MONASH BUSINESS SCHOOL

DEPARTMENT OF ECONOMICS

ISSN 1441-5429

DISCUSSION PAPER 31/15

Can Testing Improve Student Learning? An Evaluation of the Mathematics Diagnostic Testing Project

Julian R. Betts[a], Youjin Hahn[b] and Andrew C. Zau[a*]

Abstract:
Unlike state-mandated achievement tests, tests from the Mathematics Diagnostic Testing Project (MDTP) offer teachers timely and detailed feedback on their students' achievement. We identify the effects of providing feedback on student outcomes by using data from the San Diego Unified School District, a large urban school district where mandatory diagnostic tests in mathematics were implemented to some grades between 6 and 9 during 1999 and 2006. These diagnostic tests offer teachers timely and detailed feedback on their students' achievement. We find that providing diagnostic feedback improves math test scores by 0.1 to 0.2 standard deviations. All students gain from this mandated testing, but those with higher initial achievement gain the most. The gains arise in part from students being more accurately tracked into appropriate math classes. The gains decay unless students are tested repeatedly.

Keywords: Diagnostic testing; student learning; education
JEL Classification Numbers: I21

[a] Department of Economics, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA
[b] Department of Economics, Monash University, Building 11, E878, Clayton, VIC 3168, Australia
* E-mail addresses: jbetts@ucsd.edu (Julian R. Betts); youjin.hahn@monash.edu (Youjin Hahn); azau@ucsd.edu (Andrew C. Zau)

1. Introduction

Partly in response to the federal No Child Left Behind law of 2001, U.S. states have adopted statewide mandatory testing of public school students meant to measure academic achievement and progress towards universal student proficiency. These tests, often referred to as summative tests, are meant to measure overall student achievement, rather than diagnosing specific areas in which students need to improve. The existing state tests lack this level of detail, and often they take so long to grade that the school year has ended before schools receive results.

Frustrated in part with the limited ability of the existing state tests to diagnose specific learning difficulties in a timely fashion, many states have recently signed on to the "Common Core" – a new curriculum that aims to impart higher order skills to students. Accompanying the new standards are two multi-state consortia that are unveiling new tests in 2015 that aim to provide schools with much more timely and detailed reports on student progress. Both the Smarter Balanced Assessment Consortium and the Partnership for Assessment of Readiness for College and Careers (PARCC) are designing computer-adaptive tests for use in multiple states, which will provide speedy feedback on student performance. In addition to providing diagnostic information on what materials students have and have not mastered, the two consortia are developing optional interim assessments that districts can use to gauge student progress during the year.[1] These reforms reflect potentially monumental changes to the way the nation tests, with a total of 23 states belonging to PARCC and 25 states participating in the Smarter Balanced Assessment Consortium.

A major question underlying these new developments is whether diagnostic testing that provides detailed diagnosis of students' strengths and weaknesses, with rapid turnaround time, can make a tangible contribution to student learning. Very little is known about the impact of such tests.[2] We therefore study a

---

[1] Information on the two consortia can be found at www.parconline.org and www.smarterbalanced.org.

[2] A literature on formative assessment, defined as classroom assessments conducted by teachers on materials recently taught, suggests that such assessments can boost student performance. See Black and Wiliam (1998) and

form of diagnostic testing already widely used in California, and test whether its adoption at the district level leads to better academic achievement.

California, like other states, has implemented an educational accountability system designed to measure the performance of students in public schools, and to provide a series of interventions for schools that fail to improve sufficiently. At present, several mandated tests feed into the calculation of the summary Academic Performance Index (API), the single number the state publishes annually as a proxy of each school's academic performance. During the period we study these tests include the California Standards Test (CST), offered in a variety of subjects between grades 2 and 11, and the California High School Exit Exam, which high school students must pass in order to receive a diploma.

This accountability program has informed the public about average achievement levels in California and the large variations in achievement across schools and across demographic groups. However, the testing system does not provide timely feedback to teachers about their students' performance, or about the specific subareas of knowledge within a subject in which individual students need to improve. Notably, results from CST tests given in March do not become available until late summer.

This paper examines the effectiveness of providing diagnostic feedback to teachers and students. Although the public is now widely aware of API scores and the main state tests, these state tests are designed mainly for purposes of external accountability rather than to improve learning. There is a unique diagnostic test, with a quite different purpose, that has been freely offered to math teachers and their students throughout the state since 1982 in the San Diego Unified School District. The Math Diagnostic Testing Project (MDTP) is a joint program of the California State University and the University of California, and it offers free diagnostic testing to math teachers throughout California. The MDTP is designed to allow teachers to obtain diagnostic testing of their students, with speedy and detailed

Collins et al. (2011) for a discussion of formative assessment. Note however, that in the present study we focus on more formal tests that originate outside the teacher's own classroom.

feedback. Teachers typically receive printed results within days of test administration, along with overall and student-by-student information on student performance on individual topics within the subject of the test. For example, the MDTP's Algebra Readiness test, designed to be given to students before their first algebra course, gives information to teachers and their students on students' understanding of integers, fractions, decimals, percentages and six other clearly defined topics.

We identify the effects of providing feedback on student outcomes by using data from San Diego Unified School District, where mandatory diagnostic tests in mathematics was implemented to some grades between 6 and 9 during 1999 and 2006. Using variation arising from different timing and targeted grades of MDTP, we find that students who in the previous year took an MDTP test under the mandatory district-wide program gained 0.1 to 0.2 standard deviations in math performance compared to their average performance.

Quite a few studies have examined broader accountability systems such as No Child Left Behind (See Dee and Jacob 2009 for a helpful review of earlier work on accountability systems). However, there is sparse rigorous evidence on the effectiveness of low-stakes diagnostic testing on educational outcomes. One notable exception in recent literature is Muralidharan and Sundararaman (2010), which evaluates the effect of providing diagnostic feedback to teachers on student learning outcomes in India. Their paper finds that students in feedback schools where teachers received detailed test-score results of students do not perform differently from the students in control schools, while the teachers in feedback schools exerted more effort when observed in the classroom. Unlike the experimental setting in India, the mandatory MDTP test used in our study was implemented in a time and place where schools were evaluated for accountability systems, which would make teachers have higher incentive to use the diagnostic feedback to improve student learning. Our findings suggest that providing diagnostic feedback is more effective in a situation where schools and teachers are accountable for student outcomes, whereas it may not be as effective in encouraging teachers' intrinsic motivation which eventually leads to student learning, as in Maralidharan and Sundarararaman (2010).

The remainder of the paper is organized as follows: section 2 provides background information on Mathematics diagnostic testing project; section 3 describes data; section 4 discusses empirical strategy; section 5 provides baseline results and offers further checks for robustness and extensions; section 6 explores potential channels for the effect of the diagnostic testing; and section 7 concludes.

2. Background

The MDTP consists of a set of "readiness" tests designed to give students and teachers detailed feedback on the student's readiness to move on to a given course in the next academic year. During the time period of this study, the most commonly used MDTP tests in the San Diego Unified School District were the Pre-algebra Readiness Test, the Algebra Readiness Test and the Geometry Readiness Test. For example, the Algebra Readiness test assesses a student's preparation in topics that are required knowledge for a student to fare well in a subsequent algebra course, while the Geometry Readiness Test evaluates a student's understanding of first-year algebra topics that students will need to have mastered in order to do well in a subsequent geometry course. Each test lasts 45 minutes and contains, depending on the subject matter, 40-50 multiple choice questions.

The teacher will receive for each student, as well as for her class, indicators of the percentage of questions answered correctly in specific areas.[3] Areas in which students answer correctly fewer than 70 percent of questions will be flagged for teacher follow-up. The teacher receives this information individually for each student, but also receives statistics on the percentage of students scoring 70 percent or higher on each math topic, which might help him or her focus on areas for greater emphasis and clarity in the classroom in general.

[3]For example, in the Algebra Readiness test, the areas include: Data Analysis, Probability and Statistics; Decimals, their Operations and Applications; Percent; Simple Equations and Operations with Literal Symbols; Exponents and Square Roots; Scientific Notation; Fractions and Their Applications; Measurement of Geometric Objects; Graphical Representation; Integers, Their Operations and Applications.

The intent of these tests is thus quite different from statewide testing systems as mandated under the federal No Child Left Behind law. The latter are designed to measure overall school progress and results can trigger various interventions. The MDTP tests are designed to help students and teachers to work together on specific math topics that the student has yet to master. Nonetheless, the mandated MDTP test was not entirely a low-stakes diagnostic test. The test was given in May and June, and was one of many factors that teachers took into account when placing students into math classes for the subsequent year.

One key distinguishing feature of the MDTP is that unlike the state tests that comprise California's official accountability system, the MDTP tests are graded locally in one of ten regional site offices. The proximity of the grading centers confers two advantages, by speeding up turnaround time dramatically, and by enabling regional coordinators to work with local schools to interpret and use test results. The accessibility of the local MDTP offices is largely responsible for the ability of the MDTP to provide timely and detailed feedback to teachers, which they can immediately put into use by tailoring their instruction to the specific needs of individual students.[4] But perhaps the most important distinction from the state test in use at the time of our study is that the MDTP gives teachers and students very specific feedback on areas within a given curriculum where students need to improve.

In 2007-2008, MDTP scored approximately 40,000 tests for 95 SDUSD schools (including private schools in the SDUSD region). In general, most of the testing occurs in middle schools, but a significant number of high school students are also tested. For instance, of the 40,000 tests done in 2007-2008, 10,500 tests were scored at 22 schools for which the highest grade was grade 12.[5]

Beginning in 1999-2000 the district as a whole mandated the use of at least one MDTP test at the end of the school year. These mandates were 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, as students move from grade to grade, they will be subject to testing in only a few grades. Thus, the effects

[4] More details on the MDTP tests and the history of the program are available at http://mdtp.ucsd.edu.
[5] This information was provided by Alfred Manaster, who is the emeritus State Director of the MDTP.

of MDTP are identified by comparing individual student trends in achievement during years and grades with and without testing. Table 1 shows the way in which SDUSD mandated testing.

[Table 1]

Table 2 shows the actual proportions of students in each grade and school year who took an MDTP diagnostic test. It suggests fairly high, but not complete, compliance with the district mandate. The estimated effect of MDTP could be biased if there is non-random compliance. As discussed in section 5.2, we find the effects of actual MDTP taking and "intent-to-take" are similar, suggesting that non-random compliance is unlikely to bias our estimates.

[Table 2]

In addition to the mandatory testing, individual math teachers have voluntarily adopted MDTP testing early in the school year. Beginning in 2000 the district as a whole mandated the use of at least one MDTP test at the end of the school year. Some teachers have continued to provide testing at the start of the school year since that time. 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 taken with actual student data based on student identifier, birthday, name, and teachers. The nature of scanned, voluntary data leads to some students who cannot be identified because of insufficient data on identifiers. Approximately 80% of the voluntary tests could be matched to student data.

3. Data

The San Diego Unified School District (SDUSD) consists of approximately 135,000 students in pre-school through grade 12. It is the second largest district in California and as of 2015 is the 20th largest urban district in the United States. Our longitudinal data consists 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, and a rich set of covariates related to the student's class size and teacher qualifications. Teacher qualifications at the SDUSD fall under three main areas: education, credentialing, and years of experience. For most teachers, our data-set includes the highest-degree attained (bachelor's, masters, or doctorate), and subject major. We control for the level of overall credential and the math subject authorization of the student's home-room teacher (in elementary grades) and the student's math teacher in middle and high school.[6]

California has administered various standardized tests at different times. It mandated the Stanford 9 test in spring 1998 through spring 2002, and the California Standards Test (CST) in spring 2002 and later years. Our outcome of interest is a gain in standardized math CST scores. We use CST data from spring 2002 through 2007. 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.

Starting in grade 8, students take different version 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. We present in Table 3 the seven most common sequences of test-taking for cohorts who are 8th grade in 2001-2002. Up until 8th grade, students take similar courses. However, students start diverging in terms of their courses-taking behavior from 9th grade. 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 3]

A second approach would have been to standardize not just by grade and CST subject, but also by 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

[6] Credentials can be divided into two types: overall credential, and subject authorization. The credential simply 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 as 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. Further details of the credentialing system may be found in Chapter 3 of Betts, Zau and Rice (2003).

changes in Z-scores when standardized by grade and subject, but not by year, because if the mandated MDTP testing had positive effects on achievement in a given grade, these trends will be completely removed and therefore undetectable in models of such scores.[7]

4. Empirical Strategy

We model gains in CST test scores as a function of student and family characteristics, teacher qualifications, school effects and an error term. The district-mandated MDTP testing took place in May or June of the academic year, while the CST is offered in March or April. Therefore, 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*. The basic model is the following. Below MDTP refers to district-mandated MDTP testing.

$$(1)\ Y_{icgst} - Y_{icgs,t-1} = \alpha_i + \sigma_s + \beta\ MDTP_{i,t-1} + \gamma\ Family_{it} + \delta\ ClsSize_{icst} + \rho\ School_{ist} + \mu\ Teacher_{icst} + \pi\ Grade_{it} + \tau\ Year_t + \eta\ Test_{it} + \varepsilon_{icgst}$$

Y_{icgst} is a standardized test score in the California Standards Test (CST) for student *i* in (math) classroom *c*, grade *g*, school *s*, in year *t*. α_i is a student fixed effect and σ_s is a school fixed effect. As using the student fixed effect 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, β, is of primary interest in this study, and it indicates the level by which district-mandated MDTP testing shifts students' test scores from their average growth trajectory.

[7] In spite of these concerns, we did estimate models in which we standardize by grade, subject and school year. Our results on mandated MDTP testing are similar in terms of statistical significance to what is reported below, but more muted in terms of coefficients, when we standardize test scores in this way. 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.

The other control variables are as follows. A vector *Family* contains dummies for parental education and *ClsSize* has an annual average math class size for each student. *School* is a vector of time-varying school characteristics, consisting of the percentage of students on free lunch; percentage who were Asian, White, Hispanic, and African American; and percentage who were English Learner students. These are constructed by averaging variables for each school in a given year across semesters.

We also include various teacher characteristics. The *Teacher* vector 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 level. *Year* dummies are included to account for any changes in the educational system by which all students in the district are affected. 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; Algebra1, Algebra 2, Math 8/9, Geometry, High School Math, and Integrated math.

5. Results

5.1. The Effects of MDTP Taking on Test Score

Table 4 shows the main results. We show the results for the several variants of the indicator for MDTP taking, such as whether the student took a district-mandated MDTP test the prior year, two years ago, or two years in a row. Table A1 in the appendix shows the means and other summary statistics for all of the regressors.

[Table 4]

In the first model, we condition on $MDTP_{t-1}$, thus testing for an association between the individual student having taken a mandated MDTP test in May or June of the prior year and 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% 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. One possibility 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. 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.

In the third column of Table 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 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 2002-2003.[8] 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. There is essentially no depreciation in the initial estimated effect of MDTP testing if the student is tested for a second year.

As another extension, we interacted MDTP testing with indicators for prior math achievement. Table 5 shows the results of interacting lagged mandated MDTP testing with indicators for whether the student's lagged math achievement was medium or high (defined as being in the middle or top third of prior year math achievement respectively). 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

[8] 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 2 shows that the two cohorts listed above were mandated for testing two years in a row.

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 5]

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.

5.2. Identification issues

The main results are robust to the following concerns. First, one might be concerned that teacher characteristics and class size are endogenous. But the estimates are virtually unaffected after dropping these variables.[9]

Second, 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 main model in the first column of Table 4, the coefficient on $MDTP_{t-1}$ remained significant at the 1% level, but the coefficient fell from 0.219 to 0.188. In a separate model, instead of separately adding grade dummies and dummies for the CST test taken, we interacted the two. The coefficient estimate of lagged MDTP was 0.198 and was statistically significant at the 1% level.

Third, the average compliance rate for the cohorts who are affected by the mandate is about 0.7, reasonably high but not complete (Table 2). Given MDTP in the district became mandatory for certain cohorts, there would be less likely to be a selection issue than if students were volunteered for testing by their schools. However, we are concerned that students who took MDTP might be different in certain unobserved ways, thus biasing the effects. As we control for student fixed-effects, any selection problem

[9] Results are available on request.

would arise from within-student variation over time, rather than the difference between the students who took and who did not take MDTP. We investigate this issue as follows.

We first check whether there is any systematic difference in observable time-varying characteristics of students when they did and did not take the diagnostic test, using the following linear probability model:

$$(2)\ Pr\ (MDTP_{it} = 1 \mid Mandate_{it} = 1) = \alpha_i + \sigma_s + \delta Y_{i,t} + \gamma\ Family_{it} + \pi\ ClsSize_{icst} + \theta School_{ist} + \eta\ Teacher_{icst} + \mu\ Grade_{it} + \tau Year_t + \rho Test_{it} + \varepsilon_{icgst}.$$

Thus, we regress the binary indicator of MDTP-taking on student fixed effects, a set of school dummies, the standardized CST score of the current academic year (which is administered before MDTP testing), and other characteristics included in equation (1), while also conditioning on the year and grade fixed effects. School fixed effects are jointly significant, indicating that schools are different in compliance rate. However, to the extent that unobserved differences across schools are constant over the four-year period covered by our panel, we have fully controlled for variations in compliance rates by including school fixed effects in our test-score models. The coefficient on the CST score, δ, indicates whether there is dynamic selectivity bias akin to Ashenfelter's Dip, whereby a drop in test performance might later induce the school to give the student the MDTP test (Ashenfelter, 1978). The CST score in March of the current academic year *does not* predict the MDTP take-up in May or June in the same year (p-value = 0.410). Similarly, last year's CST score does not predict whether the student takes the MDTP in May or June (p-value = 0.720). This finding suggests that schools are not endogenously choosing the students to whom they give MDTP tests based on students' academic performance.[10]

As a further check, we instrument $MDTP_{t-1}$ with a dummy variable set to one if in the student's current grade and year the district had mandated that students should be administered the MDTP. If there

[10] 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 stayed 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 student's decision on when to move is unlikely correlated with the timing of the MDTP.

is selectivity bias in terms of who was tested based on unobserved characteristics other than investigated above, we can reduce this potential endogeneity by instrumenting $MDTP_{t-1}$ with a dummy variable set to one if in the student's current grade and year the district had mandated that students should be administered the MDTP.[11] Table 6 reports the results. We also obtain intent to treat estimates, in which we include the instrument itself. Both the IV estimates (column 2) and intent to treat estimates (column 3) are close to the baseline estimates and highly statistically significant. The IV estimate is slightly 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 last column of Table 6, as a placebo test, we run the same model as equation (1) but using gain in standardized reading score as a dependent variable (column 4). Since MDTP is intended to improve math performance, we should see a weaker effect of the MDTP on reading score. Confirming our expectation, the coefficient estimate on $MDTP_{t-1}$ drops significantly and it is no longer statistically significant.

[Table 6]

Lastly, as an alternative approach to the fixed effect model that models gains in test scores, we use a more general approach which is to model the *level* of test scores as a function of the lagged test score and other explanatory variables. Thus, we adopt the fixed-effect approach of Anderson and Hsiao (1981) that allows for a lagged dependent variable. To do so, we take the first difference of the following model:

$$(3)\ Y_{icgst} = \delta\, Y_{icgs,t-1} + \alpha_i + \sigma_s + \beta\, MDTP_{i,t-1} + \gamma\, Family_{it} + \rho\, ClsSize_{icst} + \pi\, School_{ist} + \eta\, Teacher_{icst} + \eta\, Grade_{it} + \tau Year_t + \theta\, Test_{it} + \varepsilon_{icgst},$$

and instrument $\Delta Y_{icgs,t-1}$ using $Y_{icgs,t-2}$. This more flexible approach yields qualitatively similar results to the main results when interpreted in gains in test scores (Table A2). The coefficient on the indicator for

[11] As shown in Table 1, in 2004, MDTP was mandated only for selected schools for grade 6 students and we treat them as a non-mandated cohort. 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 0.75 with a t-stat of 37.14.

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 0.109. This is still a sizeable coefficient. Columns (2) and (3) in Table A2 show that $MDTP_{t-2}$ has a positive estimated effect on the student test score, which is in fact quite consistent with the models shown in Table 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).[12]

5.3. Voluntary versus Mandated MDTP Testing

In this paper, we focus on the natural experiment in which the district phased in mandated district-wide MDTP testing in certain grades. But before and during this period, individual teachers in the district could voluntarily opt to use the MDTP tests for any of their classrooms. It is not clear how one could obtain a causal estimate of the impact of this voluntary testing, as it is clearly an endogenous choice made by teachers for specific classes. However, we did obtain data on this testing, and added it to our main student fixed effect model from column 1 of Table 4. We did not find a significant positive association between test-score changes and this voluntary testing, and in some specifications the coefficient was small but negative and significant. This could simply reflect endogenous choices by teachers to test classes that were performing poorly. Perhaps more importantly, we also found that the coefficient and standard error on mandated MDTP testing changed negligibly from the models we have shown above when controls for voluntary testing were added. Results are available from the authors on request.

Several conclusions emerge from this analysis. First, our earlier results are robust to controlling for voluntary MDTP testing. Second, there appears to be no positive association between this voluntary testing and student gains in math achievement, although we question whether our method could produce a causal relation. Third, one possible inference is that MDTP tests have a much greater impact when

[12] Note that because the lagged test score has a coefficient of -0.013, the predicted change in a student's Z-score after being tested for two years in a row would be -0.013(0.123) + 0.038 + 0.152 = 0.188.

conducted schoolwide and districtwide than when used in a piecemeal fashion by individual teachers for specific classes.

6. Potential Explanations for the Positive Effect of Mandated MDTP Testing

To delve further into why mandated MDTP might have a positive effect on gains in math achievement, we examined two elements related to tracking and remediation. First, the district used overall student performance during a year in addition to MDTP performance to make recommendations to parents on whether their children should attend summer school.[13] 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?

[Table 7]

Table 7 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

[13] Using a randomized field trial, Borman and Dowling (2006) find that attending summer school promotes longitudinal achievement growth.

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% increase in probability to a 2.1% 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.

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 7 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 8, in column (3) we include both lagged MDTP and summer school. Compared to column (2), the coefficient on lagged MDTP falls by 0.002 or about 1%. Thus, the effect of taking MDTP on summer school placement can explain perhaps only 1% of the overall effect of mandated MDTP. In column (4) we interact lagged MDTP and summer school, but the interaction is not significant.

[Table 8]

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 better placement, then we should expect to see the variance of test scores, within the set of students taking a certain class, fall if the students were given the MDTP in the prior spring. The first empirical test we conduct is to regress the standard deviation of the prior spring's standardized test score for each current year class on an indicator variable whether a student took MDTP the prior year, along with other control variables that were listed in the basic specification. The coefficient estimate suggests that taking MDTP the prior year reduces the within-class standard deviation by 0.035 and the effect is statistically significant at 1%. The mean standard deviation for the classroom's lagged math achievement is 0.67, so taking MDTP last year yields a modest decrease in within class variation (about 5%), by moving towards a more homogenized ability-group setting.

The second empirical exercise is therefore to see whether controlling for the dispersion of lagged test scores within the current year 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 9, the coefficient of lagged MDTP drops from 0.218 to 0.206 when the measure of within course/class dispersion is included. Thus, it appears that about 5.5% of the MDTP effect is coming from increased degree of sorting.

[Table 9]

We conducted robustness checks by estimating the results in Table 8 and 9 using the Anderson and Hsiao model (Table A3). In that model, summer school explains about 1% 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 2%. Overall, then, we conclude that about 1% of the effect of MDTP on subsequent math achievement can be explained by the role in plays in placing struggling students in summer school, and, using our measure of the variance of achievement in the actual math classroom, more accurate ability grouping of students by classroom can account for 2 to 5.5% of the effect of MDTP.

What can account for the unexplained portion of the estimated impact of mandated MDTP testing? Bachofer, Zau and Betts (2012) conducted a survey of middle and high school math teachers in SDUSD, and asked them about whether and how they made use of the results of mandated MDTP tests.

The results are striking. Fully 93 percent of teachers who reported administering MDTP tests under the district mandate reported making use of the results during that same school year, and 75% of teachers reported making use of mandated test results to assist students in the following school year.[14]

[14] Results cited here and below come from Table 3 of Bachofer, Zau and Betts (2012).

Within the year of the test, 75 percent of math teachers stated that they “reviewed results on my own to determine overall strengths and weaknesses”. Students and their families also received considerable feedback. For instance, 47 percent of teachers discussed the MDTP results with students in their classes. They also made considerable use of the student letters generated by the MDTP that are intended to inform students and their parents of students’ specific weaknesses and strengths. Almost half – 45 percent – of teachers reported distributing the MDTP student letters to students. Remarkably, given that the MDTP tests were given within one or two months of the end of the school year, 39 percent of math teachers reported that they “modified teaching to help students understand and correct misunderstandings and errors revealed by (the) test” before the end of the school year.

Given the fairly large impact of the mandated testing, given that we can explain only a portion of this effect through course placements and summer school placements, and given how late in the school year the tests took place, we should expect to see that during the school year after the test, math teachers would use the MDTP tests from the prior spring to help their current students. Bachofer, Zau and Betts (2012) indeed find that teachers made use of these tests in ways that are very much aligned with the stated goals of the MDTP. Three quarters of math teachers report that they used the previous spring’s MDTP results in some way during the current school year. Just under two thirds (58.5 percent) reported that they “reviewed results on my own to determine overall strengths and weaknesses”. Perhaps most importantly, 58.5 percent stated that they “modified teaching to help students understand and correct misunderstandings and errors revealed by (the) test”. Roughly a third of math teachers reported discussing the prior spring’s results at a formal meeting of the school’s math department and spending additional time working on areas in which his or her students performed poorly on the prior spring’s MDTP test.

We cannot prove that these factors account for the quite large achievement gains in the year following a mandated math test, but it seems clear from these survey results that math teachers actively used test results to help their students to improve in areas revealed by the tests as areas of weakness.

7. Conclusions

We find that district-mandated MDTP testing is associated with positive gains in mathematics achievement during the next year. In addition, findings suggest that just under a third of the apparent effect of mandated testing dissipates in tests a year and a half after the MDTP test, and most of the effect dissipates two years after taking the MDTP test. One inference from this pattern is that isolated MDTP testing may not be as helpful as repeated testing to keep students on track. This notion gains support from our finding that this depreciation is essentially eliminated if students take MDTP tests two years in a row.

The one-year effect sizes of 0.22 in our main model and 0.10 in an alternative approach shown in the appendix are meaningful. Gains of 0.10-0.22 standard deviations translate into gains of 4-9 percentile points in a single year for a student initially at the 50th percentile, and 3-7 percentile points for a student at the 25th percentile. Notably, Dee and Jacob (2011) estimate an effect size for the implementation of NCLB, including testing plus accountability provisions, of 0.23 in grade 4 math. Both of these estimated effects are on the same order of magnitude as the estimated impact of reducing elementary school class size from roughly 22-25 students to 13-17 students in the STAR experiment, which was 0.27 for math and 0.23 for reading (Finn and Achilles, 1990, Table 5). Note however the significant costs of reducing class size compared to the negligible costs of the MDTP's 45 minute test, offered free throughout California and online outside California for a nominal fee.

The MDTP tests were designed to help teachers identify students' specific strengths and weaknesses so that teachers can promptly provide remedial help to each student to overcome the identified misunderstandings. The district's vision in adopting MDTP testing near the end of the school year was slightly different – to use the tests to encourage remediation over the summer and to ensure that students are placed into an appropriate level of math class the following year. We find evidence that about 2 to 5.5% of the estimated impact of MDTP on math achievement can be attributed to classrooms that are more homogeneous in students' initial achievement. To the extent that the district used spring MDTP results in assignment of students to math courses, the improvements in ability grouping that resulted may have benefited students regardless of whether they were placed in a higher or lower track. For an example of such a finding, note that Duflo, Dupas and Kremer (2010) present experimental evidence that students in Kenya gain from ability grouping regardless of the group into which they are placed.

We also found that a very small amount -- roughly one percent of the overall effect -- could be explained by the increased attendance rate by weaker students at summer school if took the MDTP.

These two mechanisms, summer school and class placement, together explain roughly 2-6% of the quite sizeable effects of mandated diagnostic testing. What can account for the rest of the effect? In particular, given that the MDTP tests were given in May and June, what could have happened during the following academic year to account for the observed student gains? Based on a survey they conducted of middle and high school math teachers in the district, Bachofer, Zau and Betts (2012) find that teachers made use of the MDTP test results in numerous ways in the year of the test and, significantly, in the following year. For instance, teachers reacted to test results by speaking to students in their classes about the results, sending the MDTP student letters which detail students' results home with the students, spending time to understand their classes' overall strengths and weaknesses and, significantly, modifying their teaching to pay more attention to student misunderstandings revealed on the prior spring's diagnostic tests.

Another possible channel that is not explored in this paper is that students may work to improve their weaknesses in mathematics after the MDTP test. This is possible as students, not only teachers, receive detailed reports after the MDTP test. For instance, students may not be formally enrolled in summer school but may have been engaged in other activities that promote learning in a subsequent academic year. These activities, however, are unobserved in the data thus we could not explore this hypothesis further.

In sum, math diagnostic testing has meaningful positive effects, and two partial explanations for these effects are increased participation in summer school among students who fare poorly and in particular more accurate placement of middle school students into an appropriate level of math class in the subsequent year. Teachers report using the test results even in the next grade to understand their students' strengths and weaknesses and to adjust their teaching to address these weaknesses. These actions on the part of teachers could account for much of the observed impact of districtwide diagnostic testing. In this sense, we find that diagnostic testing, if it leads to specific interventions and more accurate tracking, has positive effects.

We believe that the context in which diagnostic testing is introduced is likely to mediate the impact of the testing. In their experimental study conducted in India, Muralidharan and Sundararaman (2010) find no effect of giving teachers test scores, but they point out that if the government of Andhra Pradesh adopted such a policy, the effects could be quite different. In the context we study, math teachers came to know that every student entering given grades had detailed diagnostic results available from the prior spring, and they appear to have used these data actively to boost the achievement of their students. [15]

Our finding seems timely given that most U.S. states have recently adopted one of two testing systems (Smarter Balanced and PARCC) that both aim to provide not only year end "summative" tests, but also more diagnostic "formative" assessments. These latter tests can be used by teachers throughout the school year to learn which skills students have mastered and which they have yet to truly learn. In this regard these formative tests are similar to the MDTP. Our MDTP results do not necessarily apply to these tests. But they offer important hope that diagnostic tests, when used systematically by a school district, can accelerate achievement.

Acknowledgements

This research was funded at arms' length by the California Academic Partnership Program (CAPP) at the California State University and the University of California. We thank Karen Bachofer,

[15] There were other important differences between the context of this study and ours. Schools in the study in India did not use the tests to place students. A second important difference is that in our case the central district administration embraced usage of the diagnostic testing rather than receiving the tests as part of an outside intervention. A third key distinction is that in our sample teachers and schools were subject to an accountability system and were incentivized to help boost test scores.

Julie Cullen, David Figlio, Hee-Seung Yang, Ron Rode, and Peter Bell, as well as participants in seminars at the annual meetings of the American Economic Association, UC San Diego, the CAPP Advisory Committee meeting, and the Asian Conference on Applied Micro-Economics/Econometrics at Sogang University for helpful suggestions. For their help answering our many questions about the MDTP we especially acknowledge the assistance provided by Alfred Manaster, the emeritus state director of the MDTP, Bruce Arnold, the current state director of MDTP, Donna Ames of the MDTP San Diego office, and David Jolly and Andrea Ball, emeritus and current directors of CAPP respectively. We also thank Hans Johnson, Richard Murnane, and Russell Rumberger for providing detailed comments on an earlier draft of a non-technical version of this paper.

Table 1: SDUSD Mandated Use of MDTP Tests by School Year, Test and Grade Levels

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	Pre-Algebra	6 at selected schools
2004	Algebra	7
2005	Pre-Algebra	6
2005	Algebra	7
2006	Pre-Algebra	6
2006	Algebra	7

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

Table2: The Proportion of Students Who Took MDTP for 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 and italics indicate the mandated years. * In school year 2004-2005, MDTP was mandated for 6th year students in selected schools.

Table 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 who was 8th grade in 2001-2002 academic year up to when they became 11th grade 2004-2005. Subject names are: AL1: Algebra 1; AL2: Algebra 2; Gmtry: Geometry; HsMath:High school math; I. Math: Integrated Math

Table 4: The Estimated Effect of Mandated MDTP on Gains in Math CST Scores

	(1)	(2)	(3)
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)
Observations	342571	311415	311415

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. Other regressors are as listed in the description of equation (1) in the text. Standard errors are clustered by school and shown in parentheses.* $p<0.10$ ** $p<0.05$ *** $p<0.01$

Table 5: 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: 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. Other regressors are as listed in the description of equation (1) in the text. Standard errors are clustered by school and shown in parentheses. 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. * p<0.10, ** p<0.05, *** p<0.01.

Table 6: IV and Intention to Treat Estimates

	(1)	(2)	(3)	(4)
	Baseline	IV	Intent to Treat	Reading
MDTP t-1	0.219***	0.277***		-0.003
	(0.019)	(0.032)		(0.008)
Mandate t-1			0.208***	
			(0.023)	
Observations	342571	342571	342571	343524

Note: Baseline estimate is from Table 4, column 1. The IV estimate is obtained by using Mandate t-1 as an instrument for MDTP t-1. The first-stage F-statistic is 1379 and corresponding p-value on exclusion of Mandate t-1 is smaller than 0.001. Dependent variable for columns (1)-(3) is gain in CST math test score, standardized by grade and test. The dependent variable for column (4) is gain in CST reading test score, standardized by grade and year. All regressions include student, grade, year, CST subject type and school fixed effects. Other regressors are as listed in the description of equation (1) in the text. Standard errors are clustered by school and shown in parentheses. * p<0.10 ** p<0.05 *** p<0.01

Table 7: 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. Other regressors are as listed in the description of equation (1) in the text. 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.* p<0.10 ** p<0.05 *** p<0.01

Table 8: The 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. Other regressors are as listed in the description of equation (1) in the text. 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 6 since dependent variable here requires two consecutive years of test score). * p<0.10 ** p<0.05 *** p<0.01

Table 9: Can the Effect of MDTP Testing on Math Achievement Be Explained by Resulting Changes in the Standard Deviation of Initial Achievement by Classroom?

	(1)	(2)
MDTP t-1	0.218***	0.206***
	(0.020)	(0.020)
SD for each class		-0.337***
		(0.027)
Observations	210635	210635

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. Other regressors are as listed in the description of equation (1) in the text. Standard errors are clustered by school and shown in parentheses. Class information for 6th to 11th graders are used (for 1st-5th graders, most students likely take one generic math course for their grade. Also, many students in 12th grade did not take the CST test so we exclude those rather than including selected students of who took CST test in later grades). * p<0.10 ** p<0.05 *** p<0.01

Appendix

Table A1: Summary Statistics

Variable	Mean	Standard Deviation	Minimum	Maximum
$MDTP_{t-1}$	0.135	0.342	0	1
$MDTP_{t-2}$	0.109	0.312	0	1
$MDTP_{t-1}$ * $MDTP_{t-2}$	0.019	0.135	0	1
Type of Math CST Test Taken				
Test==AL1	0.14	0.35	0	1
Test==AL2	0.07	0.25	0	1
Test==Math8/9	0.02	0.16	0	1
Test==Gmtry	0.11	0.31	0	1
Test==HsMath	0.03	0.18	0	1
Test==Integrated Math	0.02	0.13	0	1
Other CST test (grade-specific)	0.60	0.49	0	1
Average Math Class Size	14.53	14.08	0	47
School Demographics				
% of school Asian	16.78	14.64	0	100
% of school white	25.81	20.85	0	100
% of school Hispanic	41.42	23.11	0	100
% of school black	13.63	10.39	0	98.41
% of school free lunch	55.67	28.43	0	100
% of school English Learner	19.48	20.16	0	100
Teacher Characteristics in Math Class (or in Home Room in Elementary Classrooms or Other Self-Contained Classrooms), averaged across Semester				
Average years teaching at SDUSD	10.73	7.12	0	42
Average years teaching	12.39	7.77	0	46
Average percent bachelors in math	18.79	38.64	0	100
Average years service in math classes	7.37	9.96	0	44
Avg. SDUSD yrs service in math classes	6.35	8.93	0	39
Avg. full credential among math teachers	54.37	49.71	0	100
Average teacher intern	0.86	8.93	0	100
Average full authorization in math	0.29	0.45	0	1
Avg. supplemental authorization in math	0.12	0.32	0	1
Average female teacher	29.22	44.86	0	100
Average CLAD certificate	24.65	42.57	0	100

Average any Master's degree	28.44	44.68	0	100
Average of white teachers	41.45	48.88	0	100
Average of black teachers	3.20	17.23	0	100
Average of Asian teachers	6.67	24.54	0	100
Average of Hispanic teachers	3.19	17.33	0	100
(Potentially) Time-Varying Student Characteristics				
English learner	0.23	0.42	0	1
Grade Level				
3	0.12	0.33	0	1
4	0.12	0.33	0	1
5	0.12	0.33	0	1
6	0.11	0.32	0	1
7	0.11	0.32	0	1
8	0.11	0.32	0	1
9	0.11	0.31	0	1
10	0.10	0.30	0	1
11	0.08	0.28	0	1
School Year				
2002-03	0.20	0.40	0	1
2003-04	0.21	0.41	0	1
2004-05	0.21	0.41	0	1
2005-06	0.20	0.40	0	1
2006-07	0.19	0.39	0	1
Parental Education				
Less Than High School	0.16	0.37	0	1
High School	0.18	0.39	0	1
Some College	0.16	0.36	0	1
College Graduate	0.18	0.38	0	1
Postgraduate College	0.10	0.30	0	1
Missing	0.22	0.41	0	1

Note: N = 342,571 except for $MDTP_{t-1}$ (N=311,415) and $MDTP_{t-1}$ * $MDTP_{t-2}$ (N=311,415).

Table A2: The Estimated Effect of MDTP on Test Scores: Anderson-Hsiao Models of Overall Effect and Effect by Type of MDTP Test

	(1)	(2)	(3)
MDTP t-1	0.109***	0.142***	0.123***
	(0.014)	(0.012)	(0.013)
MDTP t-2		0.064***	0.038**
		(0.016)	(0.017)
MDTP t-1 * MDTP t-2			0.152***
			(0.023)
Lagged test gains	0.000	-0.011	-0.013
	(0.027)	(0.028)	(0.027)
Observations	302826	275763	275763

Note: Estimates are based on the first-differenced version of equation (3) in the text. Regressors are as listed in the description of equation (1) in the text. First-differenced school fixed effects are partialled out. Standard errors are clustered by school and shown in parentheses. * p<0.10 ** p<0.05 *** p<0.01

Table A3: The 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	0.100***	0.099***	0.102***	0.100***
	(0.014)	(0.014)	(0.015)	(0.014)
Summer		0.012		
		(0.008)		
SD for each class				-0.087***
				(0.020)
Observations	123946	123946	140955	140955

Note: Estimates are based on the first-differenced version of equation (3) in the text. Regressors are as listed in the description of equation (1) in the text. First-differenced school fixed effects are partialed out. Standard errors are clustered by school and shown in parentheses. Samples are restricted to grades 7 to 11 in columns (1) and (2). Samples are restricted to grades 6 to 11 in columns (3) and (4). * p<0.10 ** p<0.05 *** p<0.01.

References

Anderson, T.W. and Cheng Hsiao, “Estimation of Dynamic Models with Error Components,” *Journal of the American Statistical Association*, 76:375, pp. 598-606. 1981.

Ashenfelter, Orley. "Estimating the Effect of Training Programs on Earnings." *The Review of Economics and Statistics*, 60:1 pp. 47-57, 1978.

Bachofer, Karen Volz, Andrew C. Zau, and Julian R. Betts. “The Impact of the Use of the Mathematics Diagnostic Testing Project in San Diego Unified School District: Teacher Survey Component,” Long Beach, CA: California Academic Partnership Program, 2012

Betts, Julian R., Andrew Zau and Lorien Rice, Determinants of Student Achievement: New Evidence from San Diego, San Francisco, Public Policy Institute of California, 2003.

Black, Paul and Dylan Wiliam, “Assessment and Classroom Learning,” *Assessment in Education: Principles, Policy & Practice*, 5:1, 7-74, 1998.

Borman, Geoffrey D. and N. Maritza Dowling, “Longitudinal Achievement Effects of Multiyear Summer School: Evidence from the Teacher Baltimore Randomized Field Trial,” *Educational Evaluation and Policy Analysis*, 28, pp. 25-48, 2006.

Collins, Anne M., Antonio Cameron, Jane Gawronski, Mary Eich, Sharon McCready, *Using Classroom Assessment to Improve Student Learning*, Reston, VA: National Council of Teachers of Mathematics.

Dee, Thomas, and Brian Jacob, “The Impact of No Child Left Behind on Student Achievement,” *Journal of Policy Analysis and Management*, 30:3, pp. 418-446, 2011.

Duflo, Esther, Pascaline Dupas, and Michael Kremer, “Peer Effects and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya,” *American Economic Review*, 2010.

Finn, Jeremy D. and Charles M. Achilles, “Answers and Questions about Class Size: A Statewide Experiment,” *American Educational Research Journal*, 27:3, pp. 557-577, 1990.

Muralidharan, Karthik and Venkatesh Sundararaman, “The Impact of Diagnostic Feedback to Teachers on Student Learning: Experimental Evidence from India,” *Economic Journal*, August 2010, 120:546, pp. F187-F203.