342567	342567	342567	342567	342567	342567

NOTE: Estimates are based on equation (1) in the text. Dependent variable is Gain in CST test score, standardized by Grade and Test. All regressions include student, grade, year, CST subject type, and school fixed effects. Standard errors are clustered by school and shown in parentheses.

Significance indicators: * p<0.10, ** p<0.05, *** p<0.01.

Potential Explanations for the Positive Effect of Mandated MDTP Testing

To delve further into why mandated MDTP might have a positive, if temporary, effect on gains in math achievement, we examined two elements of tracking. First, the district used overall student performance during a year, in addition to MDTP performance, to recommend to parents whether their children should attend summer school.[6] Did MDTP scores in fact correlate negatively with summer school assignment? More importantly, can summer school attendance explain any of the estimated MDTP effect? Second, the district used letter grades, teacher recommendations, and MDTP scores to make decisions on the level of math class into which a student would be placed in the subsequent fall of middle or high school. Can we detect any evidence that this in fact took place, and that it influenced student achievement?

[6] Using a randomized field trial, Borman and Dowling find that attending summer school promotes longitudinal achievement growth. See Geoffrey D. Borman and N. Maritza Dowling, "Longitudinal Achievement Effects of Multiyear Summer School: Evidence From the Teach Baltimore Randomized Field Trial," *Educational Evaluation and Policy Analysis* 28 (1): 25–48.

Table A.9 shows linear probability models of the probability that students attended summer school before grades 7 through 11 on the corresponding sample of secondary school students. In each of the columns, we condition on whether the student took the mandated MDTP in the prior year. It appears that the district was increasing its use of summer school by about two percentage points in the years of the mandated MDTP testing, as this variable is positive and significant with a coefficient of 0.02. Column (2) shows that the math CST score in the prior year is negatively associated with attendance at summer school. Schools made recommendations on who should attend summer school before receiving these test scores, but they serve as a noisy measure of achievement, which should be positively correlated with teachers' own assessments of student achievement. As expected, weaker students were more likely to attend summer school. Column (3) tests the idea that MDTP testing would increase the probability that weaker students would be sent to summer school: The interaction between MDTP testing and that spring's CST score is negative and weakly significant. The predicted effect of a one standard deviation drop in math CST score on summer school attendance doubles, from a 1.1 percent increase in probability to a 2.1 percent increase in probability of attendance, in cases where the student was given the mandated MDTP test. We conclude that the use of mandated end-of-year MDTP testing increases the probability that low-achieving students are assigned to summer school.

TABLE A.9
Does MDTP affect enrollment in summer school?

	(1)	(2)	(3)
MDTP t-1	0.021**	0.020**	0.021**
	(0.009)	(0.009)	(0.009)
CST score t-1		-0.012***	-0.011***
		(0.002)	(0.002)
MDTP t-1 * CST score t-1			-0.010*
			(0.005)
Observations	183847	183847	183847

NOTE: Dependent variable is an indicator of attending summer school. All regressions include student, grade, year, CST subject type, and school fixed effects. Standard errors are clustered by school and shown in parentheses. Samples are restricted to grades 7 to 11, grades for which summer school data are available.

Significance indicators: * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

These findings naturally lead to the next questions: Do students who attend summer school gain more in math achievement by the following spring, and can any summer school effects at least partly explain the positive estimated effect of MDTP on subsequent test-score gains? Table A.10 models gains in the math CST test score as a function of summer school attendance and having taken the MDTP the prior spring. Column (1) suggests that summer school is associated with a 0.09 standard deviation increase in math achievement by the following spring. After replicating our basic model on this subsample in Table A.10, we include both lagged MDTP and summer school in column (3). Compared to column (2), the coefficient on lagged MDTP falls by 0.002 or about 1 percent. Thus, the effect of taking MDTP on summer school placement can explain perhaps only 1 percent of the overall effect of mandated MDTP. In column (4) we interact lagged MDTP and summer school, but the interaction is not significant.

TABLE A.10
Effect of mandate MDTP testing and summer school on achievement gains

	(1)	(2)	(3)	(4)
Summer	0.089***		0.080***	0.087***
	(0.016)		(0.016)	(0.018)
MDTP t-1		0.228***	0.226***	0.228***
		(0.023)	(0.023)	(0.023)
MDTP t-1*Summer				-0.027
				(0.030)
Observations	177612	177612	177612	177612

NOTE: Dependent variable is Gain in CST test score, standardized by Grade and Test. All regressions include student, grade, year, CST subject type, and school fixed effects. Standard errors are clustered by school and shown in parentheses. Samples are restricted to grades 7 to 11, grades for which summer school data are available. (Sample size is smaller than in Table 9 since dependent variable here requires two consecutive years of test score.)

Significance indicators: * p<0.10, ** p<0.05, *** p<0.01

Next, we examine the question of whether the student's score on the MDTP test affects placement in the following academic year. If it is true that mandated MDTP testing leads to a more accurate placement, then we should expect to see the variance of test scores, within the set of students taking a certain class, fall if they were given the MDTP in the prior spring. The first empirical test we conduct is to regress the standard deviation of prior spring's standardized test score for each class on an indicator variable for whether a student took MDTP the prior year, along with other control variables that were listed in the basic specification. Taking MDTP the prior year yields a modest decrease in within-class variation by moving toward a more homogenized ability-group setting. The effect is significant at the 1% level. Controlling for a measure of the dispersion would yield MDTP coefficient fall towards zero, if this sorting of students by ability drives the large effect of MDTP.

The second empirical exercise is therefore to see whether controlling for the initial test score dispersion within class makes the MDTP effect shrink. We add to our student fixed effects model of test scores the measure of within-class dispersion. As shown in Table A.11, the coefficient of lagged MDTP drops from 0.18 to 0.16 when the measure of within-course/class dispersion is included. Thus, it appears that about 11 percent of the MDTP effect is coming from increased degree of sorting.

We conducted robustness checks by estimating the results in Tables A.10 and A.11 using the Anderson and Hsiao model (see Table A.b in the Supplementary Tables at the end of this appendix). In that model, summer school explains about 1 percent of the MDTP effect, exactly as we found above. Including the standard deviation of prior test scores for each class reduces the lagged MDTP coefficient by 5 percent.

TABLE A.11
Can effect of MDTP testing on math achievement be explained by resulting changes in standard deviation of initial achievement by classroom?

	(1)	(2)
MDTP t-1	0.181***	0.164***
	(0.022)	(0.021)
SD for each class		-0.389***
		(0.042)
Observations	151264	149305

NOTE: Class data for 6th to 10th graders are used. For 1st–5th graders, most students likely take one generic math course for their grade. Also, many students in 11th and 12th grade did not take the CST test, so we exclude them rather than including selected students among those who took the CST test in later grades.

Significance indicators: * p<0.10, ** p<0.05, *** p<0.01.

Overall, we conclude that about 1 percent of the effect of MDTP on subsequent math achievement can be explained by the role it plays in placing struggling students in summer school, and better ability grouping of students by classroom can account for 5 to 11 percent of the effect of MDTP, using our measure of the variance of achievement in the actual math classroom.

Supplementary Appendix Tables

TABLE A.a
Summary statistics

Variable	Mean	Standard deviation	Minimum	Maximum	Number of observations
$MDTP_{t-1}$	0.135	0.342	0	1	342567
$MDTP_{t-2}$	0.109	0.312	0	1	311412
$MDTP_{t-1}$ * $MDTP_{t-2}$ *Type of MDTP Test*	0.019	0.135	0	1	311412
Pre-algebra$_{t-1}$	0.063	0.243	0	1	342567
Algebra$_{t-1}$	0.021	0.145	0	1	342567
Geometry$_{t-1}$	0.026	0.160	0	1	342567
Voluntary MDTP (VMDTP)	0.196	0.397	0	1	342571
VMDTP Sep-Mar this year	0.079	0.270	0	1	342571
VMDTP April to August last year	0.117	0.321	0	1	342571
CST Math Test Score Standardized by Grade and Test	-0.003	0.999	-4.751	8.791	475633
Difference in CST Math Score, Standardized by Grade and Test	0.049	0.782	-6.037	7.850	342571
Type of Math CST Test Taken					
Test==AL1	0.14	0.35	0	1	342567
Test==AL2	0.07	0.25	0	1	342567
Test==Math8/9	0.02	0.16	0	1	342567
Test==Gmtry	0.11	0.31	0	1	342567
Test==HsMath	0.03	0.18	0	1	342567
Test==Integrated Math	0.02	0.13	0	1	342567
Other CST test (grade-specific)	0.60	0.49	0	1	342567
Average Math Class Size *School Demographics*	14.53	14.08	0	47	342567
% of school Asian	16.78	14.64	0	100	342567
% of school white	25.81	20.85	0	100	342567
% of school Hispanic	41.42	23.11	0	100	342567
% of school black	13.63	10.39	0	98.41	342567
% of school English Learner	19.48	20.16	0	100	342567
% of school on free lunch *Teacher Characteristics in Math Class (or in Home Room in Elementary Classrooms or Other Self-Contained Classrooms(Averaged across Semesters)*	55.67	28.43	0	100	342567
Average years teaching at SDUSD	10.73	7.12	0	42	342567
Average years teaching	12.39	7.77	-1	46	342567
Average percent bachelors in math	18.79	38.64	0	100	342567

Variable	Mean	Standard deviation	Minimum	Maximum	Number of observations
Average years service in math classes	7.37	9.96	-1	44	342567
Average SDUSD years service in math classes	6.35	8.93	0	39	342567
Average full credential among math teachers	54.37	49.71	0	100	342567
Average teacher intern	0.86	8.93	0	100	342567
Average full authorization in math	0.29	0.45	0	1	342567
Average supplemental authorization in math	0.12	0.32	0	1	342567
Average female teacher	29.22	44.86	0	100	342567
Average CL:AD certificate	24.65	42.57	0	100	342567
Average any Master's degree	28.44	44.68	0	100	342567
Average of white teachers	41.45	48.88	0	100	342567
Average of black teachers	3.20	17.23	0	100	342567
Average of Asian teachers	6.67	24.54	0	100	342567
Average of Hispanic teachers	3.19	17.33	0	100	342567
(Potentially) Time-Varying Student Characteristics					
English learner	0.23	0.42	0	1	342567
Grade Level					
3	0.12	0.33	0	1	342567
4	0.12	0.33	0	1	342567
5	0.12	0.33	0	1	342567
6	0.11	0.32	0	1	342567
7	0.11	0.32	0	1	342567
8	0.11	0.32	0	1	342567
9	0.11	0.31	0	1	342567
10	0.10	0.30	0	1	342567
11	0.08	0.28	0	1	342567
School Year					
2002-03	0.20	0.40	0	1	342567
2003-04	0.21	0.41	0	1	342567
2004-05	0.21	0.41	0	1	342567
2005-06	0.20	0.40	0	1	342567
2006-07	0.19	0.39	0	1	342567
Parental Education (More Educated Parent)					
Less Than High School	0.16	0.37	0	1	342567
High School	0.18	0.39	0	1	342567
Some College	0.16	0.36	0	1	342567
College Graduate	0.18	0.38	0	1	342567
Postgraduate College	0.10	0.30	0	1	342567
Missing	0.22	0.41	0	1	342567

TABLE A.b

Estimated effect of MDTP on test scores that control for summer school attendance and within-class dispersion: Anderson-Hsiao models

	(1)	(2)	(3)	(4)
MDTP t-1	0.081***	0.080***	0.077***	0.076***
	(0.013)	(0.013)	(0.013)	(0.013)
Summer		0.015		0.017*
		(0.010)		(0.010)
SD for each class			-0.111***	-0.111***
			(0.034)	(0.034)
Observations	73142	73142	72228	72228

NOTE: Dependent variable is Level in CST Test score, standardized by Grade and Test. All regressions include grade, year and CST subject-type fixed effects. First-differenced school fixed effects are partialled out. Standard errors are clustered by school and shown in parentheses. Samples are restricted to grades 7 to 11, the grades for which summer school data are available (Sample size is smaller than in Table A.9 since the dependent variable here requires two consecutive years of test score).

Significance indicators: * p<0.10, ** p<0.05, *** p<0.01.

PUBLIC POLICY INSTITUTE OF CALIFORNIA

The Public Policy Institute of California is dedicated to informing and improving public policy in California through independent, objective, nonpartisan research on major economic, social, and political issues. The institute's goal is to raise public awareness and to give elected representatives and other decisionmakers a more informed basis for developing policies and programs.

The institute's research focuses on the underlying forces shaping California's future, cutting across a wide range of public policy concerns, including economic development, education, environment and resources, governance, population, public finance, and social and health policy.

PPIC is a private operating foundation. It does not take or support positions on any ballot measures or on any local, state, or federal legislation, nor does it endorse, support, or oppose any political parties or candidates for public office. PPIC was established in 1994 with an endowment from William R. Hewlett.

PUBLIC POLICY INSTITUTE OF CALIFORNIA
500 Washington Street, Suite 600
San Francisco, California 94111
phone: 415.291.4400
fax: 415.291.4401
www.ppic.org

PPIC SACRAMENTO CENTER
Senator Office Building
1121 L Street, Suite 801
Sacramento, California 95814
phone: 916.440.1120
fax: 916.440.1121