Title
Examining the Implementation and Impact of "Double Dose" Math Courses in a Mid-Sized, Suburban School District

Author
Lewis, Ryan W.

Publication Date
2018

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,
IRVINE

Examining the Implementation and Impact of "Double Dose" Math Courses in a Mid-Sized, Suburban School District

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Education

by

Ryan W. Lewis

Dissertation Committee:
Professor George Farkas (Chair)
Associate Professor Drew Bailey
Associate Professor Thurston Domina
Associate Professor Andrew Penner
Assistant Professor Di Xu

2018

DEDICATION

"How does one organize an expedition: what equipment is taken, what sources are read; what are the little dangers and the large ones? Your expedition will be enclosed in the physical framework of start, direction, ports of call, and return. These you can forecast with some accuracy; and in the better-known parts of the world it is possible to a degree to know what the weather will be in a given season, how high and low the tides, and the hours of their occurrence. One can know within reason what kind of boat to take, how much food will be necessary for a given crew for a given time, what medicines are usually needed—all this subject to accident, of course."
- John Steinbeck, *The Log from the Sea of Cortez*

To Linda, for being so patient while I've been sailing.
And to my parents, for supporting every expedition I take.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDIX TABLES

LIST OF APPENDIX FIGURES

LIST OF APPENDIX DOCUMENTS

ACKNOWLEDGMENTS

I would like to thank my committee chair, George Farkas, for his commitment to this project and the endless amounts of support he offered throughout the process. And to the rest of my committee, I couldn't have asked for a more supportive and caring team. Thad Domina, Andrew Penner, Di Xu, and Drew Bailey—thank you for your encouragement, your smarts, and your patience with my questions and concerns. I wouldn't have made it without you.

Special, additional thanks to Thad for employing me all these years, pushing me to develop this dissertation from its infancy, and for helping me find my way as a grad student and a researcher.

For all-around support, mentoring, and necessary good times along this journey, I'd like to thank my fellow graduate students, especially my class/homework/lunch companion Tutrang, as well as the EQUAL research team and the Monday volleyball squad.

Last but not least, extra scratches to Rooney Bear, my fuzzy sidekick and ultimate writing buddy. Thank you for keeping my feet warm, for reminding me to go outside for walks, and for always staying up as late as I did.

CURRICULUM VITAE

Ryan W. Lewis

2003	B.A. in Political Science, University of California, Los Angeles
2004-06	AmeriCorps National Service
2008	Master of Public Service, University of Arkansas, Little Rock
2008-13	Director of Research and Evaluation, 826 National
2016	M.A. in Education, University of California, Irvine
2018	Ph.D. in Education, University of California, Irvine

PUBLICATIONS

Lee, K.T.H., Lewis, R.W., Kataoka, S., Schenke, K., and Vandell, D.L. (2018). Out-of-School Time and Problem Behaviors During Adolescence. Journal of Research on Adolescence, DOI 10.1111/jora.12389.

Lewis, R.W. and Farkas, G. (2017). Using an Opportunity-Propensity Framework to Estimate Individual-, Classroom-, and School-Level Predictors of Middle School Science Achievement. Contemporary Educational Psychology, https://doi.org/10.1016/j.cedpsych.2017.08.003.

Domina, T., Lewis, R., Agarwal, P., and Hanselman, P. (2015). Professional Sense-Makers: Instructional Specialists in Contemporary Schooling. Educational Researcher, 0013189X15601644.

ABSTRACT OF THE DISSERTATION

Examining the Implementation and Impact of "Double Dose" Math Courses in a Mid-Sized, Suburban School District

By

Ryan W. Lewis

Doctor of Philosophy in Education

University of California, Irvine, 2018

Professor George Farkas, Chair

In hopes of boosting achievement for students with low math skills, a large number of schools and districts have employed the policy of “double dose” math courses (an additional math instruction period during the school day) in middle or early high school. We know very little about the effectiveness of this intervention and whether previously-reported impacts are generalizable to different settings and implementation structures. In the current dissertation, I use a mixed methods approach across three studies to focus on an individual school district that implemented, discontinued, and then re-implemented double dose math courses as an intervention for struggling middle school students.

In Chapter 1, I analyze the effectiveness of the original implementation of seventh-grade support courses on math achievement for low performing math students and find no significant impacts for students very close to the enrollment cutoff but modest, positive intent-to-treat effects on standardized test scores across all eligible students. I also find increased achievement gains for treatment students with the lowest prior math achievement and English Language Learners. In Chapter 2, I develop a case study on the decision-making processes of the school

district administrators within this time period using research on organizational sensemaking (Vaughan, 1996; Weick, 1995) and collective sensemaking in educational settings (Coburn, 2001) as a theoretical framework. Decision making regarding the policy was shown to be sensitive to changing state policy and curricular standards, student test scores, and pressures from resources and organizational structure. In Chapter 3, I document the changes to the intervention's implementation structure and test the impact of the new version. I find modest, positive benefits on standardized test scores for treated students (especially intervention students with the highest previous math achievement) but no impacts on math course grades or failure rates.

Together, these studies add depth to the literature on double dose math courses as an intervention for math achievement, analyze a unique policy environment in which two different versions of the same intervention are compared, and offer insights into the decision-making process regarding this policy and curricular policies in general. Policy considerations and ideas for future research are also discussed.

INTRODUCTION

In hopes of boosting achievement for students with low math skills, a large number of schools and districts have employed the policy of "double dose" math courses, an additional math instruction period, in middle or early high school. Across three studies, I focus on an individual school district that offers a unique setting for studying this policy. In 2008-09, this a midsized, suburban school district implemented double math courses within their middle schools as an intervention for low performing students. In 2010-11, two schools decided to discontinue offering these courses and by 2013-14 these courses were almost entirely discontinued. In 2015-16, district administrators decided to reinstate the intervention the following year with considerable structural changes to the way the courses were designed and offered from the first iteration to the second.

In Chapter 1, I use administrative data obtained from the school district to analyze the effectiveness of the original implementation of seventh-grade Pre-Algebra double dose courses on math achievement for low performing math students. I employ two quasi-experimental approaches to answer a primary research question: Were these loosely structured "double dose" math courses a successful intervention for low-performing seventh graders in this school district? In Chapter 2, I focus on the decision-making processes of the school district administrators within this time period using research on organizational sensemaking (Vaughan, 1996; Weick, 1995) and collective sensemaking in educational settings (Coburn, 2001) as a theoretical framework to construct a case study. My primary research questions for this case study are: How did district administrators engage in collective sensemaking to re-implement a double dose math course intervention? How were decisions about the altering and re-implementation of "double dose" math classes made and who made them? How and what did district administrators learn as they modified the policy?

Finally, in Chapter 3, I use administrative data obtained from the school district to study the effectiveness of the re-implemented intervention. I employ a propensity score weighting design to answer three primary research questions: Were the newly-implemented "double dose" math courses a successful intervention for low-performing seventh graders in this school district? Were the intervention courses uniquely impactful for students designated as English Learners, socioeconomically disadvantaged students, or students at different points along the prior math achievement spectrum? How do the results compare with those of the original implementation? Together, these studies add depth to the literature on double dose math courses as an intervention for math achievement, analyze a unique policy environment in which two different versions of the same intervention are compared, and offer insights into the decision-making process of this policy and curricular policies in general.

CHAPTER 1

Does More Math Instruction Always Help? Evaluating Seventh-Grade Double Dose Math Courses in a Midsized, Suburban School District

INTRODUCTION

Overview

According to the Nation's Report Card, 67% of United States eighth graders are not considered to be proficient in age-appropriate math skills (National Center for Education Statistics, 2015). Low overall math proficiency at this age is connected to another concern for educators and policy-makers: failure rates in middle and high school math courses, notably Algebra. In California, for example, 2008 state testing data showed that 44% of ninth through eleventh graders taking Algebra were repeating the course (Rosin, Barondess, & Leichty, 2009). Failing math courses during this transitional time for students is especially concerning because of established links between early high school course failure and lower high school graduation rates (Allensworth & Easton, 2007; Bottoms, 2008).

As a result, middle and high school educators face a difficult task. Most of their students are not considered proficient in the foundational math skills necessary to pass their math courses, but the stakes for failing a math class during these years are dire. Enrolling students in a "double dose" math class—a second, complementary instruction period during the school day—has become increasingly common for schools serving sixth through ninth graders (Durwood, Krone, & Mazzeo, 2010). A recent report on schools in North Carolina shows that ninety-six percent of high schools, including at least one high school in each district, and fifty-eight percent of middle schools within the state utilized double dose math courses in 2011-12 (Henry, Barrett, & Marder, 2016).

The schools or districts implementing these courses are most commonly employing them as remediation opportunities for low-performing students, although there are cases of schools using them as reinforcement for learning for all students (Henry, Barrett, & Marder, 2016). When schools enact this policy for remediation purposes, students who are assessed as below proficient on a given measure or measures are enrolled in a second math period instead of another required course or elective. In this second class period, educators expect that students will receive extra support on the current content and remedial instruction for gaps in foundational knowledge to eventually bring them up to speed with peers.

Although schools and districts are implementing double dose math courses more widely, very few empirical studies of their effectiveness as an intervention for low performing students exist. The bulk of the empirical studies on these courses are from two large, urban settings (Chicago Public Schools and Miami-Dade County Public Schools) in which the findings were generally positive (Cortes, Goodman, & Nomi, 2015; Nomi & Allensworth, 2009; Taylor, 2014). However, the unknown generalizability of this research introduces large questions for educators. Should we expect "double dose" math classes to work in all types of public schools? Are there aspects of implementing this intervention that are crucial to its success? Is the intervention more helpful for any particular types of students? I explore these questions by analyzing the implementation and effects of a double dose math course policy in a mid-sized, suburban, Southern California school district serving a racially and socioeconomically diverse student population and implementing the policy as a loosely-structured offering (e.g. considerable teacher autonomy in curriculum decisions, different teachers for traditional and intervention course, non-successive intervention periods).

Using administrative data obtained from the school district, I employ two quantitative approaches to answer a primary research question: **Were these loosely structured "double dose" math courses a successful intervention for low-performing seventh graders in this school district?** First, I utilize a regression discontinuity design to create a quasi-experimental setting in which I can compare the standardized test scores, math grades, and course failure rates for students who fell just below and above the assignment threshold for this intervention. Next, because of a unique situation in which two schools suddenly discontinued offering the courses in 2010-11, I utilize a secondary quasi-experimental approach using a difference-in-differences-in-differences design that identifies whether policy implementation impacted these outcomes across the student population of students intended for treatment. To further investigate the difference between the results from these two approaches, I conduct a follow-up analysis using basic OLS regression models to estimate whether enrollment was more impactful for students at certain points within the prior achievement distribution or for English Language Learners.

The results show that, for students within a close bandwidth to the district-designated assignment cutoff, there is no significant impact of the intervention on end-of-year standardized math assessment scores, grades in the traditional seventh-grade math course, or traditional math course passage rates. However, more inclusive approaches using difference-in-difference-in-difference and OLS regression models indicate potential test score increases for double dose students near the bottom of the prior math achievement distribution (outside of the regression discontinuity bandwidth) and English Language Learners. Although these are seemingly contradictory findings, they make sense when considering the students analyzed within each method. For students that were previously scoring high enough to almost not enroll in the course, the intervention was not impactful. However, when looking across the entire sample of eligible students, positive effects can be detected. It is likely that even a loosely structured offering of this intervention is still beneficial for very low math performers and English Learners, but those benefits may not carry over to students at the higher end of the prior achievement spectrum. These findings suggest that setting and policy implementation decisions can significantly alter the effectiveness of the intervention and should be considered more carefully.

Literature Review

The best-known research examining this type of intervention comes from analyses of the Chicago Public Schools' adoption of double dose math classes in the mid-2000's and one study on the implementation of the policy in the Miami-Dade County Public Schools later that decade. In Chicago, ninth grade double dose algebra courses were utilized in 2003 and 2004 with some unique characteristics. Teachers for double dose courses were given special professional development sessions and curricular resources. Math courses were also scheduled as much as possible with the same teacher in successive class periods for the traditional and support classes. For most students, regardless of whether they were targeted by the policy, the scheduling demands of these strict implementation guidelines caused their traditional math course placement (for treatment students this was their non-support course) to be more homogenous by peer ability level.

Varying teams of common authors completed multiple studies of the Chicago Public Schools' use of double dose algebra classes (Cortes, Goodman, & Nomi, 2015; Nomi & Allensworth, 2009; Nomi & Allensworth, 2013; Nomi & Raudenbush, 2016). As a whole, the Chicago analyses showed that double dose support courses positively influenced students' short- and long-term math skills, future math course taking, and outcomes related to high school completion and college enrollment. These analyses also showed that the impact of the policy was highly sensitive to the way students were sorted into their traditional and complementary math course placements. Cortes, Goodman, and Nomi (2015) also highlight the effectiveness of the policy for students with below average reading skills, introducing the idea that it might be more helpful for students who just need extra time because of language barriers.

Taylor (2014) analyzed double dose math classes for sixth- through eighth-graders in the Miami-Dade County Public Schools District. Miami middle schools used double dose classes as an intervention for struggling middle school students before crucial high school math classes. Compared to the implementation of these courses in Chicago, Miami-Dade County Public Schools recommended less strict scheduling guidelines. The students' second math courses were smaller in enrollment (17 pupils/class vs. 21 pupils/class) and 90% of students had a different teacher, different peers, or both for their traditional and support math courses. Traditional math classrooms were much more likely to be skills-heterogeneous environments than in Chicago. Additionally, applicable teachers were given guidelines for the intentions of the second math course but there is no mention of any professional development in Taylor's (2014) analysis.

Using the test score cutoff for double dose enrollment to employ a regression discontinuity analysis, Taylor (2014) found immediate positive effects on math test scores (.16 to .18 SD), with those gains shrinking considerably over the following two years. The study found no impacts on 9th grade Algebra I completion or Algebra II completion by the end of high school, as well as no impact on immediate reading scores or eventual high school non-math outcomes. Although the short-term effects are comparable in size, these null findings on long-term outcomes counter the positive long-term findings from Cortes Goodman, and Nomi's (2015) analyses in Chicago. Considering the difference in implementation strategies used by the two districts, this highlights the potential importance of a district's implementation decisions in regard to variation in the effectiveness of the policy.

Other studies on similar interventions further this idea. Bartik and Lachowska (2014) used a regression discontinuity design to study "double blocked" sixth grade math and reading instruction in a mid-sized Midwestern school district. These students were recommended to a second reading or math class based on prior test performance but could not enroll in both. The researchers found a significant increase in reading test scores for those taking double reading courses (+.20 SD) but no significant change for those in double math class. However, the lack of significant findings for math students could be the result of the district's implementation strategy. Since it was deemed as too much remediation to have students take double courses in both, students only took double math instruction if they passed the reading requirement. Thus, many students who received reading remediation also needed math remediation but did not take the course. It is still important to note that these students who only needed math remediation received no benefits from this intervention targeted at helping students reach math proficiency.

Fryer (2012) used difference-in-difference and instrumental variable approaches to study the impact of infusing five successful charter school methods into nine Houston public schools: longer school days, human capital development amongst school staff, use of data-informed instruction, creation of a culture of high expectations, and high dosage tutoring or doubled instruction in key subjects. The last method consisted of small-group math tutoring in sixth and ninth grade or an extra dose of reading or math instruction to students in grades 7, 8, 10, 11, and 12 who had previously performed below grade level. Students received either the double dose of math or reading based on which subject they were further below proficiency.

Students enrolled in these schools made dramatic improvements in nearly every measurable category, but the effects cannot be attributed directly to doubled instruction since other methods were also implemented. However, the author conducted difference-in-difference examinations within double dose years that compared the gains of students taking a second class to those who didn't need to. Only one subject-grade combination yielded a significant, positive result: 8^{th} grade math (+.24 SD). This examination contains sample issues similar to Bartik and Lachowska (2014), as students were enrolled in double dose for their worst subject between Math and English, even if they qualified for both. This double dose tutoring was also in conjunction with other school interventions, so the impact of the program alone cannot be isolated. Another interesting note is that the effect of these methods as a whole is lower in years when students received doubled instruction in comparison to tutoring.

Table 1.1 summarizes the settings, implementation decisions, findings, and limitations from the key studies on double math courses. These are far from definitive. The results vary from null to moderately positive and the generalizability of the positive findings outside of the studied districts is unknown. Despite wide adoption across schools of all varieties, this intervention has only been proven effective in two large, urban school districts. Empirical studies in both a mid-sized district and charter school settings were not ideal and only found one discernible effect: increased math test scores for eighth grade charter school students taking double math classes in comparison to those who did not.

Among the study settings, there is notable heterogeneity in implementation strategies and treatment effects. The strongest and longest lasting treatment effects were found in the Chicago Public Schools, where teachers were offered rigorous professional development and the implementation guidelines were the most detailed. As much as possible, the students were scheduled to be with the same teacher in successive class periods for their regular and support classes. However, this also led to more instances where treatment students were grouped with other treatment students in traditional math courses, which researchers found to be beneficial for treatment students and detrimental for non-treatment students. Whether or not treatment students are segregated into classes with similarly low ability level peers appears to be an important aspect of this policy's implementation.

One question that needs to be investigated is whether the structure and guidelines used to implement this intervention are crucial to its benefits. Is the mere act of receiving extra math instruction all that matters, or do the decisions on how these courses are offered also matter? Of the existing studies, the strongest positive impacts of the intervention were found in the district that implemented the courses with the most rigid structure (Chicago Public Schools). In a setting with less structured guidelines in place (Miami-Dade County Public Schools), impacts were still positive but smaller in size. Another research question asks whether this intervention is more successful with particular student subgroups, and less successful with others. Previous work has identified poor readers as a group for which it might be more impactful (Cortes, Goodman, & Nomi, 2015), but no further subgroups have been tested. Finally, existing studies have limited generalizability, as the intervention has principally been studied in large, urban school districts. Additional studies assessing the impact of the intervention for students in different types of districts are needed.

Current Study

In the current study, I analyze the impact of seventh-grade Pre-Algebra double dose courses on math achievement for low performing math students in a midsized, suburban school district in Southern California. This research setting is unique to the literature, as is the district's implementation of the policy. The district analyzed contains a different student demographic profile than those studied previously, a minimally structured implementation strategy (flexible curriculum, different teachers in main and double courses, less organized student sorting), and a focus on courses offered in an earlier grade level than was the case in previously studied districts.

I conducted two types of quasi-experimental identification strategies with student-level data from this district to test the following questions:

1. Using a regression discontinuity design comparing students just above and below the cutoff for enrollment, were double dose math classes impactful on standardized math test scores, main math course grades, or main math course pass rates?
2. Using a difference-in-differences-in-differences design to estimate intent to treat (ITT) effects at the policy level, were the courses impactful on these outcomes?

These analyses add much-needed depth to the literature on “double dose” math courses as an intervention for low-performing students. The loose implementation structure within this study’s setting provides a useful comparison with previous studies conducted in settings employing more rigid implementation guidelines. In this study I also investigate whether this intervention is particularly effective for population subgroups. Although it is impossible to pursue a causal answer to this question within these data, I use ordinary least squares (OLS) regression models to investigate the relationship between treatment and benefits for students at different points along the math achievement distribution as well as for English Learners.

DATA

Setting

“Sunnyside School District” serves a current enrollment of more than 45,000 students, placing it within the top twenty-five largest school districts in the state of California. Approximately 75% of students qualify for free- or reduced-price lunch, more than 50% are Hispanic/Latino, and close to 33% are Asian or Asian American. Within these ethnicity groupings, the district enrolls large groups of first- and second-generation Mexican, Guatemalan, Salvadoran, Vietnamese, Hmong, Korean, and Chinese students. This study focuses on students within Sunnyside School District’s ten intermediate schools (serving grades 7 and 8) between the academic years of 2009-10 and 2012-13. This district is an ideal setting for studying this policy because of their unique implementation decisions involving double dose math courses over this time span. In 2008-09, the district implemented double math courses within their middle schools as an intervention for low performing students. In 2010-11, two schools decided to discontinue offering these courses and by 2013-14 these courses were almost entirely discontinued.

The district implemented the courses using a very minimal structure. As noted in Table 1.1, there is notable heterogeneity in the course content and implementation decisions of districts offering these types of courses. This is not surprising; schools are serving increasingly diverse student populations and operating in unique funding and policy climates (Jackson, Johnson, & Persico 2014; Meyer, Rowan & Meyer 1978; Reardon & Owens, 2014; Reed, 2014). Two neighboring schools that are similar in many ways might be radically different settings for policy implementation because of their internal social and administrative dynamics (Bryk, Sebring, Allensworth, Luppescu, & Easton, 2010). Compared with other districts studied, Sunnyside School District introduced the intervention with far fewer guidelines. Teachers were offered resources to help guide their instruction, but no particular curriculum or instructional units were mandated. Professional development for teachers was encouraged but not required. The district also made no determined plan to match students with the same teacher or peers in their traditional and intervention math courses, although such matching did sometimes occur. Analyzing the impact of the policy in a setting that features a great deal of decision-making autonomy for principals and teachers is a strength of this study.

Data & Sample

This study is based on administrative data provided by the school district. These data, which the district collects and reports for accountability purposes, tracks student enrollment from elementary through middle school and includes: student demographics (gender, race/ethnicity, language status, free/reduced price lunch eligibility); annual California Standards Test (CST) test scores; scores from district benchmark assessments; and transcript data on middle school math course, teacher ID, grade, and course period. I used longitudinal student-level data for all seventh-grade students enrolled in the academic years from 2008-2009 through 2012-13. Double dose courses were also offered to eighth-grade students during these years. Since schools offered eighth-grade intervention courses less frequently and some students were exposed to previous treatment in seventh grade, I did not include eighth-grade students in any analyses.

Certain students and schools were excluded from the sample using in the primary analyses. Schools in the district only offered double dose courses as a complement to Pre-Algebra classes in seventh grade. Those students who had already completed Pre-Algebra or were specifically placed into special education or other intervention classes at a lower level than Pre-Algebra were excluded from the sample because they were ineligible to take the class. Additionally, schools within the district that did not offer the double dose course or offered it at a very low rate in a given year were excluded during that year. I chose to exclude schools that did not enroll at least 33% of eligible students in the course; below that rate, there is not a clear discontinuity in treatment at the recommendation cutoff nor a consistent pattern of enrollment in the course following the placement guidelines. After all exclusions, 5,946 students are present in the sample. Out of ten total middle schools, the number of included middle schools ranges from as high as seven to as low as three in the final year.

Table 1.2 displays the sample size, number of participating middle schools, demographic characteristics, and percentage of students enrolled within a double dose math course during these years for observations that remained in the sample. The students within the sample are over 60% Hispanic/Latino, close to 30% Asian or Asian-American, and more than 75% qualify for free- or reduced-price lunch. Close to one-third of the students within the sample are English Language Learners and 3% are designated as Special Education students. The demographic characteristics of the sample, those students eligible for the intervention and attending a school that participated in offering it, compare closely to those of the district as a whole. The sample is by no means state- or nationally-representative; the sample schools enroll a disproportionately large number of Latino and Asian-American students and a correspondingly small number of white and African-American students in comparison with the population of the United States.

METHODS AND MEASURES

Fuzzy Regression Discontinuity Design

To estimate a causal effect of double dose course enrollment on standardized test score changes, I utilize a regression discontinuity design (RDD) that simulates a randomized control trial of treatment versus control students within a designated bandwidth of assignment scores. An RDD provides a causal estimate of the impact of a designated treatment by using a variable that identifies a cutoff for access to the treatment. If there is a clear discontinuity in the take-up of the treatment when crossing a certain value in the predictor variable, and as long as the relationship between the predictor and the outcome is expected to be otherwise smooth across this cutoff, then the variation in outcome when crossing the cutoff can be identified as the treatment effect (Imbens & Lemieux, 2008). To account for incomplete compliance with the program I used a fuzzy regression discontinuity design (Imbens & Lemieux, 2008; Lee & Lemieux, 2010). A fuzzy RDD utilizes an instrument to predict treatment in the first-stage, and then uses the predicted value for treatment as the treatment variable in the RDD calculations.

Use of this model implies three testable assumptions. First, selection into treatment is strongly determined by the identified placement strategy, which creates a discontinuity at the decided cutoff in the percentage of students either receiving or not receiving the treatment. Second, there should be no discontinuity in student density at or around the assignment cutoff. Finally, there should be no discontinuity in other characteristics or outcomes that should not be influenced by the treatment. If these assumptions hold, the discontinuity analysis provides internally valid estimates of the causal effects of seventh grade double dose math classes for students near the assignment cutoff (Lee, 2008; McCrary 2008).

First, I tested the assumption of the functional form of a discontinuity at the cutoff by checking a loaded regression model regressing outcome math scores on squared and cubed versions of the running variable along with interactions with the treatment variable. Appendix Table 1.1 displays the estimates from this model in which none are statistically significant. Next, I tested for manipulation in density of students around the designated cutoff score using the McCrary density test (McCrary, 2008) and found no evidence of unusual grouping on either side. Figure 1.1 displays the student density on both sides of my assignment cutoff. Finally, I checked for differences in demographic characteristics between the quasi-experimental treatment and control students within my chosen bandwidth. Appendix Figure 1.1 and Appendix Table 1.2 display the visual discontinuities in demographic characteristics across the running variable scores and the results of one-tailed t-tests comparing students within the bandwidth on both sides. There is only one characteristic of concern, as students below the cutoff are significantly more likely to be English Learners than students above (47% vs. 37%). I will address this concern as I discuss results involving different bandwidths and later results from interaction models focused on English Learners.

The key variables for a fuzzy RDD are the outcome, treatment, instrument, and running (or assignment) variables. Per Lee and Lemieux (2010), I calculated the probability of treatment using the instrument as

$$\Pr(D = 1|X = x) = \gamma + \delta T + g(x - c), \qquad (1)$$

where $T = 1\ [X \geq c]$ identifies whether the running variable X is greater than the threshold for recommendation to enroll at c. The fuzzy RDD estimates are calculated using a two-equation system:

$$Y = \alpha + \tau D + f(X - c) + \varepsilon, \qquad (2)$$

$$D = \gamma + \delta T + g(X - c) + \nu \qquad (3)$$

where Y is the outcome of interest, D is the probability of treatment, and τ is the treatment effect. When completed, fuzzy RDD estimates represent the local average treatment effect (LATE); this is the effect of the treatment for compliers within the specified bandwidth.

For this analysis, the main outcomes are the California Star Test (CST) annual standardized math test score in seventh grade, the final grade (in grade point average points) in the main (non-intervention) seventh-grade math course, and a dummy variable indicating a passing grade in the main seventh-grade math course. The treatment of enrolling in a double dose math course is coded with dummy variable indicating whether the student received double dose math instruction at any point during the school year. I used a dummy variable indicating recommendation to enroll as my instrument; students were coded with a one if they were recommended to enroll in a double dose course and a zero if they were not.

The running variable and cutoff score is based on the intervention assignment strategy score as calculated by school district policy. Scores from the students' fifth grade CST standardized math assessment and first two quarters of sixth grade district benchmark assessments are rated on three individual 1-5 scales. These ratings are summed, giving the student a score between 3-15 for double dose recommendation. If the student falls below nine points, they are recommended to enroll in the course. If they score nine points or above, they are not. However, since this variable only has 13 potential values, it is not ideal to use as a running variable for a regression discontinuity design, as the values are too rigid in grouping large numbers of students that may not be similar enough for quasi-experimental purposes. Rather, I can construct an alternative assignment variable based on these test values. This zero-centered assignment variable uses only one of the three test scores involved in the original assignment strategy.

After excluding students whose third test score could not decide whether they fall on one side or the other of the recommendation cutoff, I use a zero-centered version of the 5th grade standardized test score based on how many points a student would have needed to be waived out of recommendation to the intervention. Of the three possible tests, this test has the most variation in values and was closest in size of treatment discontinuity at the cutoff to the original assignment variable. This new assignment variable allows me to decrease the bandwidth of students included in the RDD estimates to only include students whose placement on one side of the cutoff or the other could be as good as random. The distribution of assignment strategy scores around the zero-centered cutoff are displayed in Figure 1.1. Figure 1.2 shows the discontinuity in the treatment at the assignment strategy cutoff is a 24% decrease in number of students enrolling in double dose courses (dropping from 44% to 20% enrollment). Because recommendation to enroll in the program was voluntary, 57% of recommended students actually enrolled in the course at some point during the year. Figure 1.3 shows the same discontinuity in treatment at the assignment strategy cutoff for the original assignment variable used by the district. Compared to Figure 1.2, this shows that use of the alternative assignment variable does not drastically alter the relationship between assignment score and course take-up. In both figures, the relationship between the two variables and discontinuity at the assignment cutoff is very similar. Using the alternative version of the variable simply lends itself better to the assumptions of the fuzzy regression discontinuity design.

Selection of an analysis bandwidth is also required for the regression discontinuity design. I chose a bandwidth of forty-four assignment strategy points, localizing the regression discontinuity analysis to only the students that were surrounding the cutoff by that many points above and below (n=805). To select this optimal bandwidth, I used the rdbwselect package within STATA software (Calonico, Cattaneo, Farrell, and Titiunik, 2016) that follows the recommendations for selecting robust, nonparametric confidence intervals for regression discontinuity designs as laid out by Calonico, Cattaneo, and Titiunik (2014). I also confirmed that this bandwidth offered a balance of including a large enough sample (about 20% of the potential students) to detect impacts, nearly identical demographic characteristics between simulated treatment and control students, and a necessarily high F-statistic of 35.9 for instrumental variable of recommendation to enroll. I will also present estimates using bandwidths of 22 points and 88 points for comparison as robustness checks.

Although the estimates from fuzzy regression discontinuity model can be interpreted as causal, the design is limited in external validity. Treatment effects are limited to complying students within the designated bandwidth around the assignment cutoff, which only encompasses a fraction of the entire student population. To understand whether the treatment was effective for all students eligible for the intervention, I must use a different analytic approach.

Difference-in-Differences-in-Differences Design

In a second quasi-experimental design, I can exploit the fact that two schools suddenly stopped offering the double dose math course in 2010-11 to answer my second research question. Traditional difference-in-differences designs often utilize a start or stop in treatment to create a natural experiment in which the outcomes of groups who are able and unable to receive the treatment are compared (Card & Kruger, 1994; Angrist & Pischke, 2008). However, in this situation I have three main differences I can use to isolate the intent-to-treat effects of offering the intervention for the eligible district population within a difference-in-differences-in-differences (DDD) design (e.g. Baker, 2016). First, a time difference indicates the year immediately before and after the certain schools stopped offering the intervention. A second difference is whether a student is enrolled in a school that maintained or stopped offering the courses. The third difference is whether the student fell within the group of students recommended to enroll in the course or whether they did not. The estimated treatment effect is the difference between the test score change differences for eligible and non-eligible students enrolled in schools that stopped or continued offering the course.

This design relies on a main identifying assumption that the control groups for which the differences are based (schools that stopped offering the intervention, students not eligible for the intervention) provide a suitable estimate of the counterfactual condition for treated students within schools that kept offering the class. Indeed, double dose math enrollment at schools that stopped offering the course immediately drops to 0%, and since the students are seventh-graders freshly entering Junior High School there is no possibility they encountered treatment before the schools stopped. Exposure to treated eighth-grade peers is possible but minimally concerning. Contrary to the traditional difference-in-difference (DID) design, the DDD design allows for a relaxation of the parallel trends assumption as I am specifically identifying the difference between outcomes for eligible and non-eligible students compared with their peers who attend non-participating schools (see Garlick, 2016).

I compare the difference between change in test score, traditional math course grade, and traditional math course pass rate, for eligible and non-eligible students at schools that kept offering the intervention versus students at schools that stopped. The full model is:

$$Y_i = \beta_0 + \beta_1 A_i + \beta_2 Z_i + \beta_3 T_i + \beta_4 A * T_i + \beta_5 Z * T_i + \beta_6 Z * A_i + \beta_7 Z * T * A_i + \delta' \chi_i + \varepsilon_i \quad (4)$$

in which A_i indicates being enrolled in a school that did not stop treatment, Z_i indicates being eligible to receive the treatment based on the assignment strategy, T_i indicates the time point at which treatment conditions change, and β_7 is the treatment effect indicating the difference in difference-in-difference estimates for eligible and non-eligible students at schools that maintained the treatment in comparison to schools that stopped the treatment. $\delta' \chi_i$ includes the following controls added for precision: prior achievement and demographic controls (gender, race, socioeconomic status, language status, special education).

The difference-in-differences-in-difference design is limited in that it provides a big picture idea of the effects of implementing the program but does not estimate effects for students who were specifically treated. Use of the design acknowledges that students who were recommended to enroll did not always do so, and vice versa, that some students enrolled without recommendation.

RESULTS

Table 1.3 presents the findings for my first research question, in which I used a regression discontinuity design to create a natural experiment that compares the test scores, grades, and math course pass rates for students just above and below the cutoff for enrollment in a double dose math course. The results indicate that, within the optimal 44-point bandwidth around the assignment cutoff, this extra course has no significant impact on students' end-year test scores, grades in their main math course, or likelihood of passing their main math course. The estimates and local average treatment effects for each outcome are negative but none are statistically significant. Table 1.3 also presents the results when using a bandwidth that is half and then double the size of my optimal bandwidth for comparison. When the bandwidth is decreased, all estimates remain similar except for a decrease in the coefficient for main math class grade. This column helps assuage concerns about the one demographic difference between the treatment and comparison groups, as even when the sample is small enough to eliminate that significant difference the estimates are very similar. The third column presents the estimates for a doubled bandwidth that is far too inclusive, but again the estimates remain almost identical. The program had a null effect that is consistent across outcomes and bandwidth specifications.

Figure 1.6 displays the identification strategy and raw test score results for my second research question, asking whether the treatment leads to detectable intent to treat (ITT) effects at the policy level for any of the primary outcomes. Table 1.4 displays the estimates from the final models including controls for precision. The coefficient of interest is the first row; this is the estimate of the difference in the differences between eligible and non-eligible students at schools that did and did not keep offering the intervention course. Countering the regression discontinuity estimates, these are consistently positive. However, the only statistically significant positive relationship is a 13.28-point increase for seventh-grade standardized test score ($p<.05$). Although there does seem to be positive differences for grades and pass rate in the main math class for eligible students at schools offering the treatment, these differences are not very large and not statistically significant.

Follow-up analyses using OLS Regression Models

Unfortunately, due to sample size restrictions I could not use regression discontinuity models to analyze differential effects of the intervention for students at different points of the prior achievement spectrum or English learners. However, I ran OLS regression models to test whether treatment in the intervention is associated with changes in standardized math test scores, main math course grades, or main math course pass rates for treated students within these two groups. In Table 1.5 I display the results of interaction models in which the sample was limited to students eligible for the intervention and treatment was interacted with a student's placement score (ranging from 3 to 8, with 3 as the lowest group). Being in the second-to-lowest assignment grouping (4 points) and enrolling in the course was related to a positive, statistically significant increase in test scores (+11.99 points, $p<.05$) compared with students at the top of the distribution (8 points) that took the course. Similar to the findings from Cortes, Goodman, and Nomi (2015) that low-performing readers benefitted more from the double dose class, Table 1.6 shows a positive relationship between math grades and treatment for students designated as English Learners (+.19 GPA points, $p<.05$).

DISCUSSION

One of the most striking conclusions from this analysis is descriptive in nature. Figure 1.2 highlights the variation in enrollment for students with similar scores on the placement mechanism and the low overall adherence to enrolling students in the intervention. This is a clear case of the strong variation in policy implementation that can take place at schools that are administering a district policy. School principals were adapting a curricular policy that was intended to be rigid to their individual settings that presented obstacles and required flexibility. The result is a sporadic implementation of a policy that was based on a clear assignment strategy.

It is also clear that even while dealing with the usual difficulties of adapting district policies to school settings, this district also took a minimally-structured approach to the design and accountability decisions for these classes. Without set curriculum in place, and with so much autonomy given to teachers on how this intervention was carried out in the classroom, it's possible that this loose structure could have cut into the positive effects usually seen from implementing these courses. I could not conduct any proper tests of this hypothesis, as data on how individual teachers used classroom time was not available, but future studies that could measure the relationship between implementation structure and intervention effectiveness would be an asset to this literature.

Despite the implementation differences from previously studied versions of this intervention, this particular course still appeared beneficial for some of the students who need math remediation the most. Although the results from my first research question indicate that the policy was not helpful for students near the assignment cutoff, implementing the policy still appears to have helped students overall. The results from my interaction models are correlational, but they suggest that even this loosely structured math instruction time can still improve test scores for students who are near, but not at the bottom, of the math achievement distribution in seventh grade.

This also explains why the results from my two quasi-experimental approaches might be conflicted. Within the tight bandwidth of the students who barely missed or made the cutoff to enroll, these courses were not impactful. However, when expanding the sample to students across the prior achievement spectrum, larger gains by lower performing students buoy the overall treatment effect. As for English Learners, the results are congruent with previous analyses (Cortes, Goodman, & Nomi, 2015) suggesting the intervention offers students with low language skills more time to navigate the language barrier while studying math so they can engage deeper and perform better.

Overall, this study provides a cautionary tale for school districts looking to implement this policy to improve students' math performance in middle school because of positive findings from previous studies. It appears that setting and policy implementation decisions might significantly alter the effectiveness of the intervention and should be considered more carefully. Overall, even a loosely-structured course can be helpful on average and might still be helpful for very low math performers and English Learners, but those benefits may not carry over to all students.

STUDY LIMITATIONS

The possibilities for modeling decisions, especially those establishing causality, were limited based on the data available and the sample size. In the regression discontinuity and difference-in-difference designs, statistical power is a concern. Although both analyses are suitably powered to detect moderate effects, some of the smaller nuances might be masked. It was not possible to use a quasi-experimental approach to analyze student subpopulations, so all OLS models highlighting these groups are only correlational. Testing of hypotheses regarding implementation decisions, course structure, or teaching strategies were not possible in this study. Finally, these analyses are focused on only one school district, so they are not generalizable outside of that population.

REFERENCES

Allensworth, E. M., & Easton, J. Q. (2007). *What matters for staying on-track and graduating in Chicago public high schools.* Chicago, IL: Consortium on Chicago School Research.

Angrist, J. D., & Pischke, J. S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Baker, R. (2016). The effects of structured transfer pathways in community colleges. *Educational Evaluation and Policy Analysis*, *38*(4), 626-646.

Bartik, T., & Lachowska, M. (2014). *The effects of doubling instruction efforts on middle school students' achievement: Evidence from a multiyear regression-discontinuity design* (No. 14-205). Kalamazoo, MI: WE Upjohn Institute for Employment Research.

Bottoms, G. (2008). *Redesigning the ninth-grade experience: Reduce failure, improve achievement and increase high school graduation rates.* Atlanta, GA: Southern Regional Educational Board.

Bryk, A. S., Sebring, P. B., Allensworth, E., Easton, J. Q., & Luppescu, S. (2010). *Organizing schools for improvement: Lessons from Chicago*. Chicago, IL: University of Chicago Press.

Calonico, S., Cattaneo, M. D., & Titiunik, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, *82*(6), 2295-2326.

Calonico, S., Cattaneo, M. D., Farrell, M. H., & Titiunik, R. (2016). rdrobust: Software for regression discontinuity designs. *Unpublished manuscript available at: http://faculty.chicagobooth.edu/max.farrell/research/Calonico-Cattaneo-Farrell-Titiunik_2016_Stata. pdf*.

Card, D., & Krueger, A. (1994). Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania. *American Economic Review*. 84 (4): 772–793.

Cortes, K. E., Goodman, J. S., & Nomi, T. (2015). Intensive math instruction and educational attainment: Long-run impacts of double-dose Algebra. *Journal of Human Resources*, *50*(1), 108-158.

Durwood, C., Krone, E., & Mazzeo, C. (2010). *Are two Algebra classes better than one? The effects of double-dose instruction in Chicago.* Chicago, IL: Consortium on Chicago School Research.

Fryer, R. G. (2012). *Injecting successful charter school strategies into traditional public schools: Early results from an experiment in Houston*. Cambridge, MA: National Bureau of Economic Research.

Garlick, R. J. (2016). Academic Peer Effects with Different Group Assignment Policies: Residential Tracking versus Random Assignment. *Economic Research Initiatives at Duke (ERID)*.

Henry, G. T., Barrett, N., & Marder, C. (2016). *Double-dosing" in math in North Carolina public schools* (REL 2016–140). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southeast. Retrieved from http://ies.ed.gov/ncee/edlabs.

Imbens, G. W., & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, *142*(2), 615-635.

Jackson, C. K., Johnson, R., & Persico, C. (2014). *The effect of school finance reforms on the distribution of spending, academic achievement, and adult outcomes* (No. w20118). Cambridge, MA: National Bureau of Economic Research.

Lee, D. (2008). Randomized experiments from non–random selection in US house elections. *Journal of Econometrics*, 142, 675–697.

Lee, D. S., & Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature*, *48*(2), 281-355.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, *142*(2), 698-714.

Meyer, J. W., Rowan, B., & Meyer, M. W. (1978). *The structure of educational organizations.*

National Center for Education Statistics. (2015). *The Nation's Report Card: Mathematics 2015.* NCES 2015-136. Washington, DC: Institute for Education Sciences, U.S. Department of Education.

Nomi, T., & Allensworth, E. (2009). "Double-Dose" Algebra as an alternative strategy to remediation: Effects on students' academic outcomes. *Journal of Research on Educational Effectiveness*, *2*(2), 111-148.

Nomi, T., & Allensworth, E. M. (2013). Sorting and supporting: Why double-dose algebra led to better test scores but more course failures. *American Educational Research Journal*, *50*(4), 756-788.

Nomi, T., & Raudenbush, S. W. (2016). Making a success of "Algebra for All": The impact of extended instructional time and classroom peer skill in Chicago. *Educational Evaluation and Policy Analysis*, *38*(2), 431-451.

Reardon, S. F., & Owens, A. (2014). 60 Years after Brown: Trends and consequences of school segregation. *Annual Review of Sociology*, *40*, 199-218.

Reed, D. S. (2014). *Building the federal schoolhouse: Localism and the American education state.* Oxford: Oxford University Press.

Rosin, M. S., Barondess, H., & Leichty, J. (2009). *Algebra policy in California: Great expectations and serious challenges*. Oakland, CA: EdSource.

Taylor, E. (2014). Spending more of the school day in math class: Evidence from a regression discontinuity in middle school. *Journal of Public Economics*, *117*, 162-181.

CHAPTER 1 TABLES & FIGURES

Table 1.1

Summary of empirical studies analyzing the effects of double dose courses

Analysis	Setting & population	Subject of double dose course(s)	Key treatment heterogeneity	Key findings	Setting limitations
Nomi & Allensworth (2009)	Chicago Public Schools, five ninth grade cohorts between 2000-2005	Math	• Professional development and curricular resources provided for double dose teachers • Same teacher and peers in regular and support course as much as possible • Courses offered sequentially in school day as much as possible	Short-term test scores improved amongst students who were targeted and non-targeted by the policy, but Algebra course failure rates increased for both.	Consistency of curriculum across schools not verifiable, and implementation guidelines were followed less strictly in the second of the two policy years.
Nomi & Allensworth (2013)	Chicago Public Schools, two ninth grade cohorts between 2002-2005	Math	Same as in Nomi & Allensworth (2009)	Changes in the classroom ability of the traditional math class after sorting between treatment and non-treatment students were impactful on outcomes for non-targeted students but not targeted students.	Same as in Nomi & Allensworth (2009)
Cortes, Goodman, & Nomi (2015)	Chicago Public Schools, two ninth grade cohorts between 2003-2005	Math	Same as in Nomi & Allensworth (2009)	Double dose course enrollment improved short-term outcomes (Algebra grades and pass rates) and long-term outcomes (credits earned, pass rates and test scores in later math classes, ACT math scores, four-year graduation rates, and college enrollment) for targeted students, especially poor readers. Schools' level of compliance to district's recommended implementation guidelines was not crucial.	Same as in Nomi & Allensworth (2009)

Nomi & Raudenbusch (2016)	Chicago Public Schools, all ninth-grade students from sixty high schools in 2003-04	Math	Same as in Nomi & Allensworth (2009)	The degree of ability level sorting in students' traditional class placement is crucial, and it is recommended that schools do not utilize any kind of sorting by ability level in students' traditional math placements.	Same as in Nomi & Allensworth (2009)
Taylor (2014)	Miami-Dade County Public Schools, students attending 6th through 8th grade between 2003-2013	Math	• Smaller class sizes • Most students (90%) had different teacher and peers in traditional and support course • Math courses were traditionally skills-heterogeneous	Double dose enrollment resulted in immediate positive effects on math test scores (.16 to .18 SD), but gains shrank considerably over the next two years. No connections to long-term outcomes were found, and double math courses were not significantly detrimental to outcomes in other school subjects.	Amount of professional development and curricular resources for double dose teachers is unknown.
Bartik and Lachowska (2014)	6th graders in a Midwestern, mid-sized district between 2010-2013	Math and English	• Students were identified for intervention in English and Math, and placed in double English blocks if they needed intervention in both	The authors found a significant increase in reading test scores for those taking double reading courses (+.20 SD) but no significant change for those in double math class.	No information was given on ability level sorting in students' traditional course placements or amount of professional development and/or resources given to teachers.
Fryer (2012)	Grade 6-12 students in nine public charter schools in Houston	Math and English	• Public charter school setting • Double blocked math and English courses were one of five interventions taking place simultaneously in these schools • Students were placed in double math blocks if they needed intervention in both	Gains were identified in most academic and social emotional categories, but these gains cannot be attributed solely to double classes. Between grades 7-12, a second math class in 8th grade was identified as the lone, significant predictor of math achievement gains when comparing targeted and non-targeted students	No information was given on ability level sorting in students' traditional course placements or amount of professional development and/or resources given to teachers.

Table 1.2

Summary statistics of analysis sample for participating schools (n=5946)

	2008-09	**2009-10**	**2010-11**	**2011-12**	**2012-13**	**Total Sample**
Observations	1242	1426	1532	1086	660	5946
Number of participating middle schools	6	7	7	5	3	-
Student demographics						
% Female	50.4	48.5	49.6	48.3	52.6	49.6
% African American	0.7	0.4	0.6	0.8	0.8	0.6
% Asian	30.7	28.5	22.8	25.0	21.8	26.1
% Hispanic or Latino	61.3	54.8	69.6	66.9	71.2	64.0
% White	6.2	15.1	5.4	6.3	5.6	8.1
% Free- and Reduced-Price Lunch	72.4	67.5	82.3	81.4	79.4	76.2
% English Language Learners	35.6	25.5	40.9	31.2	28.3	32.9
% Special Education	3.5	2.7	2.7	3.6	2.4	3.0
Prior math achievement						
6^{th} grade Mathematics state standardized test score	352.7 (50.8)	367.8 (58.4)	363.3 (53.9)	372.3 (59.1)	366.0 (53.1)	364.1 (55.6)
Double dose math course enrollment						
% Enrolled in double dose course during 7^{th} grade	27.5	25.9	23.4	25.1	21.1	24.9
% Eligible for double dose that enrolled	55.1	58.4	59.6	59.5	49.7	57.2

Note. Students were included in the sample if their school enrolled more than 33% of eligible students in a double dose math course. Schools that did not implement the course, or did not offer the course widely, were excluded.

Table 1.3

Fuzzy regression discontinuity estimates of the treatment effect of double dose math classes for seventh grade students (n=4003)

	44 point bandwidth (optimal)	22 point bandwidth	88 point bandwidth
7th grade math test score			
Difference for treatment students	-3.64	-3.92	-3.82
	(3.94)	(5.68)	(2.93)
Localized average treatment effect	-24.0	-24.2	-27.7
	(25.99)	(34.78)	(21.62)
7th grade main math class grade (GPA points)			
Difference for treatment students	-.17	-.04	-.16
	(.13)	(.19)	(.10)
Localized average treatment effect	-1.12	-.23	-1.12
	(.90)	(1.15)	(.73)
7th grade main math class pass rate			
Difference for treatment students	-.05	-.03	-.05
	(.04)	(.06)	(.03)
Localized average treatment effect	-.32	-.19	-.34
	(.29)	(.38)	(.24)
Average, weighted discontinuity across bandwidth	15.2%	16.2%	13.8%
F statistic of eligibility instrument	35.9	20.1	149.3
Number of students included within bandwidth	805	458	1610

Note. Standard errors in parentheses.

Table 1.4

Difference-in-difference-in-difference (DDD) estimates of the intent-to-treatment effect of double dose math classes for seventh grade students (n=3745)

	7th grade math test score		7th grade main math class grade (GPA points)		7th grade main math class pass rate	
School offering x Eligibility x	10.50	13.28^{*}	0.11	0.10	0.09	0.08
Time (main DDD estimate)	(8.33)	(6.46)	(0.19)	(0.18)	(0.06)	(0.06)
Time (2011 year)	-11.85^{***}	-9.13^{***}	-0.23^{**}	-0.21^{**}	-0.05^{*}	-0.05^{*}
	(3.54)	(2.74)	(0.08)	(0.08)	(0.02)	(0.02)
School offering double dose	-1.16	4.53	0.07	0.14^{*}	-0.03	-0.02
math class	(3.14)	(2.49)	(0.07)	(0.07)	(0.02)	(0.02)
Eligibility for double dose	-75.84^{***}	-20.73^{***}	-1.24^{***}	-0.65^{***}	-0.19^{***}	-0.11^{***}
math class	(4.48)	(3.66)	(0.10)	(0.10)	(0.03)	(0.03)
School offering x Eligibility	-9.23^{+}	2.01	-0.04	0.19^{+}	-0.02	0.02
	(5.28)	(4.13)	(0.12)	(0.11)	(0.04)	(0.04)
School offering x Time	-10.02^{*}	-5.60	-0.06	0.02	0.01	0.02
	(4.32)	(3.35)	(0.10)	(0.09)	(0.03)	(0.03)
Eligibility x Time	3.27	-4.11	-0.02	-0.07	-0.07	-0.07
	(7.40)	(5.75)	(0.17)	(0.16)	(0.05)	(0.05)
Prior math assessment score		X		X		X
Student demographics		X		X		X

Note. Standard errors in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 1.5

OLS regression estimates of the relationship between outcomes and treatment for all eligible students and students at different points along the prior achievement distribution

	7th grade math test score		7th grade main math class grade (GPA points)		Pass 7th grade main math class (logistic)	
Double dose math course	5.55*** (1.35)	-2.91 (2.86)	-0.00 (0.05)	-0.05 (0.09)	-0.07 (0.09)	-0.13 (0.19)
Assignment strategy: 3 points		-28.77** (10.10)		-0.73* (0.33)		-1.26* (0.58)
Assignment strategy: 4 points		-38.65*** (3.99)		-0.44*** (0.13)		-0.82*** (0.24)
Assignment strategy: 5 points		-17.90*** (3.30)		-0.24* (0.11)		-0.34 (0.21)
Assignment strategy: 6 points		-10.73*** (2.77)		-0.14 (0.09)		-0.28 (0.18)
Assignment strategy: 7 points		-8.89*** (2.45)		-0.14 (0.08)		-0.25 (0.16)
Assignment strategy: 8 points (reference)		-		-		-
Double dose math course x 3 points		-10.68 (11.24)		0.16 (0.36)		0.25 (0.65)
Double dose math course x 4 points		11.99* (5.30)		-0.12 (0.17)		0.06 (0.32)
Double dose math course x 5 points		-0.60 (4.61)		-0.14 (0.15)		-0.20 (0.29)
Double dose math course x 6 points		-0.80 (4.16)		-0.04 (0.13)		-0.10 (0.27)
Double dose math course x 7 points		3.09 (3.98)		-0.06 (0.13)		-0.07 (0.26)
Double dose math course x 8 points (reference)		-		-		-
Prior math assessment score	X		X		X	
Student demographics	X	X	X	X	X	X
Observations	2818	2818	2818	2818	2818	2818

Note. Standard errors in parentheses. Controlled student demographics include gender, race, free/reduced lunch status, English language status, and special education / disability status.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 1.6

OLS Regression estimates of the impact of treatment for English Learner students eligible to enroll

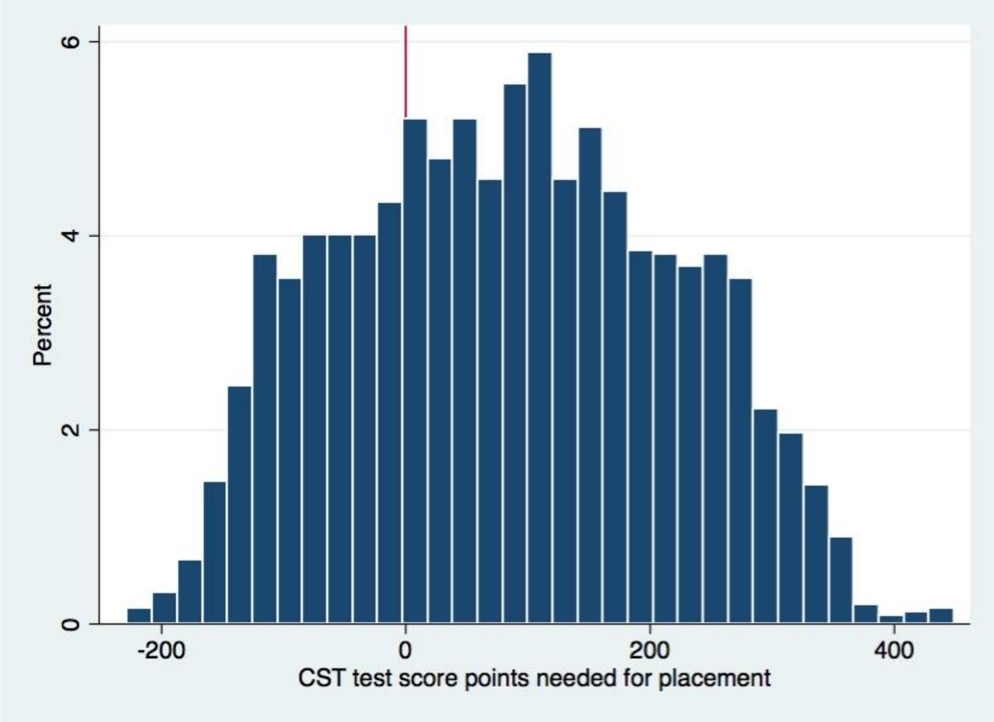
	7^{th} grade math test score	7^{th} grade main math class grade (GPA points)	Pass 7^{th} grade main math class (logistic)
Double dose math course x English Learner	-2.33	0.19^{*}	0.05
	(2.55)	(0.09)	(0.03)
Double dose math course	6.83^{***}	-0.11	-0.04
	(1.95)	(0.07)	(0.03)
English Learner	-11.12^{***}	-0.27^{***}	-0.07^{**}
	(1.90)	(0.06)	(0.03)
Prior math assessment score	X	X	X
Student demographics	X	X	X
Observations	2818	2818	2818

Note. Standard errors in parentheses. Controlled student demographics include gender, race, free/reduced lunch status, and special education / disability status.

$p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Figure 1.1

Zero-centered test score needed for placement decision for double dose math class (n=4003)



Note. Placement decision is based on three test scores. Students whose third test score can still decide their placement are included. Students who score less than zero were recommended to enroll, those who scored 0 are above are not.

Figure 1.2

Percent of students enrolled in a double dose math class based on the running test score variable (n=4003)

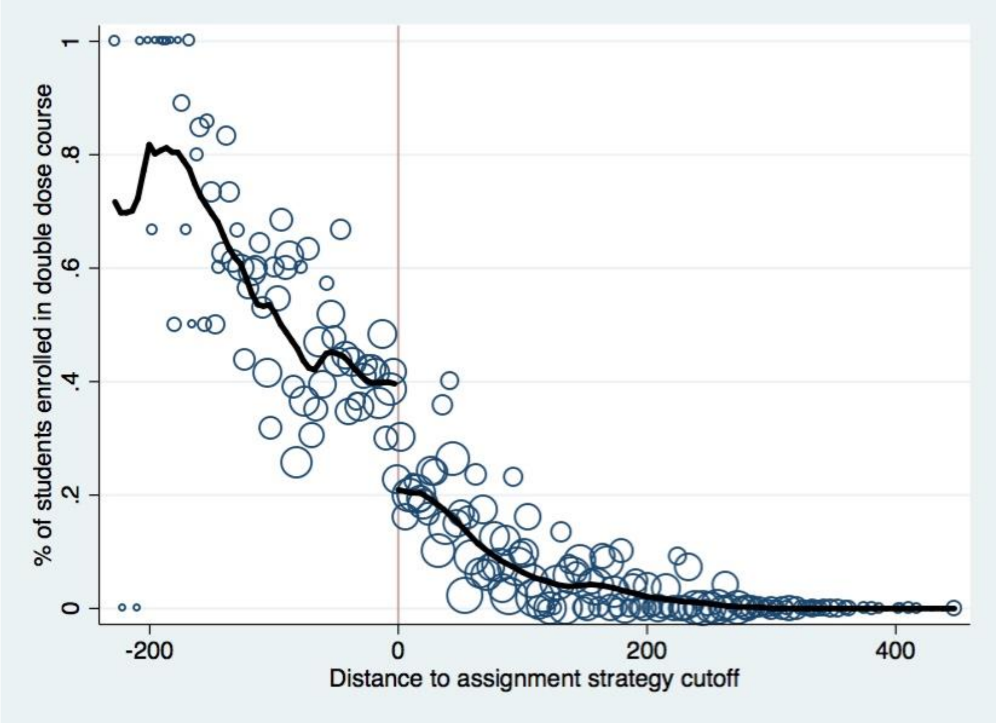


Figure 1.3

Percent of students enrolled in a double dose math class based on the original running test score variable

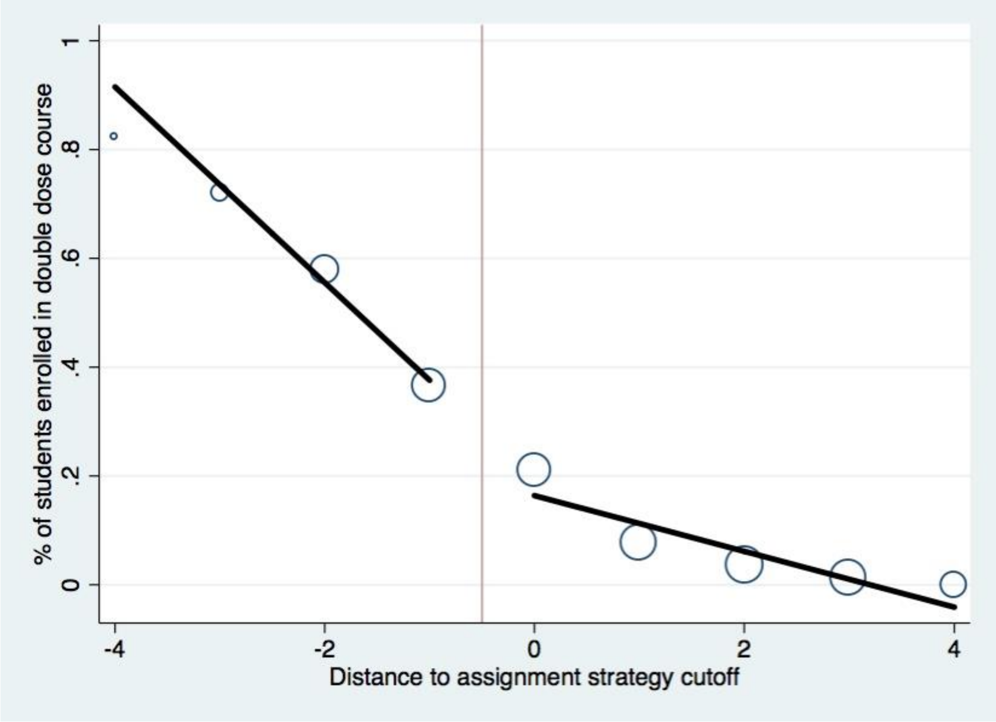


Figure 1.4

7th grade math test score outcome for students within 44 points above and below the regression discontinuity cutoff (n=4003)

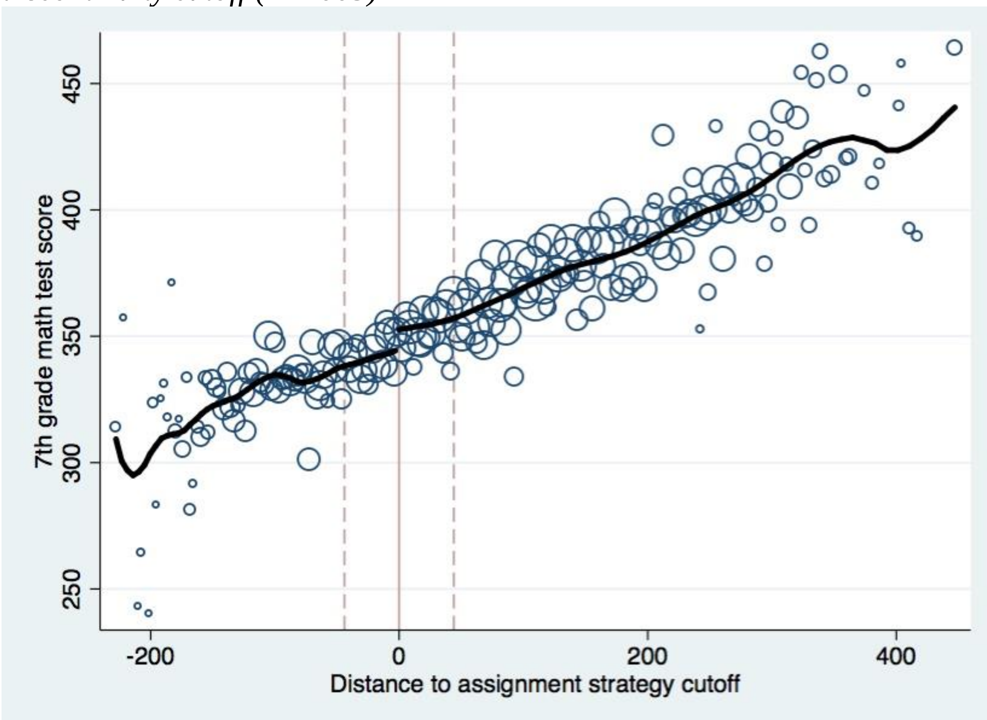


Figure 1.5
7th grade main math class grade outcome for students within 44 points above and below the regression discontinuity cutoff (n=4003)

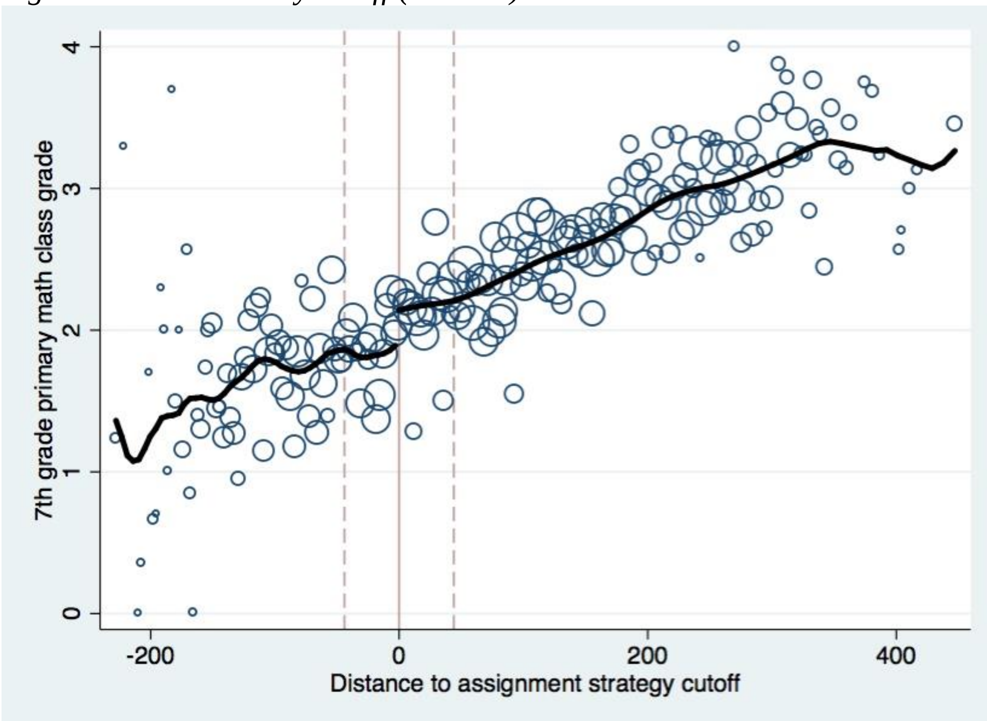


Figure 1.6
7th grade main math class pass rate outcome for students within 44 points above and below the regression discontinuity cutoff (n=4003)

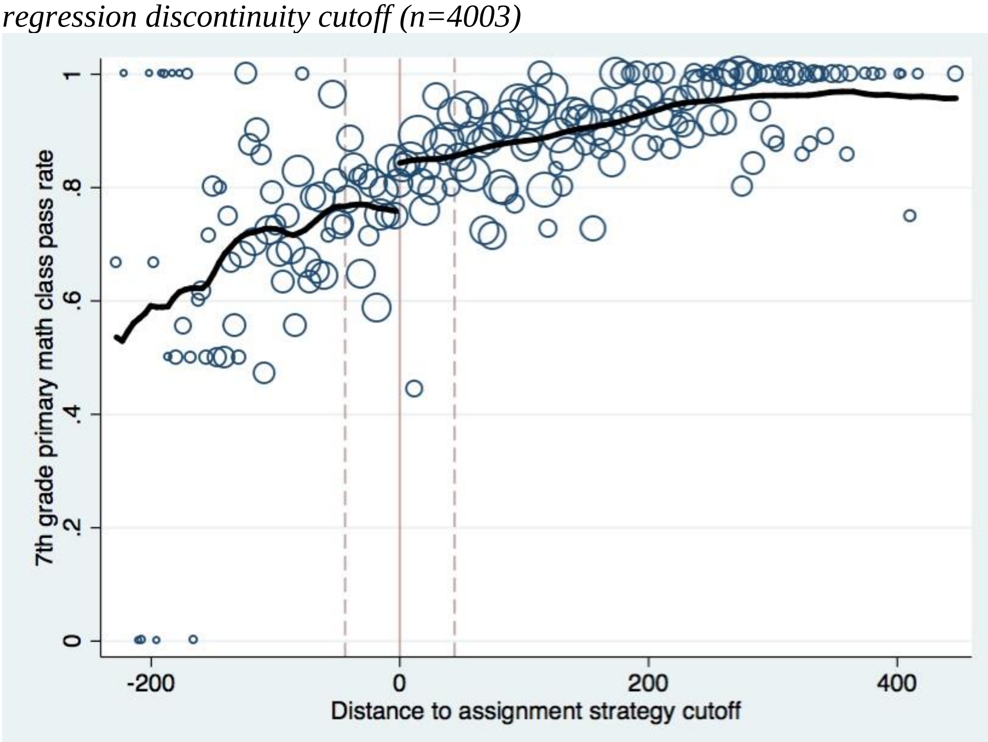
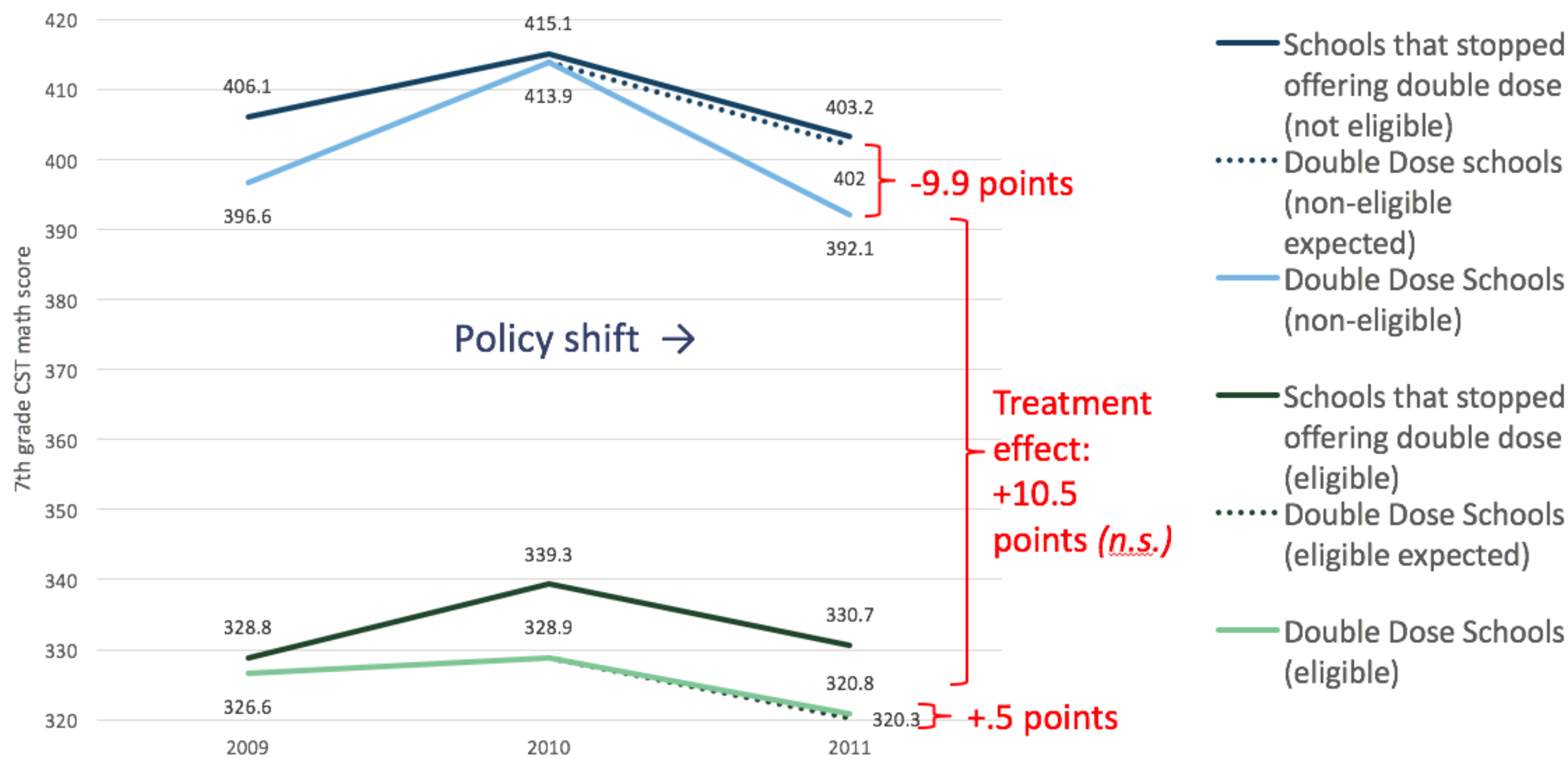


Figure 1.7

Raw test score change in 7th grade math test scores for schools that did and did not continue the intervention in 2010-11 and students who were or were not eligible to enroll (n=3745)



Appendix Table 1.1

One-tailed t-tests for groups within 44 points above and below the regression discontinuity cutoff

	44 points below	44 points above	Statistically different?
Observations	377	428	
Student demographics			
% Female	48.0	51.4	No
% African American	0.5	0.2	No
% Asian	18.6	21.0	No
% Hispanic or Latino	71.9	66.6	No
% White	8.2	10.0	No
% Free- and Reduced-Price Lunch	79.8	78.9	No
% English Language Learners	46.7	36.9	Yes, $p<.01$
% Special Education	3.4	4.0	No

Appendix Table 1.2

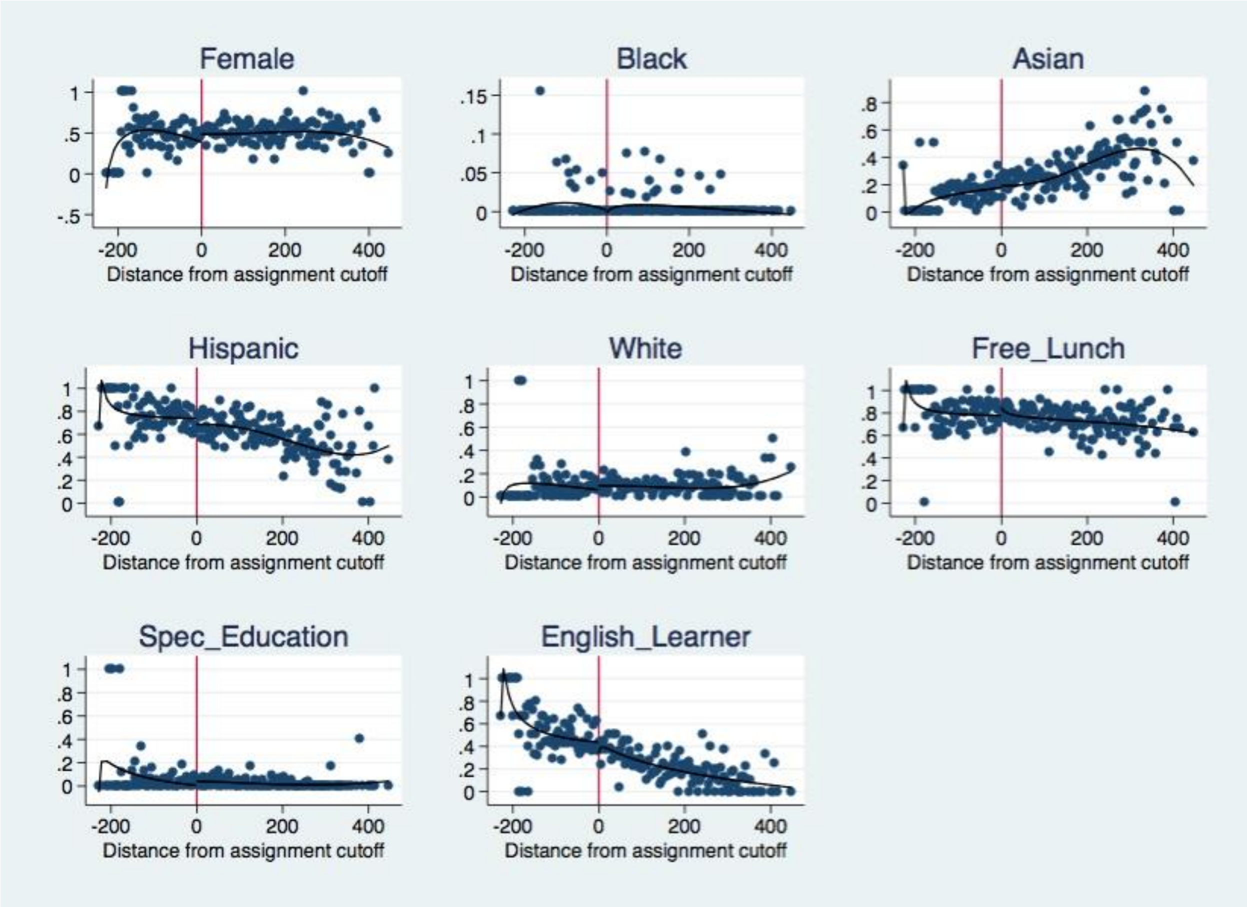
Regression check of parametric form of regression discontinuity (n=4003)

	7^{th} grade math test score
Treatment (instrumented by eligibility)	-5.80
	(7.97)
Running variable	0.22^{***}
	(0.06)
Running variable squared	0.00
	(0.00)
Running variable cubed	0.00
	(0.00)
Treatment * Running variable	-0.18
	(0.18)
Treatment * Running variable squared	-0.00
	(0.00)
Constant	348.04^{***}
	(2.79)

Note. Standard errors in parentheses.
*** $p < 0.001$

Appendix Figure 1.1

Discontinuity in demographic characteristics based on the running test score variable (n=4003)



CHAPTER 2

Doubling Down: Sensemaking During a School District's Decision to Re-Introduce a Previously Ineffective Intervention.

INTRODUCTION

One of the key ways in which school districts can impact instruction is through district-wide policy implementation. As schools are serving increasingly diverse student populations and operating in unique funding and policy climates (Jackson, Johnson, & Persico 2014; Meyer, Rowan & Meyer 1978; Reardon & Owens, 2014; Reed, 2014), district-wide policy adoptions can result from reactions to a wide range of factors including external pressures such as state or federal regulations, internal pressures from parents, employees, or student needs, or recommendations from new research (Bryk, Sebring, Allensworth, Luppescu, & Easton, 2010; Spillane, 1996). School districts are more commonly being asked to interpret and respond to pressures that impact their school sites, and this makes efforts to document the decision-making processes of school district leadership more essential. To date, there are very few case studies that offer this type of insight, especially in regard to policies that shape curricula.

This particular case study documents a unique type of decision-making within a fairly large, suburban Southern California school district serving an economically- and racially-diverse student population. In 2008-09, the district implemented "double dose" math courses, a second math course during the school day, within their middle schools as an intervention to support low-performing math students. This intervention was available to seventh- and eighth-grade students as a support for Pre-Algebra, if the student was considered to be at risk to fail Algebra the following year, or for Algebra, if the student was enrolled in Algebra but likely needed extra support to pass the course. In 2010-11, two schools decided to discontinue offering the courses and by 2013-14 these courses were almost entirely discontinued by the school district because they were determined to be ineffective.

However, by 2015-16 the district piloted a modified version of the same intervention and mandated the offering across the entire school district in 2016-17. Double dose math course support was made available to seventh- and eighth-grade students who scored below grade-level proficiency the previous academic year and were not enrolled in a Special Education or Honors level math course. In this case, not only did the school district shape its schools' curricula with a universal policy, but the decision makers returned to a previously unsuccessful policy within a short time period. To date, the literature has predominately focused on instances of "policy churn," in which policies are not fully implemented or given a chance to develop and then replaced with a different policy that suffers that same fate (Hess, 2011). This case provides an alternative view of a district that reflected upon, altered, and then reinstated the same curricular policy instead.

In this study I focus on the reasons for implementation of this policy and the changes that were made between the intervention's original offering and its reintroduction in 2015-16. I will document the decision-making processes of the school district administrators within this time period using research on organizational sensemaking (Vaughan, 1996; Weick, 1995) and collective sensemaking in educational settings (Coburn, 2001) as a theoretical framework. This case study contributes to the literature clarifying the growing role of school districts as designers of schools' instruction in the wake of increased state and national policy demands (Spillane, 1996; Spillane, 2009). The district's original version of the policy was a reaction to a California state "Algebra for All" initiative in which schools were recommended to enroll all students in an Algebra math course by eighth grade (Rosin, Barondess, & Leichty, 2009; Domina, McEachin, Penner, & Penner, 2015) and the district's return to the policy was in conjunction with California's adoption of the Common Core State Standards.

Literature Review

Although the curricular policy decisions of school districts are not discussed nearly as often as those of individual schools, school districts have a growing influence over the day-to-day classroom instruction of their sites. School districts are important, unique actors in education policy because they often interpret state and national policies. How they interpret these guidelines can influence how their educators follow these curricular reforms through direct channels such as corresponding district policies or indirect channels such as curricular resources and professional development (Spillane, 1996). Spillane (2009) specifically notes that in the wake of more detailed data collection, greater accountability pressures, and often tighter financial limitations, school districts have played an increasingly larger role in determining the curricula of their school sites over the last three decades.

Diverse student populations and demanding funding and policy climates may affect district-wide policies (Jackson, Johnson, & Persico 2014; Meyer, Rowan & Meyer 1978; Reardon & Owens, 2014; Reed, 2014). Individual district policy adoptions may also result from reactions to a wide range of other factors including internal pressures from parents, employees, or student needs, or recommendations from new research (Bryk, Sebring, Allensworth, Luppescu, & Easton, 2010; Spillane, 1996). School district administrators navigate complex environments, and although this growing influence they hold over classroom instruction is becoming more pronounced, we know very little about how curricular policy is shaped and disseminated from school district offices. Who are the primary decision makers at district offices? Why are particular decisions made by district leaders? What informs the specific curricular policies that are put into place? What evidence is used to make these types of decisions? How are policies considered for elimination or renewal?

To understand how teachers, integrate new policies into their instruction, sociologists developed the cognitive approach to policy implementation in school settings (summarized in Coburn, 2005). Derived from theories of organizational sensemaking (Vaughan, 1996; Weick, 1995), it is argued that the way in which teachers implement policy is influenced heavily by prior knowledge and understandings, the social context in which they work, and their connection to or belief in the value of the policy or reform (Coburn, 2001; Coburn, 2004; Vaughan, 1996). The practice of sensemaking involves considering these influences to interpret new instructional approaches, which often leads to reconstructing policy messages in order to reinforce previous understandings or make incremental rather than large changes (Coburn, 2001, 2005; Smith, 2000). Educators can undertake sensemaking activities individually or they can do so with colleagues. Specifically, Coburn (2001) describes the latter act as collective sensemaking in which multiple individuals within an organization interact with each other and their environment to negotiate and reach shared understandings of norms and routines involving a new policy.

Spillane, Diamond, Burch, Hallett, Jita, and Zotlners (2002) expanded this line of research when they observed similar sensemaking behaviors among principals as they interpreted and implemented policies that impact their schools. Like teachers, preexisting understandings and the social context of their environment were heavy influencers of principals' interpretations and adaptations of accountability policies. Coburn (2005) advanced this literature by connecting the sensemaking actions of principals to eventual changes in policy regarding classroom instruction in their schools. This study of elementary school principals adopting a new reading policy adds another dimension to previous work; principals not only engage in sensemaking activities of their own, but they also shape how their teachers are able to engage in their own sensemaking activities about a given policy. Principals control the teachers' access to policy ideas, guidelines for adopting the policy, and can alter conditions of the environment that a teacher must consider.

These studies define clear ideas about sensemaking activities amongst teachers and school leaders, but only a handful of studies have used collective sensemaking as a theoretical framework for understanding the policy decisions made by administrators in school district central offices. Spillane (2009) examined the instructional decision-making processes of administrators in a diverse set of Michigan school districts and noted some prominent patterns. In the district offices, administrators, staff members and representatives from school sites (principals and teachers) shaped instructional policy while school boards often provided only approval or disapproval. Recommendations from non-governmental agencies, such as nonprofit organizations, universities, and granting foundations, strongly influenced the details of policies adopted by district leaders. Additionally, the school districts used instruments, most commonly curricular materials and professional development for teachers, to implement and reinforce their policy decisions at school sites.

Coburn, Toure, and Yamashita (2009), drawing on instructional decision making in one school district over three years, suggested four factors determining sensemaking processes among district administrators. First, content knowledge and previous understandings were particularly influential in decisions on curriculum, professional development, and assessment policies. Second, the organizational structure of the central office, notably the complex and varied departments within this office, resulted in individualized sensemaking and decision making without collaboration. Third, the district was heavily responsive to resource constraints. Finally, leadership turnover at the top of the district hierarchy caused changes in the priorities of decision makers. Of these factors, the first three could be generalizable to understanding instructional decision making in other districts. The fourth factor, leadership turnover within the organization, was specific to the district studied and will not always be applicable to others.

Regarding the types of evidence consulted by district decision-makers, Coburn et. al (2009b) pointed out that resource constraints and other practical imperatives may lead to the use of less rigorous evidence than might be considered ideal by researchers. Coburn, Honig, and Stein (2009) divided evidence use into five categories: instrumental, conceptual, symbolic, sanctioning, and no role. Instrumental use is the most thorough consultation of evidence before making a decision, conceptual refers to looking to evidence for new ideas or inspiration, symbolic is looking only to evidence that reinforces preexisting understandings, sanctioning refers to incorporating evidence as way of knowing if a certain policy meets a certain requirement, and no role means that no rigorous evidence is consulted in the decision-making process. Of the categories, they found conceptual use of evidence to be the most commonly practiced.

Current Study

Although the existing literature on instructional decision making in school district offices is not robust, it provides a strong foundation for further investigating these processes. In this study, I plan to merge the findings of Coburn et al. (2009a), Coburn et al. (2009b), and Spillane (2009) into one qualitative coding framework used to analyze a case study of one midsized, suburban school district in Southern California. This endeavor will not only test the generalizability of these previous findings, but I will also use their conclusions to help interpret this district's unique decision to re-implement a previously unsuccessful intervention. Using sensemaking as a framework for understanding the organizational behavior during this case, my primary research questions for this study are:

1. How did district administrators engage in collective sensemaking to re-implement a double dose math course intervention?
2. How were decisions about the altering and re-implementation of "double dose" math classes made and who made them?
3. How and what did district administrators learn as they modified the policy?

METHODS

This is a case study of a single school district, but with the intent of revealing decision-making processes that apply more broadly. Hatch (2002) described the case study method as best used to study a "contextualized, contemporary phenomenon within specific boundaries" (p.30). Qualitative researchers prefer case studies when investigating processes that unfold over time (Merriam, 1998) or involve complex contextualization (Yin, 1994). This case study focuses on the instructional decision-making process surrounding one math intervention in one school district office over two academic years. Data collection took place over one academic year, but individuals recounted events that occurred during the preceding year.

During the years involved in this study, "Sunnyside School District" was one of the twenty-five largest districts in California. Within this district approximately 75% of students qualified for free- or reduced-price lunch, more than 50% were Hispanic/Latino, and about 33% were Asian or Asian American. Within these ethnicity groupings, the district enrolled large groups of first- and second-generation Mexican, Guatemalan, Salvadoran, Vietnamese, Hmong, Korean, and Chinese students

Data Collection & Participants

I collected data in coordination with the office of Research and Evaluation at Sunnyside School District during the 2016-17 academic year. Based on the capacity of this office, I designed a mix of focus groups, observations, and in-person interviews that was as intensive and far-reaching as the district would allow. In total, I conducted four focus groups, one professional development observation, classroom observations in six class periods with four different teachers, and two additional interviews with upper-level district management that were involved in the previous or current implementation of the courses. Table 2.1 provides full details on these activities including the participants, timing, and (where applicable) the interview protocol used.

Data collection began with a preliminary information meeting with upper-level district administrators to discuss the district's use of "double dose" math courses over the previous eight years. The participants included one teacher of the former version of the class who now works in the district's evaluation office, the director of the district's educational services for all grade levels, a former principal now serving as the director of the district's instructional technology for all grade levels, and the director of the team that designed the new version of the course. This conversation was informal and focused on the details of implementation of both versions of the course, as well as the evidence used to evaluate the course and the reasons for re-introducing the course. I was not able to record this conversation, but I kept detailed meeting notes for later reference.

I conducted a total of four focus group sessions. The first occurred in December 2016 and covered the previous implementation of the double dose math course. The conversation lasted approximately sixty minutes. I recruited four participants via convenience sampling with help from the district's research team. The participants included two teachers of the former class, one member of the team that designed the new version of the course, and the director of the team that designed the new version of the course. I used a pre-determined set of interview questions along with additional questions that came up in conversation (Appendix Document 2.1). The conversation with the focus group was recorded and transcribed in its entirety.

I conducted a second, hour-long focus group in February 2017 to discuss the current implementation of the double dose math course. Four participants, recruited via convenience sampling with help from the district's research team, attended the session. The participants included three lower-level district staff members and the director of the team that designed the new version of the course. I used a pre-determined set of interview questions along with additional questions that came up in conversation (Appendix Document 2.2). The conversation with this focus group was also recorded and transcribed in its entirety. I conducted a shorter, third focus group in June 2017 with the same group of participants. This conversation focused on lingering questions about the transition between the previous and current implementations of the double dose math course. The pre-determined set of interview questions for this session can be seen in Appendix Document 2.3.

A final focus group session was conducted in June 2017 with four teachers of the current version of the math course. This was also structured by a pre-determined interview protocol (Appendix Document 2.4). The participants were recruited via convenience sampling in conjunction with the district's research office. The conversation was sixty minutes in length, recorded, and then transcribed for coding. Following this I conducted classroom observations in six different class periods of the same four instructors as they taught the double dose math course. Each observation consisted of one full period of double dose math instruction. These observations were naturalistic in nature and I kept field notes detailing the activities that occurred and my thoughts as I observed the class. All field notes were then transcribed for coding.

I also attended a professional development session for teachers of the current version of the course in February 2017. In total, there were seventeen participants. Four participants were district staff members from the team that designed the new version of the course, ten were current double dose math course teachers, one was a student teacher, and two were district-employed Teachers on Special Assignment (TOSAs) that had been assigned to lead the professional development sessions the following year. The professional development session was not mandatory, but more than half of the current teachers of the course attended the session. This observation was naturalistic in nature and I kept field notes describing the activities that occurred and the opinions of the teachers involved. I used information from this session to confirm details provided during the focus groups and inform a list of teacher focus group questions.

Finally, in June 2017 I conducted a formal interview with the district's director of instructional technology to learn about upper-level management's views regarding the decision-making process. This conversation was sixty minutes in length and focused on the evidence used and reasoning behind decisions surrounding the course from their perspective. My questions followed the interview protocol presented in Appendix Document 2.5. All answers were recorded and then transcribed for coding. I was also able to conduct an email interview with a current principal of one of the district's middle schools. The principal was selected via convenience sampling in coordination with the school district office and all questions asked are listed in Appendix Document 2.6.

Coding Framework

In order to code my qualitative data for analysis, I created a coding framework based on the literature review for this study. This qualitative data coding framework is included as Table 2.2. As discussed in the Introduction, I based this coding framework on the primary findings from studies by Coburn et al. (2009a), Coburn et al. (2009b), and Spillane (2009). Structural codes are included for classroom, district office, and professional development session settings. Direct action codes are included for various coding regarding policy adaptation and decision making. Finally, twenty-five descriptive codes are included under the following subtitles: general sensemaking, general decision making, previous knowledge and understandings, organizational structure, resource constraints, and evidence use. These categories incorporate the findings from these prominent studies, as detailed in the introduction to this case study, and will provide categories for summarization of findings in the results section.

The purpose of this coding framework was to identify instances of sensemaking activities amongst actors within this setting, while also marking details about what type of sensemaking activities were taking place. Incorporating previous knowledge, adapting due to resource constraints or organizational structure, and using evidence to reach understandings and make decisions are all theorized components of the sensemaking process. By categorizing these activities, I was able to examine what types of behaviors are most common, most influential, and engaged in by whom.

Data Analysis

All recorded transcriptions and field notes were coded following the coding framework I created (Table 2.2). All data were coded using the qualitative data analysis software QDA Miner Lite. Because the data came from targeted focus groups, interviews, and observations, it was not as vast as data used in other qualitative studies. However, using the qualitative data analysis software still allowed for better organization of my codes and easier pattern detection within the designated categories of results. I read and manually coded each piece of collected data within the QDA Miner Lite software, then summarized the data within the categories of results I intended to report. Within the category summaries I looked for quantity of labels and patterns of labels coded simultaneously.

RESULTS

The findings from my analysis of all coded data will be split into two sections. In the first section, I will summarize pertinent data from the direct action codes (policy adaptation, decision making) and each of the categories of descriptive codes from my qualitative coding framework (general sensemaking, previous knowledge and understandings, organizational structure, obstacles and/or constraints, and evidence use). These summaries are designed to inform the reader of the quantities of assigned codes under these categories and provide examples of data that was coded within these categories. The second section will focus on answers to my proposed research questions. The study concludes with a discussion of the findings.

Coding Category Summaries

1) Direct action codes: policy adaptation and decision making

Within these data, I indicated a code under the two direct action codes for policy adaptation or decision making a total of 81 times. The most common code under policy adaptation was references to a policy adaptation at the district level (23 total). Under decision making, the most common coding was for a decision-making actor (12 times). Policy adaptations included minor changes to how the policy was implemented, such as counselors trying to group students with the same teacher for both their main and support math course, while decisions were categorized as specific choices that were made by an actor and broadcast out for others to follow. Depending on the situation, a policy adaptation could also qualify as a decision if it was enforced across multiple sites or classrooms (ex. such as a school's rule on admitting newly transferred students into the class automatically). District staff members (both upper- and lower-level administrators) and principals were the only individuals identified as decision-making actors, while those three groups plus teachers were identified as policy adaptors. Understandably, district administrators, principals, and teachers were the most likely to reference making policy adaptations at the district, school, and classroom levels, respectively.

2) General sensemaking

I indicated a code within a general sensemaking category a total of 111 times. The most common codes used were those indicating a reference to sensemaking activities among teachers, sensemaking activities among administrators, or sensemaking activities by an individual in a solo setting. I identified five types of sensemaking actors that were involved in the implementation of the policy. Upper-level district administrators (e.g. Superintendents, Assistant Superintendents, Directors of Instruction), lower-level district staff members (e.g., Curriculum Specialists), school-level Principals, school site academic counselors, and teachers.

General sensemaking activities occurred in seven different settings. Discussions about the adoption and implementation plans for the double dose math policy occurred within two of these settings: meetings amongst district administrators only, and meetings in which district staff members invited and solicited feedback from school-level personnel (namely school principals.) As it was described to me, many district-level policies, including the double dose math course policy, were items of discussion during the latter meetings and these meetings were designed to gather feedback from school principals and administrators on policy effectiveness. However, there was a clear delineation that decision-focused conversations about policy changes, adoption, and basic implementation details for these policies occurred amongst upper-level district administrators only.

Dialogues about how to specifically implement the policy within school sites, what adaptations should be made, and whether or not the policy was congruent with educator beliefs primarily took place in five other settings. A team of four district staff members underneath the upper-level administrators (all former teachers and school-level administrators) were tasked with implementing the new double dose math courses and this team worked for three years to design the course's curriculum, train teachers on the new materials, and support school sites with placement, data analysis, and policy modifications as needed. This team engaged in sensemaking activities around the policy when it met, and they also organized meetings and professional development sessions with their teachers at the district office. The final three settings where sensemaking activities occurred were in teacher's meetings with principals or counselors at their school site (either as a group or individually), when teachers within the same school informally met and discussed the policy, and finally, teachers engaging in reflection by themselves.

Teachers were the only actor that engaged in sensemaking about the policy in a solo setting; all other sensemaking activities took place in a group of individuals.

As introduced by Coburn (2005), there were also situations in which certain actors dictated other actors' access to or ability to engage in sensemaking activities about the policy. This was particularly prevalent in the lower-level district staff members' conversations with double dose math teachers and school principals. In this case, district staff members framed conversations in hopes of generating "buy-in" for the policy. As the district staff members directly responsible for the implementation and success of the policy at the site level, these actors were interested in keeping the policy easy to implement and viewed in a positive light. For example, one district staff member described doing additional work to provide unplanned support for the school sites on how to place incoming seventh-grade students. "[We] did so much work identifying incoming seventh graders… We provided every school with incoming seventh graders we felt met the criteria that they could place in a [double dose intervention] class. That helped a lot; that got more schools doing it." Another district staff member described a goal of professional development sessions about the course as "rebranding." The district team training the teachers often referred to the course as "Version 2.0" to signal that it was new and improved, with the goal of increasing acceptance of the return to the policy.

Finally, Coburn (2001) indicates that a common goal of sensemaking activities is to develop new, shared understandings amongst others that also interact with the policy. In the situations described to me, this primarily occurred in conversations between lower-level district staff members and school site employees (principals, counselors, and teachers) or amongst those school site employees at school sites. During the professional development session, I observed one example as district staff members and teachers engaged in sensemaking around how to best grade students within the intervention course. As the course does not have regular assignments or exams, the teachers expressed their ambiguity in how to assign grades for the course on any metric besides effort. District staff members offered their vision, that they thought the ideal grade would be based on effort and participation, and together the district staff and teachers agreed that no other assignments or assessments should be considered in the grading process. In another example, focus group teachers described having similar types of conversations with principals surrounding school placement decisions.

3) Referencing previous knowledge and understandings

Referencing previous knowledge and understandings is a specific sensemaking activity that refers to including prior preferences during the sensemaking process (Coburn, 2001; Coburn, 2005, Smith, 2000). Within the data, I indicated a total of 21 individual codes that fall within this category. The only code that was used was a reference to a previous understanding utilized in the sensemaking process. Indication of defying a previous understanding never occurred in these data. Personnel at all levels made reference to or acknowledged previous knowledge and understandings about a variety of topics as they made decisions and adaptations regarding this policy.

Starting from the highest level of authority, upper-level district management acknowledged that a familiarity with double courses as a previous intervention in math and English Language Arts was a main reason in deciding to bring them back: “It’s a model that the district has used for multiple things… and the fact that we already had an existing structure.” Lower-level district staffers, those in charge of designing and guiding school sites in implementing the course, often cited previous understandings to justify the curricular and placement decisions throughout the re-design of the policy. For example, in discussing a new decision to only recommend students to take the course if they demonstrated top marks in behavior from the previous year, one district staffer was interpreting that this type of course could only be beneficial for students who struggled in math but were otherwise motivated to succeed. “We only recommend students… that showed us they had the will… [identified by] effort and responsibility for learning [grades] on their report card.”

For both school site leadership and teachers, the district offered guidance and recommendations but not firm guidelines for how the policy must be implemented. At the school level, principals and counselors demonstrated references to previous understandings as they interpreted the new policy through their commitments to older placement guidelines. While the district staff members recommended that transfer students should now be individually evaluated before they were placed into a double dose math class, leadership at some school sites maintained their old placement guidelines of initially placing every transfer student into the intervention. “Some of the counselors are very used to the way they’ve always done things” a district staff member told me. In this example, school leadership interpreted their view of the new policy as needing perceived successful elements from the last iteration.

Teachers exhibited the most frequent references to incorporating previous knowledge and understandings as they interpreted the policy and considered how to teach it within their classrooms. Commonly, teachers referred to adapting the course curriculum, provided by the district, to their own interpretation of what these types of students needed. “They give us all the resources that we can use,” one teacher described, “and then we can pick according to our class and how our class runs.” However, teacher preferences influenced more than just instruction. One conversation between two teachers revealed that they had different previous understandings, and made differing recommendations to their site counselors based on these understandings, about the merit of having the same teacher for both the traditional and double dose match class:

> Teacher 1: And that's another thing is I requested this coming year, if there's any way I could have the [double dose] kids in my regular class… I need to monitor their progress a little better. I hear about the improvements teachers see, but I don't hear it consistently.
>
> Teacher 2: And I'll play devil's advocate. I'll say that it's nice not having them [twice]. I think there's a certain amount of advantage -- depending on who your staff is, depending on how strong your math team is -- that they do hear it a different way. It is good to hear it a different way.

4) Referencing consideration of organizational structures

Considering organizational structure is another common sensemaking activity identified by Coburn et al. (2009a), but unlike in their setting, this was not a frequently discussed topic among the sensemaking actors within this case study. Individuals did not commonly refer to the roles of others or the hierarchical structure of the school district. The organizational structure of the school district office was not brought up, nor was the organization of channels for making or disseminating those decisions. Sensemaking actors were clear about their roles but did not reference any adaptations or considerations that were made because of the school district's organizational structure.

5) Referencing consideration of obstacles or constraints

Data were coded to identify three types of constraints (funding, school/classroom structure, pressure from policies or standards) as well as miscellaneous obstacles discussed. In the sensemaking process, Coburn (2001) acknowledged that considering and adapting to constraints involving the policy is a primary sensemaking activity. In total, I used codes within this category 29 times. The codes for reference to obstacles and constraints were used more than any other category of specific sensemaking activities, with references to considerations of accountability standards, school/classroom structure limitations, and funding limitations used most often.

Of these three limitations, pressures from accountability standards and new national policies were referenced the most (16 times in total). District administrators described strong sensitivities to these pressures. Notably, the impetus for deciding that students needed a math intervention in general was in direct response to policy and accountability changes, and this pressure waxed and waned multiple times within only a few years. As described by multiple administrators, the original implementation of the policy was a direct reaction to California's "Algebra for All" policy. In 2008, the California Board of Education voted to make Algebra proficiency the standard of proficiency for eighth graders, thus incentivizing schools to enroll as many students as possible in Algebra by the end of middle school (Rosin, Barondess, & Leichty, 2009; Domina, McEachin, Penner, & Penner, 2015). Doubled math courses were implemented in this district to help students succeed in the mandatory Algebra classes in eighth grade.

Part of the reason these classes were eventually dropped was, as the district adopted the Common Core State Standards (California Department of Education, 2018), district staff members interpreted a stronger focus in the new standards on English Language Arts. As one administrator described it, "There was more of a concern to focus in on wherever the new standards were going to be around Common Core, which had a lot to do with the literacy standards… there was an urgency for schools to say, 'we're in this ambiguous time, let's see what we can do to support students.' We used these [double dose math] classes for a long time, which really came out of a need to address student success versus being defined by standardized testing, and then we moved out of that."

Yet, only a few years later, pressures from the new Smarter Balanced Assessment Consortium (SBAC) assessment results were at the forefront of the district administrators' minds.

> Administrator: We had traditionally done better [in mathematics] when we looked at our standardized test scores... we would actually do better in many grade levels in mathematics than the county as a whole. But we wouldn't do as well in English Language Arts (ELA). Then we get SBAC, and all of sudden that switches. So now we're doing better in ELA, and we're matching the county in all grade levels. But we're not doing as well in mathematics. In fact, we're underperforming in mathematics. We were doing better at mathematics against the county. Now we're not. So that's a big reason why we're investing more time and energy into the [double dose math] classes.

Another district staff member echoed this sentiment, describing the period when No Child Left Behind (NCLB) accountability policies gave way to adoption of the Common Core Standards, SBAC assessments, and Every Student Succeeds Act (ESSA) as an "easy break" from old policies. She described the period in between district guidelines of offering the double dose math course as unstructured from the district's perspective. Schools were allowed to pursue whatever avenue they felt "best supported their kids for the new standards."

Constraints regarding school or classroom structure were also suggested as an influencing factor by multiple individuals. Because a double dose math class required an extra math period, this caused pressure on schools' staffing balance. District administrators were aware that requiring double dose math classes could be a burden on smaller school staffs and class sizes for non-intervention math courses. However, the limits of the school day also pushed administrators toward prescribing the double course over different interventions such as after-school support or extra tutoring. One district staff member acknowledged that an intervention within the school day could be preferable because "we've had conversations about getting students to actually come consistently after school, which can be quite difficult." Another district staff member referenced the school day model as more accessible. "Sometimes after-school stuff is brought into that equity and access banner, because of different obligations that students may or may not have… or families may or may not have."

The obstacle of funding was referred to less than policy pressures or school structure, but its inclusion was notable. In describing what factors went into determining the feasibility of offering the double dose math courses, an administrator listed funding as a primary factor that influenced other components of the decision.

> Administrator: You've got declining enrollment taking place… so you're going to have some limited staffing there. We're allotted funds and it's shrinking over time because… schooling has become more expensive with technology. So, as you're buying materials now, there's a technology component that comes along with it. As we're evaluating resources, money, time, and staffing, that's going to constantly be something that year to year to what extent can schools offer the [double dose math] classes and to what benefit they are to students.

This same administrator also described the variety of reasons for the original decision to eliminate the double dose math course policy as a mixture of pressure from funding, structure, and standards.

> Administrator: You had three factors hitting all at once. You had a bad budget, schools were struggling with staffing, and class sizes were large. You had accountabilities shifting as you were going from No Child Left Behind to ambiguity about whatever was to replace it, which is now ESSA.

6) Referencing use of evidence in decision making or sensemaking activities

Within the final category of specific sensemaking activities, I coded individuals' references to consulting evidence as they interpreted the double dose math course policy. I also coded for instances in which it was specifically mentioned that no evidence was used. In total, I identified 31 references to evidence use and only three instances of non-evidence use.

District administrators and staff members referenced a variety of types of evidence and forums for collecting evidence as they implemented, altered, and evaluated the policy. This evidence broadly fell into two categories, either student assessments or feedback from teachers and principals. Students' standardized assessments were often consulted to understand if the course was producing its desired effect of improving student's math skills. This was consistent from the previous iteration of the course into the current version and appeared to be the norm for understanding many policies within the district. A former teacher, now district staff member, described the tools used to evaluate the course as "the typical" combination of benchmark scores, grades, and state standardized test scores that the district often uses. The same staff member said the district was also experimenting with student surveys about the course, although the surveys were not mandatory and not widely administered.

District employees also consistently referred to collecting feedback from principals and teachers, which was done in both formal and informal channels. The district hosted meetings between staff members and principals; checking in about the double dose courses was a planned discussion during some of the meetings. The recommendations from the principals to district decision-makers were used to identify issues with implementing the courses. District employees held similar meetings with teachers (although less frequently), during which the challenges of teaching the course and their curricular pacing were directly discussed. However, although these meetings took place, there were no references made to decisions that were made or altered based on feedback from these formal meetings.

Lower-level district employees, those in charge of implementing the courses and supporting the school sites, were more likely to collect feedback from principals informally while at their school sites or attending other meetings. They had more traditional channels for collecting feedback from teachers during professional development sessions and site visits. During the professional development session I observed, one of three sessions planned for that year, brainstorming or sharing best practices about obstacles faced was a large component of the allotted time. This is a notable amount of time designated for sensemaking amongst teachers and district staffers, as some teachers do not have peers also teaching a double dose course at their school. Informal surveys of teachers also occurred during site visits and individual coaching sessions. For example, a district staff member and the classroom teacher discussed a team quiz activity the students were doing during one of the classroom observations I conducted. The district staff member had never heard of this particular activity and the teacher was explaining why they thought it was helpful for their intervention students.

Because much of the feedback the district uses as evidence for evaluating the course is coming from school principals, it seemed pertinent to know how principals would describe the evidence they use to formulate opinions about the policy. Principals tended to mirror the district employees in this regard. The principal I interviewed listed course grades, student assessments, and teacher observations of a student's perceptions of math as the main pieces of evidence regarding student performance they were tracking. Multiple current teachers of the course also indicated that their principals offered time to give feedback about the course during staff and department meetings.

Citing use of evidence to understand the policy within the district was far more common than referencing decisions made without use of any evidence, although two observations regarding the way in which evidence was actually used were notable. One member of the district team in charge of implementing the course was responsible for keeping and tracking assessment data regarding the course's effectiveness. During the teacher professional development session, I observed a presentation by this staff member on double dose math students' benchmark test scores through the midway point of the school year. The presentation pointed to improved student performance from the first to second assessment of the year. However, this conclusion was based on change in test score for double dose math students only, without any comparison to whether students outside the intervention were experiencing similar growth. In this case it is possible that some of the evidence used to evaluate the policy was interpreted incorrectly.

The other notable reference to evidence collection came during a focus group with district staff members. This focus group took place during the school year in which the double dose class was being administered, and I asked about what measures would be used to evaluate and learn more about the course. As mentioned earlier, grades and test scores were listed as definite sources of evidence. However, staff members were unsure of what other pieces of evidence they might use or collect. One district staff member described it this way:

> Right now, honestly, I don't know if we're 100% sure what's the best thing to look at, but we're trying to look at all of it and see do we notice trends, are there any patterns we're recognizing.

This reference was important, as it indicated that district staff members might not understand the policy thoroughly enough to clearly identify and collect specific evaluation evidence from the beginning of the year. Evidence used and cited would be determined as the policy unfolded, not beforehand, demonstrating symbolic rather than instrumental evidence use as defined by Coburn at al. (2009).

Primary Research Questions

1) How did district administrators engage in collective sensemaking to re-implement a double dose math course intervention?

Previous studies highlighted four specific kinds of sensemaking activities prevalent amongst district administrators: referencing previous knowledge and understandings, acknowledging organizational structure, navigating obstacles and/or constraints, and incorporating evidence and data. In this case, district exhibitors frequently engaged in three of these activities. There were hardly any references to district organization structure within these data. For two of the other three activities, there were notable differences in the most common examples between different hierarchical levels of district employees. As presented in the examples in the previous section, upper-level district administrators (Assistant Superintendents, Curriculum Directors) engaged in references to previous knowledge and understandings and navigating constraints in different ways than lower-level district employees (e.g. Curriculum Specialists) rather than as a unified set of district employees. However, use of evidence in sensemaking was similar between the two groups.

These two groups also experienced different timelines in their sensemaking activities about this policy. Upper-level district administrators engaged in almost all of their sensemaking activities prior to the decision to reintroduce the policy (e.g. responding to various pressures, mirroring an earlier policy) while the lower-level district employees charged with implementing the policy were assembled after this decision was made. After this team was in place, upper-level district administrators were primarily concerned with program evaluation while the lower-level district employees were closely connected to school sites and made necessary adaptations. The structure of the district office, although it was hardly referenced by the sensemaking actors themselves, was a clear driver of differentiated sensemaking activities amongst these two groups.

It is also notable that the sensemaking activities of school-level employees (principals, counselors, and teachers) were deeply embedded into the sensemaking activities of district officials. This was also differentiated between the two groups of district employees. Upper-level district employees exclusively invited principals and teachers to meetings, where they incorporated feedback into their evaluation of the policy. There was no indication that shared sensemaking took place in these meetings; the upper-level district employees were seeking information. Principals and teachers engaged in sensemaking about the policy, then brought these ideas to the meetings, where their feedback was used as evidence about the policy by district administrators. However, lower-level district employees were more likely to engage in collective sensemaking with school principals and teachers. During site visits and organized professional development sessions, lower-level district employees exchanged ideas, brainstormed solutions to problems, and worked toward new, shared understandings with school staff members.

2) How were decisions about the altering and re-implementation of "double dose" math classes made and who made them?

To understand the timing, actor, and influences of the key decisions made during the time period under study, Table 2.3 summarizes these processes as they were described by the informants. As previously described, there is a clear delineation in the timing of decisions made by upper-level vs. lower level district staff members. Upper-level district administrators were responsible for earlier and larger-picture decisions about the policy, then lower-level district staffers and school staff took over decision-making processes as the policy was implemented and adapted. A pattern of timing also emerged in regard to the transitions between iterations of the policy as well. Each period of transition occurred as the district was responding to either changing state policy or curricular standards that the district was following.

District administrators describe the original implementation of the policy as a direct response to California's Algebra-for-All policy. The district believed the doubled math instruction would help more low-performing students pass the now-mandatory Algebra course in eighth grade. They also specifically reference the intended outcome of the course as Algebra readiness. For seventh-graders taking a Pre-Algebra course, this meant being able to take and pass Algebra the following year, hopefully without the additional support course in eighth grade. For eighth graders, this meant being supported enough to pass Algebra that year.

Next, the decision to formally discontinue the course occurred as the district was interpreting the early stages of the new Common Core State Standards. California's Algebra-for-All policy was no longer in place and the district viewed the new standards as more heavily focused on English Language Arts. Additionally, although the district's students in general were exhibiting above-average scores on standardized math testing, the district did not feel that the double dose math courses were contributing to this success. District administrators considered the benchmark assessment and year-end standardized math test scores of students enrolled in double dose courses and determined that the courses were ineffective. The combination of new standards perceived to emphasize English Language Arts and lack of improvement in test scores for double dose math students drove the decision to phase out the policy.

However, the district administrators were surprised by their students' performance on the new standardized math tests under the Common Core State Standards that were rolled out the following year. Although district administrators expected an overall drop in scores as students and teachers adjusted to the new assessments, they described feeling particularly concerned that low-performing students would have continued troubles with the transition. It was not stated whether this concern was based on any specific data or evidence about the performance of the below-proficient students. This caused the district to seek out a way to offer extra support to below-proficient students in math and led to the re-implementation of the double dose math policy with major modifications ("Version 2.0"). Overall, the decisions regarding this policy across the eight-year span were extremely sensitive to pressures from state policy, changes in state curricular standards, and student assessment performance during these policy transitions.

The decisions of the district administrators were also highly motivated by pressures from resources and existing school/classroom structure. These two items were listed as central components to the decision to eventually return to the double dose course policy, as it would require less money and no additional staffing to implement a policy that had previously been in place and it would be hard to guarantee student attendance in an intervention that was structured differently. These three types of pressures, factored in simultaneously, were continuously stated as the reasoning behind returning to a policy that was previously deemed unsuccessful. More so than in previous sensemaking analyses, the policy decisions in this case study were direct responses to policy and standards changes with accommodations for constraints within the district.

Although lower-level district employees, school principals, and teachers were all consulted about the decisions, ultimately upper-level district administrators at the Assistant Superintendent level were the sole decision makers about this policy. Within the district in this case, the Assistant Superintendents in charge of secondary instruction direct the creation of course guides that are given to the school sites. These guides indicate priority classes that schools should offer and what types of students should be enrolled. Although school sites do have autonomy to adjust and adapt from the recommendations, they are expected to closely follow them. In the original implementation, the removal, and then the return to offer these classes, upper-level district employees were responsible for the final decisions to emphasize or remove them from the guides accordingly.

3) How and what did district administrators learn as they modified the policy?

District administrators referenced various items that were modified in the reintroduction of the policy and the reasons behind these modifications. The changes can be categorized broadly within these areas: student placement, curricular structure, space for teacher sensemaking and collaboration, and overall goals for the intervention. These categories and the main modifications under each are summarized in Table 2.4.

Student Placement

In the reintroduction of the double dose math course policy this district altered its target population of students. Within the first policy implementation, all students that fell below math proficiency were recommended for enrollment. Additionally, students who transferred to the district and did not have full standardized test data for their placement were recommended to enroll in doubled math instruction math and then get re-evaluated by their math teacher at the semester break. District administrators learned that this system was problematic for admitting too many students into the intervention, notably students who were too far below proficiency for this intervention to provide enough support, students who had extreme behavior issues (and correspondingly low math performance), students who were just below proficiency and did not need such an intensive support, and too many transfer students who did not need the intervention but could not switch out until the semester.

Under the new placement guidelines, students at the lowest and highest ends of the below-proficiency achievement spectrum were not recommended to enroll in the intervention. Neither were students with poor citizenship grades on their report cards. These modifications to student placement were made by the lower-level district administrators based on principal and teacher feedback about student performance in the original courses. The district also encouraged school sites to only admit transfer students who they could identify as a good fit for the course, although some sites were less cooperative with this recommendation than others. In the end, the district modified the student placement policy to encourage a more finite group of students with a targeted range of math skills and better classroom behavior to be enrolled in the course.

Curricular Structure

District administrators cited the lack of consistency in the curriculum offered during the course as a main flaw of the first iteration of the policy. Previously, the district provided only curricular resources to teachers without set guidelines, pacing instructions, or planned units of material to cover. As they described it, this caused variance in the style of instruction and material covered from classroom to classroom, as teachers were given complete freedom to interpret the goal and best methods for the intervention.

> Former Teacher: I think that one of the problems I found from the class was that there wasn't so much direction given, so that's why you'll hear that every single teacher would have a totally different experience with that. We were not given too much information. We were not given many directives or expectations, so every different teacher is going to be totally different.

In the new version of the course, considerably more curricular structure is embedded. Teachers are still allowed a degree of autonomy in selection of activities, daily schedules, and instructional methods; however, the new course follows planned units in conjunction with the students' traditional math course. The district also implemented a weekly activity allotment to ensure diversified instruction, including guidelines for days spent on traditional lecture, computer-based learning, and project-based learning.

Space for Teacher Sensemaking and Collaboration

Although they didn't use the exact term, district administrators learned that the previous iteration of the course lacked space for double dose math teachers to engage in sensemaking and collaboration:

> Administrator: You would find that with our [intervention] classes, sometimes it became a challenge because there wouldn't be another [intervention] teacher [at that school]. Or if there was another [intervention] teacher, that [intervention] teacher would be at the different grade level. So [the teachers] could talk in general, maybe some things that [they] wanted to do, but you couldn't share lesson plans like you could if you were teaching the same content.

District-organized professional development sessions, which were optional and infrequent in the previous iteration, were formalized to be used as this space. Now, professional development sessions are scheduled consistently and feature large portions of time dedicated to sharing best practices, working through common obstacles, and brainstorming solutions to common problems. One of the lower-level district staff members explained, "There's a lot of solo people at their sites and they can't collaborate with anybody... They find it very valuable that they can get time to actually talk to each other and share ideas."

Overall Goals for the Intervention

In the transition between the two iterations of the intervention, the district also modified its vision and overall goals for how this intervention should be supporting students. Previously, the district viewed the course like a supportive boost to ensure students were ready for or supported during their Algebra course. As one former teacher described it, "That was the focus of the class, it was very much academic. It was making sure that we were catching them up and getting them prepared for their other class." In the new version, although the reiteration of the policy was a response to academic concerns, the district administrators describe the goals of the course much differently.

> Administrator: Our goal is to not only have kids who are better at math, but kids who feel better about math. That was really one of the big things… we wanted kids to see themselves as mathematicians versus 'I'm never going to be good at math.' [We are] hopefully trying to affect if we can, even on a small scale, the fail cycle in high school. These kids who feel like they can't do it then can get into this horrible cycle. But if maybe we can get them to think, 'I can do this, I can put the effort in,' maybe it will have some effects there too.

These new goals for the course are described to include outcomes outside the realm of standardized math assessment scores and Algebra readiness. Confidence and new skills in math are a large part of the goal, but district administrators also see this course as part of the larger picture of improving graduation rates, self-confidence, and student motivation. Yet, although broader goals for the course are discussed, new metrics for determining success in reaching these goals were not specifically determined prior to offering the course.

DISCUSSION

Even though this case study centers on one district's decision-making process regarding an individual policy, it illuminates important issues regarding how and why curricular policy decisions can be made at the district level in general. It also documents a unique type of decision-making in which school district officials returned to a previously unsuccessful policy and then reinstated it with aims for improvement. This is contrary to the narrative of "policy churn" that is common in the literature; in this case, school officials led themselves through a period of reflection, improvement, and reintroduction of the same policy rather than seeking out something new. Why and how they did so is powerful to know and a contribution to the literature. Besides answering the research questions proposed at the beginning of this study, I was also able to observe some facets of this case study that are new to collective sensemaking and district policy literatures.

First, district staff members participated in a layered nature of sensemaking about multiple policies at the same time. More so than previous observations of district administrators, this case demonstrates that the sensemaking activities regarding one policy are often inseparable from sensemaking activities and decisions on other policies or considerations that overlap with this particular policy. In this case, administrators' sensemaking activities and decisions surrounding the "double dose" math policy were also part of their reactions to changes in state curricular standards in general. Particularly, their decisions to discontinue and then reintroduce the course were direct responses to the evidence of student test performance that was informing their sensemaking on the new curricular standards. In this case, sensemaking about the course policy was embedded within sensemaking activities about larger-scale policies.

This case also highlights a unique aspect of this district's operations regarding the decisions and implementation of the restructured math course policy. Specifically, the activities of decision-making and then course implementation involved two unique, non-overlapping groups of district staff members. Decisions regarding the math course policy were informed by information from lower-level district staff members (as well as principals and teachers) but made strictly by upper-level district administrators. Implementation decisions (curricular structure, professional development for teachers, placement guidelines) were then made strictly by lower-level district staff members and reported to higher-level staffers. District staff members described this as a common decision-making and implementation structure within this district. This structure differs from previous cases (Coburn et al., 2009a) that noted hierarchy within the district office as an influence on policy decisions. In this case, almost no references to the hierarchy were made but the structure was clearly delineated in the roles of each group of district staff members. This is possibly a function of size of the district's staff as this district is quite large in comparison to most U.S. school districts.

Among the team of lower-level district employees charged with implementing the policy, these decision-makers were open to seeking outside expertise on curricular materials for the course but not any information on how other districts had administered the policy. This provides more nuance to the Spillane's (2009) findings that non-government agencies, universities, and granting foundations could be heavy influencers on the details of policies adopted by district leaders. In this case, district staffers referenced consulting outside curricula and feedback from math specialists at a nearby university about the structure of the course's units and scheduling for the year. However, they indicated that they found no evidence of what guidelines other districts used to implement similar courses (e.g. student placement guidelines, using simultaneous periods for the traditional and intervention courses, teacher selection). How thoroughly this search was conducted is not known; however, models of how other districts have implemented these courses and whether or not they were effective are documented in many research analyses (e.g. Bartik & Lachowska, 2014; Nomi & Allensworth, 2013; Taylor, 2014). The district employees were either willing to look outside the organization for curricular materials but not for examples of implementation guidelines, or, unable to find examples of the latter because they are located outside their traditional pool of resources (accessible available through research journals and research databases).

Another clear pattern emerged surrounding evidence use within this case district. Returning to Coburn et al.'s (2009b) classification of evidence used during policy sensemaking as either instrumental, conceptual, symbolic, sanctioning, or no role, the district in this case exhibiting signs of evidence use in all five ways. However, since performance metrics or benchmarks were usually not clearly laid out beforehand, most instances could be categorized as a hybrid of instrumental (thorough consultation of evidence before making a decision) and symbolic (looking only to evidence that reinforces preexisting understandings). The district maintained thorough data and referenced this data in their decision-making process, but the data they chose to use was not determined beforehand and subject to bias based on previous understandings. In an era where school districts have growing amounts of data to choose from in their evaluation of curricular policies, this strikes a warning bell for the importance of proper evaluation methods and pre-determined, theory-based metrics of policy effectiveness that are not subject to useful interpretations.

Finally, akin to previous studies detailing that some sensemaking activities among classroom educators are made possible through access granted by principals (Coburn, 2005), I found this also to be the case for access granted by district staff members. In this case, I observed that district administrators granted access in three ways. First, more structure provided by the district office proved to be a necessary foundation for any collective sensemaking among teachers to occur. Second, dedicated professional development sessions allowed teachers from other schools, who are regularly unable to engage in collective sensemaking, to do so. Third, district staff members framed the teachers' perceptions of the policy in similar ways as principals have been shown to do. District staffers specifically referenced instances of framing the policy to their teachers to garner a more positive or supportive reaction to implementation. Teachers' access to, and how they are able to, make sense of a policy is shaped by both principals and district staff members in unique but similar ways. Additionally, district leaders have power to bring teachers together across school sites, which school site principals usually do not.

REFERENCES

Bartik, T., & Lachowska, M. (2014). *The effects of doubling instruction efforts on middle school students' achievement: Evidence from a multiyear regression-discontinuity design* (No. 14-205). Kalamazoo, MI: WE Upjohn Institute for Employment Research.

Bryk, A. S., Sebring, P. B., Allensworth, E., Easton, J. Q., & Luppescu, S. (2010). *Organizing schools for improvement: Lessons from Chicago*. Chicago, IL: University of Chicago Press.

California Department of Education (2018) Common Core State Standards website: accessed February 19, 2018 at https://www.cde.ca.gov/re/cc/

Coburn, C. E. (2001). Collective sensemaking about reading: How teachers mediate reading policy in their professional communities. *Educational Evaluation and Policy Analysis*, *23*(2), 145-170.

Coburn, C. E. (2004). Beyond decoupling: Rethinking the relationship between the institutional environment and the classroom. *Sociology of Education*, *77*(3), 211-244.

Coburn, C. E. (2005). Shaping teacher sensemaking: School leaders and the enactment of reading policy. *Educational Policy*, *19*(3), 476-509.

Coburn, C. E., Honig, M. I., & Stein, M. K. (2009a). What is the evidence on districts' use of evidence? In J. Bransford, L. Gomez, D. Lam, & N. Vye (Eds.) Research and practice: Towards a reconciliation. Cambridge: Harvard Educational Press.

Coburn, C. E., Touré, J., & Yamashita, M. (2009b). Evidence, interpretation, and persuasion: Instructional decision making at the district central office. *Teachers College Record*, *111*(4), 1115-1161.

Domina, T., McEachin, A., Penner, A., & Penner, E. (2015). Aiming high and falling short: California's eighth-grade Algebra-for-All effort. *Educational Evaluation and Policy Analysis*, *37*(3), 275-295.

Hatch, J. A. (2002). Deciding to do a qualitative study. *Doing qualitative research in education settings*, 1-36.

Hess, F. M. (2011). *Spinning wheels: The politics of urban school reform*. Brookings Institution Press.

Jackson, C. K., Johnson, R., & Persico, C. (2014). *The effect of school finance reforms on the distribution of spending, academic achievement, and adult outcomes* (No. w20118). Cambridge, MA: National Bureau of Economic Research.

Merriam, S. B. (1998). *Qualitative research and case study applications in education. Revised and expanded from" Case Study Research in Education."*. Jossey-Bass Publishers, 350 Sansome St, San Francisco, CA 94104.

Meyer, J. W., Rowan, B., & Meyer, M. W. (1978). *The structure of educational organizations*.

Nomi, T., & Allensworth, E. M. (2013). Sorting and supporting: Why double-dose algebra led to better test scores but more course failures. *American Educational Research Journal*, *50*(4), 756-788.

Reardon, S. F., & Owens, A. (2014). 60 Years after Brown: Trends and consequences of school segregation. *Annual Review of Sociology*, *40*, 199-218.

Reed, D. S. (2014). *Building the federal schoolhouse: Localism and the American education state.* Oxford: Oxford University Press.

Rosin, M. S., Barondess, H., & Leichty, J. (2009). *Algebra policy in California: Great expectations and serious challenges*. EdSource.

Smith, M. S. (2000). Balancing old and new: An experienced middle school teacher's learning in the context of mathematics instructional reform. *The Elementary School Journal*, *100*(4), 351-375.

Spillane, J. P. (1996). School districts matter: Local educational authorities and state instructional policy. *Educational Policy*, *10*(1), 63-87.

Spillane, J. P. (2000). Cognition and policy implementation: District policymakers and the reform of mathematics education. *Cognition and instruction*, *18*(2), 141-179.

Spillane, J. P. (2009). *Standards deviation: How schools misunderstand education policy*. Cambridge. MA: Harvard University Press.

Spillane, J. P., Diamond, J. B., Burch, P., Hallett, T., Jita, L., & Zoltners, J. (2002). Managing in the middle: School leaders and the enactment of accountability policy. *Educational Policy*, *16*(5), 731-762.

Taylor, E. (2014). Spending more of the school day in math class: Evidence from a regression discontinuity in middle school. *Journal of Public Economics*, *117*, 162-181.

Vaughan, D. (1997). *The Challenger launch decision: Risky technology, culture, and deviance at NASA*. University of Chicago Press.

Weick, K. E. (1995). *Sensemaking in organizations* (Vol. 3). Thousand Oaks, CA: Sage.

Yin, R. K. (1994). *Case study research: Design and methods*. Thousand Oaks, CA: Sage.

CHAPTER 2 TABLES & FIGURES

Table 2.1

Summary of data collection for Chapter 2

Data collection activity	**Participants**	**Interview protocol document (if applicable)**	**Date completed**	**Details**
Preliminary information meeting about district use of the course	4 district staff members		August 2016	1 hour in length, detailed notes taken
Focus group on previous implementation of courses	2 former teachers, 2 district staff members	Appendix Document 2.1	December 2016	1 hour in length, recorded and transcribed
2 focus groups on the re-design and current implementation of courses	4 district staff members	Focus Group #1: Appendix Document 2.2 Focus Group #2: Appendix Document 2.3	February 2017 – June 2017	Focus Group #1: 1 hour in length, recorded and transcribed Focus Group #2: 30 minutes in length, recorded and transcribed
Professional development session for teachers of current course	6 district staff members and 11 teachers		February 2017	Two hours in length, field notes of naturalistic observations taken
Classroom observations of current course	Multiple class periods with 4 different teachers		June 2017	One hour in length per period, field notes of naturalistic observations taken
Focus group with teachers of current course	4 current teachers	Appendix Document 2.4	June 2017	60 minutes following an interview protocol, recorded and transcribed
Interview with upper-level district administrator overseeing the revamped course	1 interview with district director	Appendix Document 2.5	July 2017	60 minutes following an interview protocol, recorded and transcribed
Email interviews	1 current Principal	Appendix Document 2.6	July 2017	1 set of interview questions answered over email

Table 2.2
Qualitative coding framework

	Code	Definition	Examples
Structural codes			
	CLA	Classroom setting	Activities described take place within a classroom or school setting
	DIS	District office setting	Activities described take place within the district administrative offices
	PD	Professional development setting	Activities described take place within a professional development setting for intervention teachers
Direct action codes			
Policy adaptation	ADAPT-CLA	Adaptation in policy implementation at classroom level	Indication of an autonomous adaptation from the designed policy at the classroom level
	ADAPT-DIST	Adaptation in policy implementation at district level	Indication of an adaptation from the designed policy at the district level
	ADAPT-SCHOOL	Adaptation in policy implementation at school level	Indication of an autonomous adaptation from the designed policy at the school level
Decision making	DM-ACT	Decision-making actor	An individual that is indicated to be involved in or responsible for making a decision about the policy
	DM-MADE	Decision made	Indication of a clear decision made about the policy
	DM-CON	Decision-making consult	Indication that outside materials or individuals were consulted in the decision-making process
	DM-MISC	Decision-making (miscellaneous)	Other notable reference to decision making
Descriptive codes			
General Sensemaking	SM-ACCESS	Sensemaking access	An individual that is involved in defining someone else's access to understanding the policy
	SM-ACT	Sensemaking actor	An individual that is involved in sensemaking about the policy
	SM-ADMIN	Sensemaking among administrators	Indication of sensemaking activities within a group of district administrators

	SM-SOCIAL	Sensemaking in a social setting	Indication of sensemaking activities within a group of individuals
	SM-SOLO	Sensemaking in a solo setting	Indication of sensemaking activities with no other individuals
	SM-TEACH	Sensemaking among teachers	Indication sensemaking activities within a group of teachers
	SM-UNDER	Sensemaking reference to working toward shared understandings	Indication of a reference to building new, shared understandings
	SM-MISC	Sensemaking (miscellaneous)	Other notable reference to sensemaking activities
Previous knowledge and understandings	PRE-DEFY	Reference to defying previous knowledge	Indication of a reference to defying previous knowledge or prior orientation while decision making
	PRE-KNOW	Reference to previous knowledge	Indication of a reference to previous knowledge or prior orientation used in decision making
	PRE-MISC	Previous knowledge (miscellaneous)	Other notable reference to previous knowledge
Organizational structure	ORG-DEPT	Organizational structure - department	Reference to a department within the organizational structure of the district office
	ORG-STRUC	Organizational structure - general	Reference to the organizational structure of the district office
Resource constraints	RES-FUND	Resource constraints regarding funding	Reference to consideration of school/district funding
	RES-SCHOOL	Resource constraints regarding school or classroom structure	Reference to consideration of school/classroom structure (e.g. class size, population)
	RES-STANDARDS	Resource constraints regarding accountability standards	Reference to consideration of accountability standards / standardized testing
	RES-MISC	Resource constraints of any other kind	Reference to consideration of uncategorized organizational structure
Evidence use	EV-DM	Decision-making evidence	Indication that evidence was used to inform a decision-making process about the policy
	EV-INST	Instrumental use of evidence	See Coburn, Honig, and Stein (2009)

	EV-CONC	Conceptual use of evidence	" "
	EV-SYM	Symbolic use of evidence	" "
	EV-SANCT	Sanctioning use of evidence	" "
	EV-NO	Reference to no use of evidence	" "
	EV-MISC	Reference to non-category use of evidence	Does not fall within a category classified within Coburn, Honig, and Stein (2009)

Table 2.3

Timeline and actors of key decision-making processes

Chronological Order (Year)	2013-14	2015	2015-16	2016-ongoing
Key Decision	Discontinue previous version of the intervention	Develop and pilot revised version of the intervention	How intervention will be revised and implemented	Necessary adaptations for school sites
Decision-Making Actor(s)	Upper-level district administrators	Upper-level district administrators	Lower-level district administrators	School principals
Who/what was consulted?	Principals, teachers, student standardized test scores, changing state standards policies, resource pressures	Student standardized test scores, changing state standards policies, resource and structural limitations	Upper-level administrators, School site staff (principals, counselors, teachers), outside expertise (local University's math curriculum experts), resource limitations	School staff (teachers and counselors), Student needs, resource limitations, scheduling demands

Table 2.4
Key Revisions to Intervention Structure

Intervention Structure	Previous implementation	New implementation
Student placement	Students scoring 3-8 points on a 3-15 scaled based on three prior assessments	Students scoring 4-7 points on a 3-15 scaled based on three prior assessments; Students must also have satisfactory classroom behavior grades in the prior year
Curricular Structure	Instructional resources offered but no structured curriculum or guidance on specific instruction; optional professional development sessions on learning activities	Curriculum coordinated with traditional math course; guidance on weekly variation of instruction techniques; Instructional resources offered; professional development sessions (some mandatory, some optional) on structured curriculum
Space for Teacher Sensemaking and Collaboration	Optional professional development sessions	Professional development sessions (some mandatory, some optional); site visits and individual coaching with district staff members
Stated Goals for the Intervention	Improving math achievement and preparedness for Algebra	Improving math achievement and confidence in math abilities; Improving larger-scale outcomes such as graduation rates, academic self-confidence, and student motivation

APPENDIX DOCUMENT 2.1

“Double Dose” Math Class (Previous implementation) Focus Group Protocol

What was the curriculum for the double dose class? Was it aligned with students’ traditional math courses?

What was a typical week of instruction like? Was homework offered?

Were students assessed regularly in their double dose class? On what were they assessed?

Did you often teach your students in both their double dose and traditional class? Were students usually grouped with double dose students in their traditional class?

Were teachers offered special professional development, materials, or guidance on how to teach these courses? If so, what kind?

How did students report feeling about the double dose classes?

What about the courses did you feel was helpful for your students?

What about the courses did you feel was not helpful for your students?

Were there any obstacles or issues you regularly faced while teaching the double dose class(es)?

How did you and other teachers feel about the courses as you were teaching them?

How did you principals and/or administrators feel about the courses?

APPENDIX DOCUMENT 2.2

"Double Dose" Math Class (new implementation) Focus Group Protocol

1) What is the curriculum for the double dose class? How is it aligned with students' core math courses?

2) What is a typical week of instruction like? Is homework offered?

3) Are students assessed regularly in their double dose class? On what were they assessed?

4) Do teachers teach double dose students in both their double dose and core class? Are students usually grouped with fellow double dose students in their core class or spread out?

5) Are teachers offered special professional development, materials, or guidance on how to teach these courses? If so, what kind?

6) How are teachers assigned to teach the double dose courses? Are they typically teaching more than one in a day?

7) Are there any ways in which you try to keep teachers organized in pacing, material covered, or strategizing around obstacles? Essentially, does your team try to oversee these courses now that they're in place?

8) Recap: What has changed about these courses from the previous iteration?

9) Where did your team look for inspiration and ideas on how to change the double dose courses?

10) Have you heard any reports about how students are feeling about the double dose classes?

11) What do you feel will be most helpful about the new double dose courses for your students?

12) Have you heard any reports about how teachers are feeling about the new version of the double dose classes?

13) Have teachers reported any obstacles or issues they are regularly facing while teaching the double dose classes?

APPENDIX DOCUMENT 2.3

"Double Dose" Math Class District Administrator Focus Group Protocol

<u>Original setting</u>

What was the original impetus for offering double dose math classes?

Why were double dose math courses chosen instead of other types of support?

What kind of evidence was used to evaluate the effectiveness of the course?

Who was responsible for evaluating the course?

Who eventually decided to discontinue the course?

<u>Current setting</u>

Why was the policy chosen to be reintroduced?

Foundation relationship: interest in double dose classes from them as well? Or purely a district initiative?

What evidence was used to determine that the classes could be successful if altered and re-implemented?

Has it been tough to get buy-in from schools to try again?

<u>Re: Further interviews</u>
Who is best to speak to about:

- Evaluation of previous policy?
- Grant-writing process

APPENDIX DOCUMENT 2.4

"Double Dose" Math Class Current Teacher Focus Group Protocol

Intro, site/experience, taught double dose before?

How would you describe the curriculum for the double dose class? How is it aligned with students' traditional math courses?

What is a typical week of instruction like? Is homework offered?

Are students assessed regularly in their double dose class? On what are they assessed?

How would you describe the professional development, materials, and guidance on how to teach these courses that is provided by the district?

Do you utilize the resources and guidance provided by the district office to teach your course?

Are you in contact with other double dose course teachers during the year about their content, pacing, or teaching strategies? Other teachers in general?

Are you in contact with the district team during the year about content, pacing, or teaching strategies?

Are there any obstacles or issues you regularly face while teaching the double dose class(es)?

Is it possible to adapt your instruction to counter these obstacles? If so, how do you do so?

How do you think the students feel about the double dose classes?

Are there aspects of the courses that you feel are particularly helpful for your students? Not helpful?

Additional, if time:
Do you often teach your students in both their double dose and traditional class?

APPENDIX DOCUMENT 2.5

"Double Dose" Math Class District Administrator Focus Group Protocol

<u>Original setting</u>

What kind of evidence was used to evaluate the effectiveness of the course?

Who was responsible for evaluating the course?

Were other recommendations/feedback solicited through formal or informal channels?

Who eventually decided to discontinue the course or how was this decision made?

<u>Current setting</u>

Decision to reintroduce: Anyone else involved that would be good to discuss with?

Are there formal plans for evidence or feedback collection that will be used to evaluate the effectiveness of the course?

Is there anyone at the district that is directly responsible for evaluating the course?

Is there anything laid out about how future usage of the course will be decided upon? Decisions of whether to keep offering?

APPENDIX DOCUMENT 2.6

“Double Dose” Math Class Current Principal Email Questions

1) How do you make final decisions on which students are placed/enrolled in your math double dose courses?

2) How do you make decisions on how many total double dose class periods to offer each year? Are you balancing offers of these courses alongside other support courses/programs you have in place?

3) What do you see as the largest benefits of offering the math double dose classes at [your school]?

4) Do you encounter any obstacles to incorporating or utilizing the double dose classes in your school? If so, have you had to make any adaptations to how you offer them?

5) What kind of information/evidence do you use to interpret the effectiveness of the double dose math courses in your school?

6) Do teachers ever report to you on successes and challenges of teaching the courses? Do you utilize teacher feedback in your decisions on how you offer the double dose courses, and if yes, how so?

CHAPTER 3

Remediation Recycled: Evaluating a School District's Attempt to Restructure Seventh-Grade "Double Dose" Math Courses

INTRODUCTION

Overview

Enrolling students in a "double dose" math class—a second, complementary instruction period during the school day—has become an increasingly common strategy for schools hoping to boost the math proficiency levels of sixth through ninth graders (Durwood, Krone, & Mazzeo, 2010). A 2011-12 report on schools in North Carolina showed that ninety-six percent of high schools, including at least one high school in each district, and fifty-eight percent of middle schools within the state were utilizing double dose math courses in 2011-12 (Henry, Barrett, & Marder, 2016). When schools enact this policy for remediation purposes, students who are assessed to be below proficient on a given measure or measures are enrolled in a double dose math period instead of another required course or elective. In this second class period, educators expect that students will receive extra support on the current content and remedial instruction for gaps in foundational knowledge to eventually bring them up to speed with peers.

Although schools and districts are implementing double dose math courses more widely, there are only a few empirical studies of their effectiveness as an intervention for low performing students. The bulk of the empirical studies on these courses are from two large, urban settings (Chicago Public Schools and Miami-Dade County Public Schools) in which the findings were generally positive (Cortes, Goodman, & Nomi, 2015; Nomi & Allensworth, 2009; Taylor, 2014). However, other studies have uncovered mixed results and the literature as a whole covers a variety of implementation styles and settings (see also Bartik & Lachowska, 2014; Fryer, 2012;

Chapter 1). There is only limited evidence on the details of this intervention and on its effectiveness.

This study contributes to the literature on the effectiveness of double dose math course policies by studying the impact of this intervention on low-performing seventh-grade students in a mid-sized, suburban, Southern California school district. It is also the follow-up to a study on the same policy within the same school district. In Chapter 1, I used a regression discontinuity design to examine the effectiveness of the district's loosely-structured implementation of this education strategy. I found no significant impact of the intervention on end-of-year standardized math assessment scores, grades in the traditional seventh-grade math course, or traditional math course passage rates for students close to the cutoff for assignment to the course. However, more inclusive approaches using difference-in-difference-in-difference and OLS regression models indicated potential test score increases overall and specific benefits for double dose students near the bottom of the prior math achievement distribution. Estimates of the relationship between main math course grade and treatment were significantly higher for English Language Learners as well.

In this study, I am able to document what elements of the policy were revised and whether the courses had a different impact on a similar group of students in this second implementation. Using administrative data obtained from the school district, I employ a propensity score weighting design to answer three primary research questions: **Q1. Were the newly-implemented "double dose" math courses a successful intervention for low-performing seventh graders in this school district? Q2. Were the intervention courses uniquely impactful for students designated as English Learners, socioeconomically disadvantaged students, and/or students at different points along the prior math**

achievement spectrum? Q3. How do the results compare with those of the original implementation?

As a preview, the findings show that there is a significant association between enrollment in a double dose math course and increased math assessment scores at the end of that year, but interactions between treatment in the course and student demographic characteristics do not yield significant results. The courses are possibly more impactful in increasing standardized test scores for students at the highest end of the prior achievement spectrum that still qualifies for enrollment. Identical models using math course grades and math course passage as the outcomes yield no significant effects. Despite substantial changes to the course implementation, these results are not drastically different from those found for the first program iteration in Chapter 1, but they do seem to indicate less specific benefits across student subgroups. Additionally, differences in the population of targeted students and the timeframe of the two implementations could be important and are examined further below. Key strengths of this study involve an addition to the effect estimates of double dose math courses as well as an opportunity to examine the implementation of a different version of the same district policy within a very limited time period.

Literature Review

As detailed in the literature review of Chapter 1, the varied settings, implementation decisions, target population, and effects from analyses of different offerings of double dose math courses make it difficult to draw definitive conclusions about program effectiveness. The results of these analyses range from null to moderately positive and the generalizability of the positive findings outside of the studied districts is unknown. The bulk of the empirical studies on these courses are from two large, urban settings (Chicago Public Schools and Miami-Dade County Public Schools) in which the findings were generally positive (Cortes, Goodman, & Nomi, 2015; Nomi & Allensworth, 2009; Taylor, 2014). However, other studies have uncovered mixed results. Empirical studies in both a mid-sized district and charter school settings were not ideal and only found one positive and statistically significant effect: increased math test scores for eighth grade charter school students taking double math classes in comparison to those who did not (Bartik & Lachowska, 2014; Fryer, 2012).

Additionally, in Chapter 1 I analyzed the loosely-structured first implementation of double dose math courses for seventh-grade math students in the same setting as this study. Using a quasi-experimental regression discontinuity design to estimate a causal effect of the program, my results showed no significant impact of the intervention on the end-of-year standardized math assessment, grades in the traditional seventh-grade math course, or traditional math course passage rates for treatment versus control students within a close range around the enrollment cutoff. However, in more inclusive approaches using difference-in-differences-in-differences and OLS regression models that could incorporate the entire student sample, I found modest, positive overall intent-to-treat effects on standardized test scores that possibly stemmed from increases for double dose students near the bottom of the prior math achievement distribution (outside of the regression discontinuity bandwidth). I also found a significant, positive relationship between main math course grade and treatment interacted with English Language Learner designation (+.19 GPA points, $p<.05$).

One potential area of consistency within the literature is that this intervention is particularly helpful for students within different subpopulations. Cortes, Goodman, and Nomi (2015) highlighted the effectiveness of the policy for students with below average reading skills, introducing the idea that double dose math courses might be more helpful for students who just need extra time because of language barriers. I mirrored this finding in Chapter 1 and found differential impacts for students in the lowest category of math achievement. I also tested for differential impacts by socioeconomic status but found no significant difference.

I concluded Chapter 1 by suggesting that setting and policy implementation decisions might alter the effectiveness of the intervention and should be considered more carefully. One potential hypothesis that can be made from the previous literature and my first study is that this intervention is most successful when it is implemented with a well-developed structure. Thus far, the strongest and longest-lasting treatment effects on this intervention to date were found in the Chicago Public Schools, where teachers were offered rigorous professional development and the implementation guidelines were the most detailed. However, researchers were only able to isolate one component of implementation within the Chicago Public Schools, showing the program to be more effective for treatment students when they were matched in groups of students with the same teacher for both the traditional and intervention courses (Nomi & Allensworth, 2013).

Compared to the Chicago Public Schools, the district I studied implemented the original version of the intervention with less structure including a flexible, teacher-directed curriculum, optional, minimal professional development for teachers (rather than mandatory training sessions) on how to teach the course, and no determined plan for student sorting into math courses with the same peers. Although the potential impact of targeted curriculum and professional development factors have never been studied in regard to this policy, they have been studied within the literature on math instruction and student achievement in general. Specifically, learning from teachers with adequate training in delivering a specific curriculum designed to target math standards has been linked to increased math standardized test scores for middle schoolers in multiple settings (e.g. Reys, Reys, Lapan, Holliday, and Wasman, 2003; Tatt et al., 2008).

Much work still needs to be done to further understand the effectiveness of this intervention in different settings and how different elements of this intervention's implementation might relate to its impact. A setting in which multiple versions of the policy were offered in succession with different implementation strategies makes for an ideal setting for contributing to this literature.

Current Study

In the current study, I compare two recent versions of the same intervention, seventh-grade Pre-Algebra double dose courses for low performing math students, in a midsized, suburban school district in Southern California. Notably, I document the implementation decisions that were changed across the two different iterations of the intervention and whether the impact for students was different. This research setting is unique to the literature, and to literature on interventions of all kinds, as it allows for comparison of two different versions of the same intervention, across a similar student population, within a nine-year time period. I also continue the conversation of whether this intervention is particularly impactful for specific student subgroups. Besides testing whether the intervention is especially helpful for English Learners, as previous studies have done, I also run interaction models involving socioeconomically disadvantaged students and prior math achievement as I did in Chapter 1.

I conducted a propensity score weighting analysis using student-level data from this district to test the following questions:

1) Were the newly-implemented “double dose” math courses a successful intervention for low-performing seventh graders in this school district?
2) Were the intervention courses uniquely impactful for students designated as English Learners, socioeconomically disadvantaged students, or students at different points along the prior math achievement spectrum?
3) How do the results compare with those of the original implementation?

DATA

Setting

“Sunnyside School District” serves a current enrollment of more than 45,000 students, placing it within the top twenty-five largest school districts in the state of California. Approximately 75% of students qualify for free- or reduced-price lunch, more than 50% are Hispanic/Latino, and close to 33% are Asian or Asian American. Within these ethnicity groupings, the district enrolls large groups of first- and second-generation Mexican, Guatemalan, Salvadoran, Vietnamese, Hmong, Korean, and Chinese students. This study focuses on students within Sunnyside School District’s ten intermediate schools (serving grades 7 and 8) during the academic year of 2016-17.

This district is a special setting for studying this policy because of their unique implementation decisions involving double dose math courses across a short time span. In 2008-09, the district implemented double math courses within their middle schools as an intervention for low performing students. In 2010-11, two schools decided to discontinue offering these courses and by 2013-14 these courses were almost entirely discontinued. In 2015-16, district administrators decided to reinstate the courses with significant structural changes and piloted the new version of the course in two schools. In 2016-17, the district expanded its offering of the new course across eight middle schools.

The district also made considerable structural changes to the way the courses were designed and offered from the first iteration to the second. Originally, the district implemented the courses using minimal structure. Upon reintroducing policy, district administrators made modifications that can be categorized broadly within these areas: student placement, curricular structure, space for teacher sensemaking and collaboration, and overall goals for the intervention. To aid the reader, these categories and the main modifications from the original course to the reintroduced course are summarized in Table 3.1.

Within the first policy implementation, all students that did not qualify as proficient in math skills by sixth grade were recommended for enrollment in the intervention. Additionally, some students who transferred to the district and did not have full standardized test data for their placement were recommended to enroll in doubled math instruction math and then get re-evaluated by their math teacher at the semester break. Under the new placement guidelines, students at the lowest and highest ends of the below-proficiency achievement spectrum were recommended to other support programs, as were students with poor citizenship grades on their report cards. The district also encouraged school sites to only admit transfer students who they could identify as a good fit for the course, although some sites were less cooperative with this recommendation than others.

Previously, the district provided only curricular resources to teachers without set guidelines, pacing instructions, or planned units of material to cover. In the newer version of the course, considerably more curricular structure was provided. Teachers are still allowed a degree of autonomy in selection of activities, daily schedules, and instructional methods; however, the new course follows planned units based on current math standards and timed in conjunction with the students' traditional math course. The district also implemented a weekly activity allotment to ensure diversified instruction, including guidelines for days spent on traditional lecture, computer-based learning, and project-based learning.

District administrators also felt that that the previous iteration of the course lacked space for double dose math teachers to collaborate and workshop problems about teaching the course. District-organized professional development sessions, which were optional and infrequent in the previous iteration, were formalized to be used for this purpose in the new iteration. Now, professional development sessions are scheduled consistently and feature large portions of time dedicated to sharing best practices, working through common obstacles, and brainstorming solutions to common problems.

Finally, in the transition between the two iterations of the intervention, the district modified its vision and overall goals for how this intervention should be supporting students. Previously, the district viewed the course like a one-time boost to ensure students were ready for Algebra. The new goals for the course are described to include outcomes outside the realm of standardized math assessment scores and Algebra readiness. Confidence and new skills in math are a large part of the goal, but district administrators also see this course as part of the larger picture of improving graduation rates, self-confidence, and student motivation. However, no specific data was collected to measure these immediate outcomes, and time will have to pass to estimate any differences in long-term outcomes, so it is currently impossible to test any outcomes besides math achievement and grades.

Data & Sample

This study is based on administrative data provided by the school district. These data, which the district collects and reports for accountability purposes, tracks student enrollment from elementary through middle school and includes: student demographics (gender, race/ethnicity, language status, free/reduced price lunch eligibility); annual Smarter Balance Assessment Consortium (SBAC) standardized test scores; scores from district benchmark assessments; and transcript data on middle school math course, teacher ID, grade, and course period. I used student-level data for all seventh-grade students enrolled in the 2016-17 academic year. Double dose courses were also offered to eighth-grade students during this academic year, but participating schools offered eighth-grade intervention courses less frequently and some students were exposed to previous treatment within a pilot version of the course in seventh grade. Therefore, I did not include eighth-grade students in any analyses.

Certain students and schools were excluded from the sample in the primary analyses. Schools in the district only offered double dose courses as a complement to Pre-Algebra classes in seventh grade. Those students who had already completed Pre-Algebra or were specifically placed into special education or other intervention classes at a lower level than Pre-Algebra were excluded from the sample because they were ineligible to take the class. Additionally, two out of ten middle schools within the district did not offer the double dose course. After all exclusions, 1522 students from the 2016-17 academic year are present in the sample.

Table 3.2 displays the sample size, number of participating middle schools, demographic characteristics, and percentage of students enrolled within a double dose math course during these years for observations that remained in the sample. The students within the sample are over 70% Hispanic/Latino, close to 20% Asian or Asian-American, and more than 75% qualify for free- or reduced-price lunch. More than 40% of the students within the sample are English

Language Learners and over 8% are designated as students with a disability. The demographic characteristics of the sample, those students eligible for the intervention and attending a school that participated in offering it, compare closely to those of the district as a whole. The sample is by no means state- or nationally-representative; the sample schools enroll a disproportionately large number of Latino and Asian-American students and a correspondingly small number of white and African-American students in comparison with the population of the United States.

MEASURES AND METHODS

Measures

For these analyses, the three main outcomes are the Smarter Balance Assessment Consortium (SBAC) annual standardized math test score from the end of seventh grade, the final grade (in grade point average points) from the students' main (non-intervention) seventh-grade math course, and an indicator of a passing grade in that main math course. The SBAC assessment is administered by the state of California and possible values range from 2000 to 3000. The district's grade point average point system follows a traditional high school scale in which an A is coded as 4 points, an A- is coded as 3.7 points, a B+ is coded as 3 points, a B equals 3 points, a B- is 2.7 points, a C= is 2.3 points, a C is 2 points, a C- is 1.7 points, a D+ is 1.3 points, a D is 1 point, a D- is .7 points, and an F is coded as 0 points. Any grade besides an F is considered a passing grade within the district. The treatment of being enrolled in a double dose math course is coded with dummy variable indicating whether the student received double dose math instruction at any point during the school year.

Student-level data provided by the school district identify the gender and race/ethnicity of each student in the sample. The data also include whether each student was designated as an

English Language Learner, if they qualified for free- or reduced-price lunch, if they were diagnosed with a learning disability, and whether they maintained satisfactory marks in classroom behavior in the previous year. These four variables were converted to dummy variables with a value of 1 if the student held the designation and 0 if they did not.

The school district also provided four measures of prior math achievement from the previous school year, including the SBAC standardized math test score in sixth grade and three math benchmark assessments administered by the district. The SBAC assessment is administered by the state of California and possible values range from 2000 to 3000. The three benchmark assessments were administered in sixth-grade math classrooms in October, February and April of the previous academic year. Possible scores on all three benchmark assessments range from 0 to 100.

Methods

Propensity Score Weighting Design

Table 3.3 displays the demographic characteristics of students who were and were not enrolled in the double dose math course from the sample. Asian students were less likely to be enrolled in the intervention across the analysis sample while White, English Learner, and better-behaved students were more likely to be enrolled. These differences indicate patterns of selection into treatment that could bias any estimates obtained from a basic treatment-control comparison. To accurately estimate the impact of enrollment in the double dose math course, I utilized a propensity score weighting design using inverse-probability-of-treatment weights (IPTW) to account for selection into treatment (Murnane & Willett, 2011; Imbens & Wooldridge, 2009; Wooldridge, 2007). As long as propensity scores are estimated correctly, using IPTWs to balance treatment and control groups is more accurate than propensity score matching for counteracting selection bias and maintaining sample size (Murnane & Willett, 2011; Imbens & Wooldridge, 2009).

I followed Imbens and Wooldridge's (2009, Equation 18, p.35) instructions to calculate an IPTW for each student within the sample. First, I used a logistic regression model to calculate the likelihood of being enrolled in the intervention for each individual student based on their gender, race/ethnicity, free/reduced lunch status, English language designation, disability status, classroom behavior qualification, their previous standardized math score test, and all three previous math benchmark assessments. Figure 3.1 displays the common support in propensity scores among treated (grey) and non-treated students (red) within the sample. Then I use this probability value to calculate the IPTW for each individual student, such that the weight is equal to one over the probability of treatment for treatment students (1/p^) and one over one minus the probability of treatment (1/(1-p^)) for control students.

After using the IPTWs to the weight individual cases within the treatment and control groups, Table 3.4 displays the balance in student characteristics. Between the two groups, there are no discernible differences in demographic characteristic percentages once the weights are applied. To further check my weighting strategy, I also examined higher order sample balance by comparing the ratio of standard deviations between treated and non-treatment students for each of these variables before and after applying the weights (e.g. Hill, 2008). Table 3.5 displays these ratios, where a ratio closer to 1 indicates better balance. Before weighting, the ratios ranged from .60 to 1.23 and after weighting the range tightened to .82 to 1.11. Additionally, each ratio was equal in distance or closer to 1 in the weighted sample versus the unweighted sample except the ratios for Hispanic, other race, and learning disability (decreases in absolute value of .02, .09, and .11 respectively).

Then I used separate ordinary least square (OLS) regression models that incorporate the inverse probability treatment weighting of treatment and control students to test the effect of enrollment in the double dose math intervention on my three outcomes of interest. These propensity-score-weighted regression models are considered doubly robust, as they also include statistical controls for every variable included in the probability estimation:

$$Outcome_i = \beta_0 + \beta_1(Treatment)_i + \delta' X_i + \varepsilon_i \qquad (1)$$

In this equation, $Outcome_i$ reflects the outcome of interest for individual i in seventh grade, $Treatment_i$ is an indicator of enrollment in the double dose math course, X_i captures the vector of demographic and prior achievement variables used to calculate the propensity scores and then again as covariates, ε_i is the error term, and all cases are weighted by their individual IPTW. The formula above is the base layer model (Model 1) in all three of my main analyses. I also estimated versions with variables identifying the interaction between treatment and the indicator for English Learner status (Model 2), the indicator for free/reduced lunch eligibility (Model 3), and dummy variables for quartiles of performance on prior math achievement on the SBAC math standardized assessment (Model 4), and then all three of the interactions together (Model 5) for both outcomes.

RESULTS

Tables 3.6 through 3.8 present the results that pertain to the first two research questions of this study. These models test whether this version of double dose math courses was a successful intervention for low-performing seventh graders in this school district and if the intervention courses were uniquely impactful for socioeconomically disadvantaged students, students designated as English Learners, or students at different points along the prior math achievement spectrum. Tables 3.6 and 3.7 display coefficients using a standardized version of the outcome variable and Table 3.8 presents the findings in the terms of odds ratios. For Models 4 and 5 in all three tables, the grouping of highest performing students still eligible for the intervention is used as the reference category for the estimates of the prior achievement interactions to be consistent with my analysis in Chapter 1.

Table 3.6 displays the estimates for models predicting the standardized value of an individual's end-of-year standardized SBAC math score. While accounting for selection into treatment by weighting treatment and control cases using IPTWs, Model 1 indicates a significant association between enrollment in a double dose math course and math assessment score at the end of that year ($\beta = .11, p<.01$). Models 2 and 3, which test interactions between treatment in the course and student characteristics, do not yield significant results. Enrollment in the course appears to be slightly less beneficial for English Language Learners and slightly more beneficial for socioeconomically disadvantaged students in this sample, but these relationships are not statistically significant. Model 4 indicates that students within the middle groupings of prior achievement experience significantly lower impacts from the program than students in the top grouping ($\beta = -.27, p<.05$ for the second-lowest group, $\beta = -.25, p<.05$ for the third-lowest group). In the final Model, where all interaction terms are included together, students within the second-lowest grouping of prior achievement still experience significantly lower impacts from the intervention than students in the top grouping ($\beta = -.27, p<.05$) while all other interaction terms are non-significant.

Table 3.7 displays the estimates for identical models predicting the unstandardized value of grade point average points for the students' traditional (non-intervention) math course. These models also account for selection into treatment by weighting treatment and control cases using IPTWs. These estimates are similar to those in Table 3.6 but less pronounced and no estimates achieve statistical significance. Model 1 indicates a non-significant but positive relationship between enrollment in a double dose math course and end-year grade. Models 2 through 5, testing that same interactions used for the previous outcome, do not yield significant results. Again, Enrollment in the course appears to be slightly more beneficial for socioeconomically disadvantaged students and slightly less beneficial for English Language Learners and those with higher prior math achievement in this sample, but these relationships are not statistically significant and cannot be interpreted as associations. The same interpretation can be used for students in the lowest tier of prior math achievement, where a positive impact is possible but not statistically detectable.

Table 3.8 displays the estimates of the same set of propensity-score-weighted logistic regression models predicting the odds ratios of passing the traditional 7th grade math course, and there are no significant differences between the intervention and non-intervention students across each iteration. Overall, I find a .36 increase in the odds of passing the traditional math course for individuals within the double dose math course intervention, however this coefficient is just shy of reaching statistical significance. Statistical power is problematically low in all of these models. This is partially due to sample size and partially due to a low overall number of students that fail the course, Results in models 2-5, including interaction terms for specific student populations within the intervention, mirror the results from end-of-year math grade models but are not close to reaching statistical significance.

In the third research question for this study, I asked how the results for this revamped version of the course compare with those of the original implementation. In Chapter 1, I found that the extra course has no significant impact on students' end-year test scores, grades in their main math course, or likelihood of passing their main math course for treatment and control students within a tight bandwidth around the assignment cutoff. Using a more inclusive model, I found an overall intent-to-treat (ITT) effect of a 13.28-point increase (effect size = .24) for treatment students on the seventh-grade standardized test score *($p<.05$)* but no significant differences for math course grades or math course pass rates. Correlational models showed that test score increases were possibly concentrated among lower achieving math students and that treatment for English Language Learners was positively related to higher main math course grades (+.19 GPA points, $p<.05$).

In the second study, I was not able to use identical methods to the first and there are notable differences in the student sample between the two offerings of the course, but I reached similar results. Using a propensity score weighting design, I found an average treatment effect of +9.3 math test score points (effect size = .11, $p<.01$), but no significant impact on traditional math course grade or pass rates. I did not find any differential impacts for English Learners, as I uncovered in Chapter 1, or socioeconomically disadvantaged students. However, my findings on students at different levels of the prior achievement spectrum were flipped. In Chapter 1, I found the intervention to be more helpful for students at the lowest end of the prior achievement spectrum. In this study, the opposite occurred as I found negative associations for the middle groupings of students in comparison to those in the highest grouping of prior achievement. I will further interpret how these studies compare in the Discussion section.

DISCUSSION

This study is the first of its kind within the literature on double does math course interventions, and unique to studies on education programs in general, as a comparison of two different implementations of a double dose math course intervention within the same setting. Overall, the findings corroborate previous studies that have found modest, positive effects on standardized test math scores when students take an additional math period during the school year. For this population, a targeted group of seventh-grade students within midsized, suburban school district, I did not see these benefits carry over to improved grades or pass rates in their main (non-intervention) math course. I also failed to find, as previous studies including my first study have suggested, evidence that these courses were especially impactful for students designated as English Learners.

Additionally, the intricacies of comparing the studies of the two versions of the intervention offer insights about its potential impact in different environments. Despite considerable changes to how the course was offered, both analyses yielded similarly modest, overall positive relationships between taking the course and increased standardized test scores but not math course grades or passage rates. However, while the first version of the intervention was recommended to any student that was below proficiency standards at the end of sixth grade, the new version of the course was targeted for a more concentrated group of students. The second implementation excluded students who scored the lowest on the placement assessments (those who were shown in Chapter 1 to significantly benefit from the course), students who fell just below the proficiency cutoff (those who were shown in Chapter 1 to not benefit from the intervention), and students within the targeted ability levels that exhibited poor classroom behavior.

After these changes in the course and the targeted student population, the intervention was related to significantly higher test scores for students who fell in the top-quartile of prior math achievement, rather than the lowest levels or prior math achievement as I found before. One possible explanation is that the new, more structured version of the course was more suited to the skills of students with more previous success in math, whereas the less structured version was more helpful to students who needed math support of any kind. It also possible that this implementation of the intervention excluded some lower-performing students that previously improved more because of the course (as seen in Chapter 1). In Table 3.6, we see that the change in standardized test score for the lowest group of prior achievement is still similar to the highest group but not quite large enough to be significantly different than the middle groups. It would have been most useful to perform a direct comparison with the same student population across these two studies; whether these implementation changes would have been more or less impactful for excluded students that previously benefitted from the course is unknown.

Overall, a more structured version of the intervention that was offered to a more targeted group of students seemed to yield slightly more universal impacts; I did not see specific benefits for English learners as suggested by prior studies. Yet, I was surprised that the implementation changes made by the district (standardized curriculum and resources, professional development and space for teacher collaboration) did not largely alter the intervention's impact. The structural changes to the course did not seem to yield great changes in student outcomes, and to date the only structural aspect of this program that has ever been shown to impact intervention effectiveness for treatment students is matching groups of students with the same teacher for both the traditional and intervention courses (Nomi & Allensworth, 2013).

From the comparison standpoint, it should also be noted that the analyses in Chapter 1 cover a five-year span of offering the course while this analysis only covers the first year of the new implementation. There is some overlap in teachers across implementations, however the majority of teachers in the second version have no prior exposure to teaching the course and did not have the opportunity to gain up to five years of course teaching experience along the way. It is possible that the impacts of the two versions could become more differentiated if future years of the new implementation could be incorporated in the analyses. This would make for an intriguing follow-up study to this analysis. Future studies should also be designed to include outcomes outside of math achievement, such as measures of confidence and perceived ability in mathematics, as district administrators believe the intervention might lead to non-cognitive impacts as well.

STUDY LIMITATIONS

The possibilities for modeling decisions, especially any establishing causality, were limited based on the data available and the sample size. The estimates from all models are correlational in nature and not to be interpreted as causal relationships. Ideally, I wanted to better match the sample and study design from Chapter 1 in these analyses, but that was not possible with the available data. Finally, these analyses are focused on only one school district, so they are not generalizable outside of that population.

REFERENCES

Bartik, T., & Lachowska, M. (2014). *The effects of doubling instruction efforts on middle school students' achievement: Evidence from a multiyear regression-discontinuity design* (No. 14-205). Kalamazoo, MI: WE Upjohn Institute for Employment Research.

Cortes, K. E., Goodman, J. S., & Nomi, T. (2015). Intensive math instruction and educational attainment: Long-run impacts of double-dose Algebra. *Journal of Human Resources*, *50*(1), 108-158.

Durwood, C., Krone, E., & Mazzeo, C. (2010). *Are two Algebra classes better than one? The effects of double-dose instruction in Chicago.* Chicago, IL: Consortium on Chicago School Research.

Fryer, R. G. (2012). *Injecting successful charter school strategies into traditional public schools: Early results from an experiment in Houston*. Cambridge, MA: National Bureau of Economic Research.

Henry, G. T., Barrett, N., & Marder, C. (2016). *Double-dosing" in math in North Carolina public schools* (REL 2016–140). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southeast. Retrieved from http://ies.ed.gov/ncee/edlabs.

Imbens, G. W., & Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, *47*(1), 5-86.

Hill, J. (2008). Discussion of research using propensity-score matching: Comments on "A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003" by Peter Austin. *Statistics in Medicine*, *27*(12), 2055-2061.

Murnane, R. J., & Willett, J. B. (2010). *Methods matter: Improving causal inference in educational and social science research*. Oxford University Press.

Nomi, T., & Allensworth, E. (2009). "Double-Dose" Algebra as an alternative strategy to remediation: Effects on students' academic outcomes. *Journal of Research on Educational Effectiveness*, *2*(2), 111-148.

Nomi, T., & Allensworth, E. M. (2013). Sorting and supporting: Why double-dose algebra led to better test scores but more course failures. *American Educational Research Journal*, *50*(4), 756-788.

Reys, R., Reys, B., Lapan, R., Holliday, G., & Wasman, D. (2003). Assessing the impact of" standards"-based middle grades mathematics curriculum materials on student achievement. *Journal for Research in Mathematics Education*, 74-95.

Tarr, J. E., Reys, R. E., Reys, B. J., Chavez, O., Shih, J., & Osterlind, S. J. (2008). The impact of middle-grades mathematics curricula and the classroom learning environment on student achievement. *Journal for Research in Mathematics Education*, 247-280.

Taylor, E. (2014). Spending more of the school day in math class: Evidence from a regression discontinuity in middle school. *Journal of Public Economics*, *117*, 162-181.

Wooldridge, J. M. (2007). Inverse probability weighted estimation for general missing data problems. *Journal of Econometrics*, *141*(2), 1281-1301.

CHAPTER 3 TABLES & FIGURES

Table 3.1
Key revisions to intervention structure

Intervention Structure	Previous implementation	New implementation
Student placement	Students scoring 3-8 points on a 3-15 scaled based on three prior assessments	Students scoring 4-7 points on a 3-15 scaled based on three prior assessments; Students must also have satisfactory classroom behavior grades in the prior year
Curricular Structure	Instructional resources offered but no structured curriculum or guidance on specific instruction; optional professional development sessions on learning activities	Curriculum coordinated with traditional math course; guidance on weekly variation of instruction techniques; Instructional resources offered; professional development sessions (some mandatory, some optional) on structured curriculum
Space for Teacher Sensemaking and Collaboration	Optional professional development sessions	Professional development sessions (some mandatory, some optional); site visits and individual coaching with district staff members
Stated Goals for the Intervention	Improving math achievement and preparedness for Algebra	Improving math achievement and confidence in math abilities; Improving larger-scale outcomes such as graduation rates, academic self-confidence, and student motivation

Table 3.2

Summary statistics of analysis sample (n=1522)

	Observations	**Mean**	**SD**	**Min**	**Max**
Student demographics					
% Female	1522	50.2	-	-	-
% African American	1515	0.9	-	-	-
% Asian	1515	19.2	-	-	-
% Hispanic or Latino	1515	70.3	-	-	-
% White	1515	8.3	-	-	-
% Other Race	1515	1.3			
% Free- and Reduced-Price Lunch	1522	75.4	-	-	-
% English Language Learners	1522	42.9	-	-	-
% Students with Disability	1522	8.3	-	-	-
Prior math achievement					
6^{th} grade Mathematics state standardized test score	1522	2506.6	71.8	2235	2697
6^{th} grade 1^{st} benchmark test score	1476	56.4	15.7	6.1	100
6^{th} grade 2^{nd} benchmark test score	1492	57.0	16.9	0	100
6^{th} grade 3^{rd} benchmark test score	1501	52.3	16.3	0	95.7
Double dose math course enrollment					
Met behavior requirement for recommendation to enroll	1522	78.2	-	-	-
% recommended to enroll in double dose course during 7^{th} grade	1522	42.5	-	-	-
% enrolled in double dose course during 7^{th} grade	1522	17.2	-	-	-

Note. Students were included in the analysis sample if they attended a school that offered the course, enrolled in the appropriate main math course for the intervention, and had a valid outcome score for 7^{th} grade mathematics standardized test score.

Table 3.3
Selection into double dose math course enrollment

Characteristic	Not enrolled (n=1260)	Enrolled (n=262)	Statistically significant difference?
Female	51%	48%	No
Asian	20%	13%	Yes**
Black	1%	.3%	No
Hispanic	70%	73%	No
White	8%	12%	Yes*
Other race	1%	2%	No
Free/Reduced Lunch	76%	74%	No
English Learner	40%	55%	Yes***
Learning Disabled	8%	8%	No
Classroom Behavior Qualification	76%	87%	Yes***

Note. Two-tailed test of statistical significance. * p<.05 ** p<.01 *** p<.001

Table 3.4
Demographics balance after propensity score weighting

Characteristic	Not enrolled (n=1260)	Enrolled (n=262)	Statistically significant difference?
Female	50%	48%	No
Asian	19%	15%	No
Black	0.9%	0.8%	No
Hispanic	71%	74%	No
White	8%	9%	No
Other race	1.2%	0.9%	No
Free/Reduced Lunch	75%	74%	No
English Learner	43%	44%	No
Learning Disabled	8%	10%	No
Classroom Behavior Qualification	78%	77%	No

Note. Two-tailed test of statistical significance.

Table 3.5

Ratio of standard deviations for demographic variables before and after propensity score weighting

Characteristic	Sample	Not enrolled (n=1260)	Enrolled (n=262)	Ratio of standard deviations (Enrolled : Not enrolled)
Female	Unweighted	.50	.50	1.00
	Weighted	.50	.50	1.00
Asian	Unweighted	.40	.34	.85
	Weighted	.39	.36	.92
Black	Unweighted	.10	.06	.60
	Weighted	.10	.09	.90
Hispanic	Unweighted	.46	.45	.98
	Weighted	.46	.44	.96
White	Unweighted	.26	.32	1.23
	Weighted	.28	.28	1.00
Other race	Unweighted	.11	.12	1.09
	Weighted	.11	.09	.82
Free/Reduced Lunch	Unweighted	.43	.44	1.02
	Weighted	.43	.44	1.02
English Learner	Unweighted	.49	.50	1.02
	Weighted	.50	.50	1.00
Learning Disabled	Unweighted	.28	.28	1.00
	Weighted	.28	.31	1.11
Classroom Behavior	Unweighted	.43	.33	.77
Qualification	Weighted	.41	.42	1.02

Table 3.6

Regression models after propensity score weighting, predicting end of year math score

	M1 End-year standardized math score	M2 End-year standardized math score	M3 End-year standardized math score	M4 End-year standardized math score	M5 End-year standardized math score
Double Dose Course	.11** (.05)	.14* (.06)	.08 (.06)	.29** (.09)	.26** (.10)
Double Dose x English Learner Status	-	-.05 (.09)	-	-	-.09 (.09)
Double Dose x F/R Lunch eligible	-	-	.05 (.09)	-	.09 (.09)
Double Dose x Prior Achievement Group 1	-	-	-	-.07 (.14)	-.00 (.06)
Double Dose x Prior Achievement Group 2	-	-	-	-.27* (.12)	-.27** (.12)
Double Dose x Prior Achievement Group 3	-	-	-	-.25* (.12)	-.24 (.12)
Double Dose x Prior Achievement Group 4	-	-	-	(reference)	(reference)
Observations	1522	1522	1522	1522	1522

Notes. Outcome variable is standardized. Standard errors in parentheses. Treatment and control students are weighted using IPTWs. Missing data is handled using a missing value dummy variable adjustment to maintain sample size. Prior achievement groups are based on the previous end-year math standardized math score, with the lowest scoring quartile of students placed in Group 1, second lowest in Group 2, second highest in Group 3, and highest scoring quartile of students in Group 4. All models include controls for all variables used to predict the propensity score, including previous Math Benchmark assessments 1 -3, the previous end-year math standardized math score, Classroom behavior, Gender, Race, English Learner status, Disability status, Free/reduced lunch status, and Learning disability status.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3.7

Regression models after propensity score weighting, predicting end of year math grade in traditional math course

	M1 Traditional Math Course GPA points	M2 Traditional Math Course GPA points	M3 Traditional Math Course GPA points	M4 Traditional Math Course GPA points	M5 Traditional Math Course GPA points
Double Dose Course	.13 (.08)	.11 (.11)	.27 (.14)	.20 (.21)	.29 (.24)
Double Dose x English Learner Status	-	.06 (.16)	-	-	.03 (.17)
Double Dose x F/R Lunch eligible	-	-	-.19 (.17)	-	-.20 (.18)
Double Dose x Prior Achievement Group 1	-	-	-	.18 (.25)	.26 (.30)
Double Dose x Prior Achievement Group 2	-	-	-	-.19 (.26)	-.15 (.29)
Double Dose x Prior Achievement Group 3	-	-	-	-.07 (.27)	-.01 (.29)
Double Dose x Prior Achievement Group 4	-	-	-	(reference)	(reference)
Observations	1522	1522	1522	1522	1522

Notes. Outcome variable is unstandardized. Standard errors in parentheses. Treatment and control students are weighted using IPTWs. Missing data is handled using a missing value dummy variable adjustment to maintain sample size. Prior achievement groups are based on the previous end-year math standardized math score, with the lowest scoring quartile of students placed in Group 1, second lowest in Group 2, second highest in Group 3, and highest scoring quartile of students in Group 4. All models include controls for all variables used to predict the propensity score, including previous Math Benchmark assessments 1 -3, the previous end-year math standardized math score, Classroom behavior, Gender, Race, English Learner status, Disability status, Free/reduced lunch status, and Learning disability status.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3.8

Logistic regression models after propensity score weighting, predicting passing math grade in traditional match course

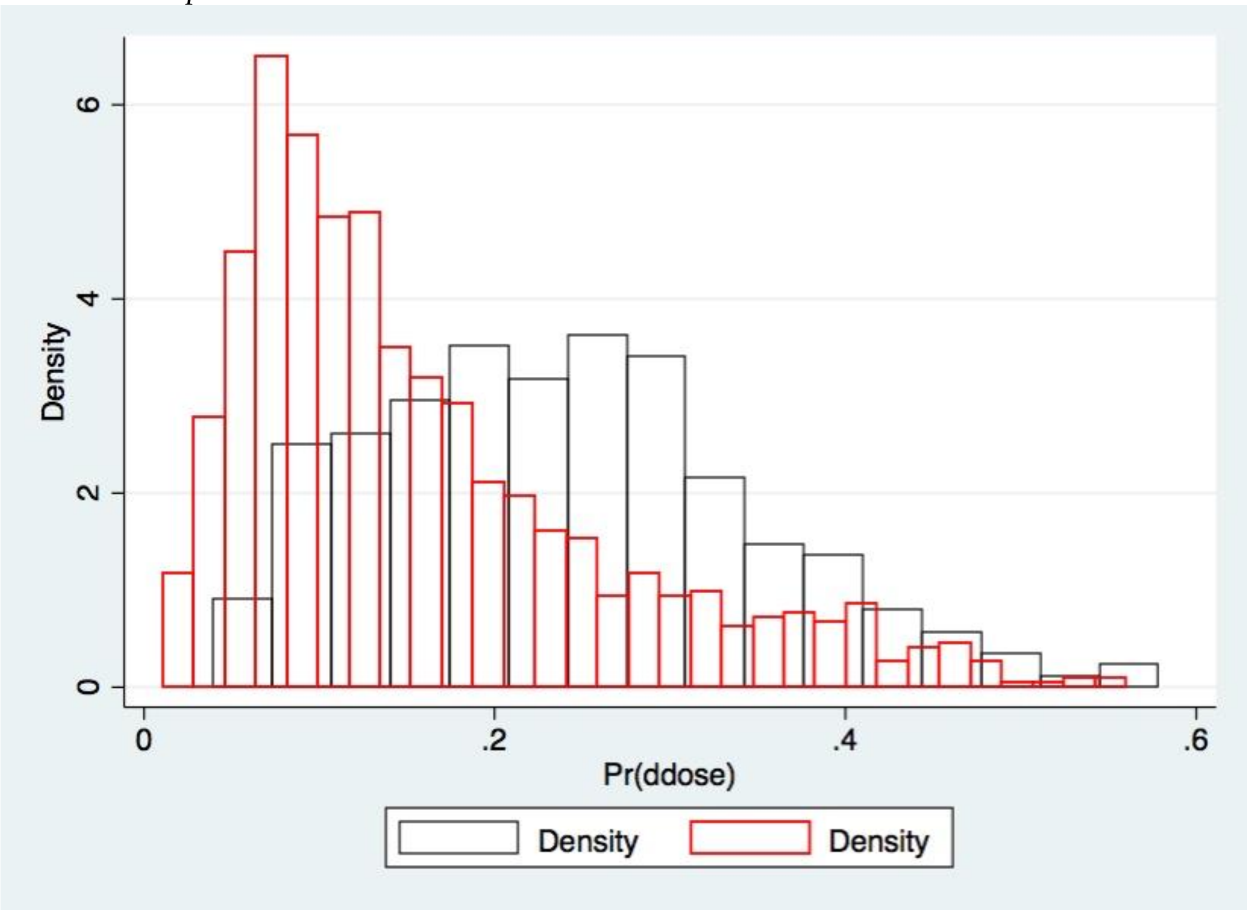
	M1 Traditional Math Course Pass	M2 Traditional Math Course Pass	M3 Traditional Math Course Pass	M4 Traditional Math Course Pass	M5 Traditional Math Course Pass
Double Dose Course	.36 (.20)	.13 (.30)	.82 (.52)	-.12 (.71)	-.08 (.82)
Double Dose x English Learner Status	-	.43 (.40)	-		.42 (.43)
Double Dose x F/R Lunch eligible	-	-	-.55 (.57)		-.66 (.62)
Double Dose x Prior Achievement Group 1	-	-	-	.63 (.78)	.95 (.90)
Double Dose x Prior Achievement Group 2	-	-	-	.65 (.80)	.87 (.92)
Double Dose x Prior Achievement Group 3	-	-	-	.24 (.83)	.70 (.94)
Double Dose x Prior Achievement Group 4	-	-	-	(reference)	(reference)
Observations	1522	1522	1522	1522	1522

Note. Standard errors in parentheses. The outcome variable is a [1,0] indicator of receiving a passing grade. Missing data is handled using a missing value dummy variable adjustment to maintain sample size. Prior achievement groups are based on the previous end-year math standardized math score, with the lowest scoring quartile of students placed in Group 1, second lowest in Group 2, second highest in Group 3, and highest scoring quartile of students in Group 4. All models include controls for all variables used to predict the propensity score, including previous Math Benchmark assessments 1 -3, the previous end-year math standardized math score, Classroom behavior, Gender, Race, English Learner status, Disability status, Free/reduced lunch status, and Learning disability status.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Figure 3.1

Common support in propensity scores among treated (grey) and non-treated students (red) within the sample



CONCLUSIONS

Considered together, these three chapters offer points to consider about double dose math courses as an intervention for low-performing math students and, more generally, how districts enact, shape, and evaluate similar types of policies. Specific to this type of policy, Chapters 1 and 3 offer further evidence that offering a second instruction period of math is likely to help increase standardized test scores for enrolled students in that year of instruction (similar to Cortes, Goodman, & Nomi, 2015; Nomi & Allensworth, 2009; Taylor, 2014) although impacts on math course grades are still inconsistent. The samples studied featured only seventh-grade students, but considered alongside the three studies cited above, that result is consistent across varied student grade levels (sixth- through ninth-grade), math courses that are being complemented (Grade 6 Math, Grade 7 Math, Pre-Algebra, Algebra), study settings, and implementation structures. Still, the effect sizes of positive outcomes within these studies have been consistently modest and should be considered in comparison to other possible interventions or unconsidered ramifications of enrolling students in the course.

These analyses also provide further information for schools or districts implementing double dose math courses to support targeted subgroups of students. Specifically to English Learners and students with differing prior achievement in math within this setting, the impact of the course for these subgroups of students was sensitive to the structural changes made between the two iterations. Findings from Chapter 1 support the developing evidence that this type of intervention is especially useful for students who are learning the English language or low-skilled in English Language Arts. Schools or districts wishing to improve math performance among this student population might find success with this model.

Particular attention should be paid to the prior achievement of recommended students and the assignment strategy used to decide which students are recommended to enroll. In this district, it appeared that the first, less-structured version of the course was possibly more impactful for some of the lowest-performing students while the more targeted, structured version was more helpful to the highest-performing group of eligible students. Both versions seem to have positive effects for students falling below the mark of proficient on California state standardized assessments; this stands as a reasonable cutoff point for students that will benefit from this support. However, I did not find significant benefits of these courses for students who fall barely underneath the proficiency level. On the other hand, schools or districts adhering to an assignment strategy that excludes students in the lowest levels of prior achievement, as was seen in Chapter 3, might be denying access to students who can benefit the most.

Because a double dose math support course also usually occupies a class period for another subject, costs of taking the course should also be considered in future work. Taylor (2014) found no impact on treatment students' reading scores, and no significant decreases in course-taking in Physical Education, music, or arts; the only subject that seemed to be hampered by the intervention was foreign language courses, as double dose math students took significantly less courses in a foreign language. Taylor (2014) also proposed other costs to consider including teacher workload, school resources, and achievement in other subjects such as science or history but was unable to test them. It was not possible to study these types of outcomes within this dissertation either, but the idea that the modest improvements in math achievement from this intervention are balanced with other possible detriments is important to consider. Outcome measures such as a student's overall happiness, feelings of academic selfefficacy, and school attendance are other possible outcomes of interest for future studies on doubled math instruction.

Additionally, longer-term outcomes (if possible and applicable) and math-related socio-emotional outcomes should continue to be considered in future studies. In Chapter 2, district administrators referenced their hope that the revamped version of this intervention was contributing to improved graduation rates, greater self-confidence, and student motivation in future math courses. These outcomes, in addition to achievement in later years, offer the chance to identify lasting impacts from this program that were not possible to incorporate in these studies. From their study on double dose algebra classes in the Chicago Public Schools, Cortes, Goodman, & Nomi (2015) found that the intervention positively influenced students' short- and long-term math skills, future math course taking, and outcomes related to high school completion and college enrollment. However, no other studies have found lasting, long-term effects from this intervention. Future researchers and their collaborators should account for and incorporate more of these types of outcomes on top of traditional math achievement, as they reflect changes in the perception of the purpose of this type of intervention.

As a whole, these chapters also show how school district administrators carried out the continuous improvement of a district-created curricular policy. Mostly covered in Chapter 2, there are many examples of how the district was engaging in a thoughtful, committed attempt to improve a math intervention that was the best fit for their setting. In order to best improve the course, the district employed a team of lower-level district administrators that were mostly former teachers of this original version of the intervention to be directly responsible for designing changes and supporting teachers. These staff members were strongly committed to their charge and uniquely qualified for this special role. Their assignment showed a commitment of targeted, extra resources by the district, as they were pulled from other responsibilities to focus specifically on supporting this intervention.

Although one could argue the merits of the individual decisions made regarding this policy, this was also an example of a school district that was attempting to be responsive to the needs of their students in the face of changing state standards and assessments. District administrators consistently cited state standards and student test scores as drivers of major decisions about this policy, and the timing of their decisions corroborates this rationale. The district was making interpretations of these changing standards and the ramifications for their students within a relatively quick response time. Overall, these district staff members demonstrated a strong awareness of changes in large-scale policy that might impact their students and a willingness to update prior assumptions and decisions in accordance with these interpretations.

However, these chapters also present an important chain of events regarding district data usage and evidence that is considered within policy modifications. Specifically, there were many instances when specific data were not consulted or only included symbolically in making crucial decisions about the policy. There were also contradictions between my findings in the difference-in-difference-in-difference and OLS regression models in Chapter 1 and the district administrators' decision-making highlighted in Chapters 2 and 3. The school district cited lack of gains in test score performance for intervention students as a reason for discontinuing the first iteration of the policy. Yet, my findings show that the policy was at least modestly working toward its intended purpose at the time. Additionally, district administrators cited the same test score evidence as the reason for changing the student placement policy within the second iteration to exclude students at the lowest end of the prior math achievement spectrum. My analyses from Chapter 1 show the policy was possibly helping these students more than other groupings based on prior achievement.

Within Chapter 2, I uncovered other instances where the school district maintained thorough data and referenced this data in their decision-making process, but the data they chose to use was not determined beforehand and subject to bias. On the surface, these issues regarding data analysis and evidence-based decisions appeared to stem from a lack of internal capacity for supporting this work. However, I was not privy to information on the priorities and direction of the district in these areas. Either way, in an era where school districts have growing amounts of data to choose from in their evaluation of curricular policies, this strikes a warning bell for the importance of proper evaluation methods and pre-determined, theory-based metrics of policy effectiveness that are not subject to useful interpretations. As school district's familiarity with collecting and analyzing data continues to grow, there are great opportunities for school districts to increase their capacity to do this research internally or partner with outside groups, such as university researchers or consultants, to provide thorough, unbiased interpretations to inform policy decisions.