Departmentalized Instruction and Elementary School Effectiveness

Ben Backes
James Cowan
Dan Goldhaber

April 2024

WORKING PAPER No. 298-0424

CALDER
National Center for Analysis of Longitudinal Data in Education Research

AIR®

Departmentalized Instruction and Elementary School Effectiveness

Ben Backes
American Institutes for Research / CALDER

James Cowan
American Institutes for Research / CALDER

Dan Goldhaber
American Institutes for Research / CALDER
University of Washington

Contents

Acknowledgments

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A210008 to American Institutes for Research (AIR). The opinions expressed are those of the authors and do not represent the views of the Institute, the U.S. Department of Education, or the Massachusetts Department of Elementary and Secondary Education. The authors thank Dana Ansel for conducting interviews with school principals as part of this study; Aubree Webb, Pierre Lucian, and participants at the 2023 Association for Education Finance and Policy Annual Conference for helpful comments; and Elana McDermott and the Massachusetts Department of Elementary and Secondary Education for making the data available for this study.

CALDER • American Institutes for Research
1400 Crystal Drive 10th Floor, Arlington, VA 22202
202-403-5796 • www.caldercenter.org

Departmentalized Instruction and Elementary School Effectiveness
Ben Backes, James Cowan, & Dan Goldhaber
CALDER Working Paper No. 298-0424
April 2024

Abstract

Departmentalized instruction, in which teachers specialize in one or more core subjects and instruct multiple groups of students in a day, has become increasingly prominent in elementary schools. Using 8 years of data from Massachusetts and a difference-in-differences design, we estimate the effects of departmentalization on student achievement. We find that departmentalization has positive effects in English language arts (ELA) and science and mixed evidence of positive effects in math. These positive effects are not driven by teacher productivity improvements: Consistent with prior findings on teacher specialization, teachers are less effective when specializing in math and no more effective in ELA than when teaching self-contained classrooms. Rather, consistent with the theoretical underpinnings for specialization, departmentalized schools tend to assign teachers to their stronger subjects.

1. Introduction

Teaching requires both specialized content knowledge in a field of study and the pedagogical skill to adapt instruction to the learning needs of individual students (Hill et al., 2005; Shavelson & Stern, 1981). The importance of domain-specific knowledge suggests students may benefit when their teachers specialize by subject area. However, assigning students to multiple teachers in different subjects may also impede their teachers' ability to target instruction to specific learning needs and thereby limit the potential gains from specialization. This trade-off between specialization and coordination frictions is a recurring theme in the labor economics literature (e.g., Becker & Murphy, 1992; Deissen & Santos, 2006) and leads to different organizational structures across grade levels. In middle and high schools, teachers tend to specialize in specific fields and organize into disciplinary departments. In elementary schools, teachers are typically generalists and instruct a self-contained class of students across multiple subjects. Relative to their colleagues in higher grades, elementary teachers more often tailor their instruction to a smaller group of students than specialize in a narrower set of subject areas.

There has been a long-standing debate over whether this is an optimal organization of elementary classrooms (Anderson, 1962; Chan & Jarman, 2004; Jacob & Rockoff, 2011). About 15% of elementary teachers in the United States work in departmentalized assignments, in which they instruct multiple classes in fewer subjects, with the proportion increasing by about two-thirds between 2003 and 2015 (National Center for Education Statistics [NCES], 2007, 2017, 2023). The theoretical benefit of content-area specialization in elementary schools arises from the divergence between teachers' general and subject-specific knowledge and pedagogical skills. Although general instructional skills are important, research has found that individual teachers vary in their proficiency in teaching different subjects (Condie et al., 2014; Goldhaber et al., 2013). Departmentalized models take advantage of this form of comparative advantage by allowing teachers to specialize and work only in their subjects of relative effectiveness. In schools where individual teachers vary in their effectiveness across subjects, trading instructional tasks across subjects might increase student achievement without changes in the teacher faculty.

Recent evidence, however, has called into question the theoretical benefits of teacher specialization. Self-contained classrooms pair students with a single teacher for all major subjects, which may be more appropriate for younger children (Brobst & Markworth, 2019). Working with fewer students may help teachers better identify students' needs, individually tailor instruction, or provide social-emotional support, effects which may offset the benefits of specialization. An experimental evaluation of departmentalization by Fryer (2018) found significant negative effects on student test achievement. Other studies have found that specialization reduces the productivity of individual elementary school teachers (Bastian & Fortner, 2020; Hwang & Kisida, 2022).

In this paper, we revisit the empirical debate on departmentalized instruction (DI) using a difference-in-differences (DID) design in Massachusetts elementary schools. Relative to the existing literature, our study makes two main contributions. First, we explore the effects of DI as undertaken voluntarily by schools at scale. The experimental evaluation described in Fryer (2018) randomly assigned schools to switch models, but also constrained their scope for planning and changing teacher assignments. We consider settings where schools have more latitude to adjust assignments and instructional teams to exploit teachers' comparative advantage. Second, the observational literature on DI has mainly focused on its effects on individual teacher productivity. Because a primary motivation is exploiting teachers' comparative advantage by trading instructional tasks, we focus on DI as an organizational model rather than a feature of teacher assignments, as in Bastian and Fortner (2020) and Hwang and Kisida (2022).

Our results suggest that DI may have more beneficial effects than indicated in previous studies. Switching to a departmentalized model increases student achievement in English language arts (ELA) by 0.03–0.07 standard deviations (SDs), math by 0.00–0.04 SDs, and science by 0.04–0.06 SDs. As in prior research, we find null or negative effects of specialization on the productivity of individual teachers (Bastian & Fortner, 2020; Hwang & Kisida, 2022). We complement this finding with survey evidence showing that students report worse participation and instructional environments under DI. These effects, however, are offset by departmentalized schools strategically assigning teachers to their stronger subjects. We find that the largest benefits of DI accrue to schools with greater potential for comparative advantage, as measured by teachers' subject-specific skills. By contrast, we find little evidence that increased specialization, as measured by the number of distinct subjects teachers instruct, leads to larger improvements in achievement in departmentalized schools. Overall, our results suggest greater potential for reassigning teachers by subject area to improve student learning than has been found previously.

2. Motivation and Literature Review

Although teachers proficient in one subject tend to perform well in others, part of teaching skill is subject-specific. For instance, at the elementary level, the correlation in teacher effects on student achievement across subjects is about 0.7-0.8 (Condie et al., 2014; Goldhaber et al., 2013). This observation has long been an argument for departmentalizing elementary classrooms (Anderson, 1962; Lobdell & van Ness, 1963; Otto, 1931). More recently, empirical studies about teacher effectiveness and its heterogeneity across subjects have led to renewed calls for DI (Jacob & Rockoff, 2011). New York City, for instance, explicitly pointed to these findings when it launched an elementary school DI initiative that was described as a "sweeping city experiment in overhauling how elementary school students are taught" (Zimmerman, 2018).

Critiques of the DI model typically focus on the potential that less time spent with each student reduces teachers' ability to form productive relationships and target instruction to student needs (Parker et al., 2017). Research on student-teacher interactions suggests the importance of teacher-student relationships that facilitate personalized, in-depth assessment of student learning and tailored instruction, and self-contained classrooms may be more conducive to building these kinds of relationships (Davis, 2003; Reeve, 2006; Thomas & Oldfather, 1997). This hypothesis is consistent with the research on the transition to middle schools, which are more frequently departmentalized. Middle school students report weaker relationships with teachers, less motivation in school, and increased social isolation (Anderman, 2003; Anderman & Maehr, 1994; Harter et al., 1992; Marks, 2000; Pellegrini & Bartini, 2000; Wentzel, 1997). Recent studies have also found that at the elementary level, students tend to fare better academically the second time they are assigned to a particular teacher (*looping*), which may reflect teachers' better knowledge of their students' academic needs (Hill & Jones, 2018; Wedenoja et al., 2022).

These findings suggest DI may involve important trade-offs between teacher specialization and student-teacher relationships. Yet there is only sparse empirical literature about the effects of DI. Fryer (2018) examines an experiment in which 46 elementary schools in Houston were randomized to either switch to DI or continue using self-contained classrooms. He estimates that DI reduced student achievement by an average of about 0.05 SDs across both years of the experiment.[1] Consistent with common critiques of DI, teachers in departmentalized schools were less likely to report knowing their students or giving them individualized attention.

Although it provides strong evidence for negative effects of DI in an experimental condition, there are two reasons to question whether these findings generalize to more routine

[1] Fryer (2018) reports results on an index of math and ELA test scores that sums student test results across both subjects.

settings. First, the experimental protocol placed some restrictions on the extent of teacher specialization within schools; for instance, teachers were required to be reassigned within grade levels, which rules out assignments in which a teacher instructs a single subject to multiple grade levels. Second, implementation occurred over just a 3-week period preceding the initial school year. During this time, the district randomized schools to treatment assignment, first notified principals of the existence of the experiment, and requested that principals in treated schools reassign teachers in consultation with the research team. Proposed teacher reassignments were not submitted until 4 days before the school year, and teachers were not notified of their new assignments until they returned to work for the new year.[2] This timeline may limit the organizational changes that typically coincide with DI. Several policy or qualitative studies suggest that switches to DI include significant planning prior to implementation, including pilot trials in one or more subjects or grade levels; the formation of new instructional teams to coordinate instruction across classrooms; and changes to student assignment policies (Chan & Jarman, 2004; Haley, III, 2018; Parker et al., 2017; Strohl et al., 2014). These changes may be difficult to implement in the limited time available to schools between randomization and the beginning of the school year.

Two other studies (Bastian & Fortner, 2020; Hwang & Kisida, 2022) assess the effects of teacher specialization (but not DI directly) at the elementary school level. Each paper uses teacher fixed-effects models to examine the extent to which teacher specialization affects the productivity of individual teachers. They define teachers as specialized when they teach two or fewer of the four core academic subjects (ELA, math, science, and social studies). Both studies find that teachers are less effective when they specialize: Working as a specialist reduces value

[2] Principals were informed of the experiment—and their randomization status—on August 8, 2013. Schools switching to DI then submitted proposed teacher assignments to the district for approval on August 22, 2013.

added (VA) to math test scores by 0.04 SDs and to reading test scores by 0.01 SDs. These negative effects do not dissipate over time as teachers gain experience as specialists. Both studies then assess the association between school achievement and the percentage of teachers who are specialized and find no relationship outside of science.

The studies described above are well-designed to assess the effects of teacher specialization on the productivity of individual teachers and provide credible evidence about important trade-offs for DI. But it is important to recognize that DI is motivated by changing teacher assignments so that teachers work in their stronger subject(s). These effects would not be captured by examining within-teacher variation in specialization, a potentially important channel through which DI might affect student outcomes. Indeed, the DI calibration exercises in Condie et al. (2014) and Goldhaber et al. (2013) are based entirely on changes to teacher assignments and not on the potential for specific human capital accumulation, as in Ost (2014) or Blazar (2015). In this study, we rely on statewide data and examine changes in student outcomes following switches in instructional models.

3. Data

A. Student Data

The sample includes all students enrolled in grades 3–6 in elementary schools in Massachusetts between 2012 and 2019 as well as their teachers in the four core academic subjects (ELA, math, science, and social studies). The administrative data include student demographic information (race, gender, special education status, English language learner status, and eligibility for free and reduced-price lunch programs), and math and ELA end-of-grade test results in grades 3–6 and science tests in grade 5.[3] We use these tests as the core academic outcomes.

We also use student survey data to help assess the implications of DI for student engagement and for students' relationships with teachers. Students' perceptions of their learning environment come from the Views of Climate and Learning (VOCAL) survey (Backes et al., 2022; Massachusetts Department of Elementary and Secondary Education [DESE], 2019). VOCAL is administered with end-of-grade testing in Massachusetts in grade 5 in 2018 and grades 4–5 in 2019. The surveys elicit students' views on three dimensions of school climate: engagement, safety, and environment. While intended to measure overall school climate, the surveys include several questions about student-teacher relationships that could plausibly be affected by school instructional models. For example, the survey asks students about the extent to which they feel a social connection with teachers, about the extent to which adults support the emotional needs of students, and about whether the environment is supportive of learning.

We complement the student survey data with administrative data on student attendance, suspensions, and grade promotion to assess student motivation and engagement (Jackson, 2018; Ladd & Sorensen, 2017). Following Jackson (2018), we also construct a nontest composite index using an exploratory factor analysis. We combine log absences, days suspended, and an indicator for grade promotion to produce a single behavioral factor. Teacher effects on this measure predict their effects on high school completion and college enrollment (Backes et al., 2023).

[3] Since 2012, Massachusetts has implemented multiple versions of the state assessment—the Massachusetts Comprehensive Assessment System (MCAS)—as well as the Partnership for Assessment of Readiness for College and Careers (PARCC) assessment. Because the assessments change over time and across grades, we standardize assessments by grade and year for this analysis.

B. Teacher Data

We construct two measures of teachers' subject-specific skills and qualifications. The first set of qualifications is based on license field. About 96% of the teachers in elementary schools hold a generalist license in elementary education that requires passing a test of the elementary curriculum. Massachusetts also offers licenses in English, science, history, and mathematics that require teachers to pass specialized content knowledge tests.[4] We construct indicators for whether the teacher holds an active license aligned with each of the four core subjects (including the math/science and English/history license) in the given school year. We also separately indicate whether a teacher holds only a generalist license (elementary or early childhood) without a corresponding content-area license.

Second, we construct subject-specific estimates of teaching effectiveness in math and ELA. We estimate the following VA model:

$$A_{ijt} = X_{ijt}\beta + \alpha_j + \epsilon_{ijt} \tag{1}$$

where *i* indexes students, *j* indexes teachers, and *t* indexes years. The vector X_{ijt} includes cubic polynomials in prior achievement in math and ELA, student race/gender, limited English proficiency status, special education, and classroom and school means of these characteristics. Following the convention in the literature, we then calculate an estimate of teacher quality using the combined residuals from Eq. (1), $\tilde{A}_{ijt} = A_{ijt} - X_{ijt}\hat{\beta}$ (Chetty et al., 2014; Kane & Staiger, 2008). Because teacher quality is estimated with error, we form minimum mean squared error predictions of teacher VA using the empirical Bayes procedure in Kane and Staiger (2008). We estimate both an annual version of the VA measure (i.e., using student data only from the current

[4] These licenses include English (grades 5–12), science (grades 1–6 or 5–8), history (grades 5–12), mathematics (grades 1–6, 5–8, or 5–12), and two combined licenses that cover middle school math and science (grades 5–8) and middle school English and history (grades 5–8).

school year) and an average VA measure that combines data from all available school years. We standardize the teacher VA measures by the estimated SD in teacher effects, so that a one-unit increment corresponds to one SD in the teacher VA distribution. To avoid potential endogeneity between instructional model and teacher effectiveness, we use only data from self-contained schools to estimate Eq. (1).[5] We also construct a measure of teacher comparative advantage by differencing the math and ELA VA estimates. We normalize this measure by the current subject assignment so that positive estimates of comparative advantage indicate higher VA in the subject corresponding to the current assignment.

C. Categorizing Instructional Models

We use student- and teacher-level schedule data to categorize school instructional models. Our categorization works as follows. We first identify students in self-contained classrooms. We identify students as self-contained if they have only one teacher in the four core academic subjects or if their teachers all teach the exact same set of classes to that student.[6] We then average the percentage of students in self-contained classrooms to the school-grade-year level. We define a school-grade-year cell as departmentalized if fewer than 50% of the students are in self-contained classrooms. As shown in Figure 1, most school-grade-year cells are either almost entirely self-contained or entirely departmentalized. Among departmentalized schools, 67% have no students in self-contained classrooms, and only 12% have more than 20% of their

[5] For instance, teachers may differentially improve over time when they teach fewer subjects. Using only self-contained classrooms for estimation of teacher VA ensures that comparative advantage is measured at baseline.
[6] Among cases where students have multiple teachers in the same self-contained classroom, 83% of the teachers are identified by the district as co-teaching the class. Some of these arrangements may involve teachers splitting subjects; however, we consider team teaching as an arrangement distinct from DI.

students in self-contained classes. The results are robust to alternative choices of the cutoff used to distinguish DI schools.[7]

Our definition of DI comprises two distinct forms of specialization. In one form of partial departmentalization, sometimes referred to as *subject-as-a-special*, students are assigned a single teacher for most subjects and a specialist in the remaining subject. Because they may avoid some of the negative effects of weaker student-teacher relationships, we consider these a distinct model in some analyses. We categorize students as belonging to a *partial DI* model if they are assigned a single teacher for three of their core subjects and a single-subject specialist in the remaining subject. Typically, these are schools in which one teacher is responsible for either science or writing instruction and the remaining core subjects are taught in a self-contained classroom. We categorize all other departmentalized students as belonging to a *full DI* model.[8] As with the overall DI indicator, we average the organizational indicators by school, grade, and year and assign individual cells to the most common assignment pattern. As shown in Figure A1 (Panel B), the schools we categorize as partial DI have almost all students assigned to a single specialist.

We display summary statistics by instructional model in Table 1. Departmentalized schools tend to serve students with similar specialized educational services (English language learners and special education students). But academic and non-academic outcomes are somewhat lower in DI schools than in self-contained schools. Current achievement is lower by about 0.07 SDs, and the nontest index is lower by about 0.06 SDs than in self-contained schools.

---

[7] We formally explore the sensitivity of these results to the definition of departmentalization in Appendix B. We show that the results are not sensitive to reasonable perturbations of the threshold used to distinguish instructional models.

[8] Among DI schools, 86% are classified as full DI.

The major differences between DI and non-DI schools are in school organization and teaching assignments. About 99% of students in non-DI schools are assigned to a self-contained classroom, while only about 7% of students in DI have a self-contained assignment. Students in non-DI environments have about 1 teacher on average across all core subjects compared to 2.4 teachers per year in DI schools. Very few teachers in non-DI schools are specialists (those teaching two subjects or fewer), while nearly half of all teachers in DI schools specialize. Similarly, about 1% of teachers in non-DI schools are licensed to teach one of the core subject fields (rather than possessing a generalist license). In DI schools, in-field license rates are 11% across all subjects. Finally, we show the average comparative advantage across students' math and ELA teachers. In non-DI schools, comparative advantage is very close to zero given that teachers are typically responsible for both subjects. In DI schools, teachers appear to be more effective in the subjects they are currently teaching by about 0.08 SDs in the teacher VA distribution on average.

We display the incidence of DI by year at the school-grade level in Table 2. As was the case nationally, DI became more common in Massachusetts over the 2010s: The rate of DI rose from 21.1% in 2012 to 29.9% in 2019. In addition, in any given year, there is an average of 88 school-grade cells switching into DI and 62 switching out of DI.

4. Research Design

We estimate the effects of DI on student outcomes using a DID design and school-grade-year data. We thus compare changes in student outcomes for school-grade cells switching to (or from) DI models to changes in outcomes in school-grade cells with either a self-contained or pre-existing departmentalized model. We estimate a basic two-way fixed-effects (TWFE) specification with school-by-grade and grade-by-year fixed effects on a sample of school-grade-year cells:

$$\overline{Y_{gst}} = D_{gst}\delta + \alpha_{gs} + \lambda_t + \epsilon_{gst}, \tag{2}$$

where $\overline{Y_{gst}}$ is the mean of some outcome Y for grade g in school s in year t, D_{gst} indicates treatment (DI) status, α_{gs} is a school-by-grade fixed effect, and λ_t is year fixed effect. We weight all models by student enrollment so that the estimated treatment effects are computationally identical to those estimated from student-level data. We retain a balanced panel of school-grade cells and obtain standard errors by clustering at the school level.[9]

The DID specifications described in Eq. (2) use two sources of within-school variation to identify the effect of DI: schools that newly switch into DI during the sample window and schools that switch from DI to self-contained classrooms. The effects of these kinds of switches may differ. For instance, if there are negative adjustment effects the first time a school adopts a new model of instruction, then the estimated effects of DI identified from switches from DI back to self-contained classrooms may be more positive than those identified from schools newly adopting DI as a model.

We therefore use an instrumental variables approach to isolate variation in instructional models arising from switches to DI that occur during our sample period. We construct an instrument equal to one for each year after a school-grade panel is first identified as implementing DI. We then instrument for current DI status using the switching instrument in Eq. (2). This approach scales the effect of DI for newly switching schools by the proportion of schools remaining departmentalized in any future school year.

Recent research has shown that the TWFE specification in Eq. (2) can provide biased estimates of the average effect of DI on student outcomes when there is variation in treatment

[9] About 15% of the school-grade observations are dropped due to missing years of data. These are primarily small schools and a subset of schools missing student assignment data in 2017. The results are not sensitive to their exclusion.

timing and heterogeneous treatment effects (de Chaisemartin & D'Haultfœuille, 2020; Goodman-Bacon, 2021). Staggered adoption is less likely to be a concern in our setting for two reasons. First, our sample includes a large proportion of never-treated (58%) or always-treated (12%) panels as controls. Therefore, relatively little of the identification in our application comes purely from variation in treatment timing. Second, our sample is based on a convenience sample when student scheduling data are available and there is no particular reason to believe that schools are selecting into DI during the sample period in such a way that would generate significant differences in DI across treatment cohorts. In other words, the early adopters of DI *in our sample* are not early adopters overall because the sample period begins when approximately one-fifth of schools had already departmentalized.[10]

Nonetheless, we address these concerns using two alternative, heterogeneity-robust approaches. Because we observe switches into and out of DI, we rely primarily on the dynamic DID method proposed by de Chaisemartin and D'Haultfœuille (2022). As with the standard TWFE estimator, this estimator uses panels that switch into or out of DI during the sample period to identify the effects of departmentalization. The estimator is a weighted aggregation of two-period differences in differences. For panels switching into or out of DI, the pool of comparison schools includes those with identical initial treatment status that have not switched models. We also use the imputation approach of Borusyak et al. (2024). Their estimator uses the untreated data to estimate Eq. (2) and then imputes the counterfactual achievement outcomes using the fitted values. By necessity, we drop all schools that are initially treated. We also drop observations following a switch back to self-contained instruction for any school that switches

[10] Using the Goodman-Bacon (2021) decomposition based on the first observed year of DI implementation, the timing groups receive relatively little weight, and treatment effects are not substantively different.

models twice or more during the sample. As with the TWFE estimator, we weight observations by student enrollment and cluster standard errors at the school level.

The key identifying assumption embedded in the research design is that schools switching instructional models would have experienced similar achievement trends as those maintaining a consistent model. Our research design implicitly assumes that the common experimentation with instructional models documented in Table 2 provides a natural experiment for assessing the effects of DI. We defer a fuller discussion of the identifying assumptions to Section 7, but we briefly discuss two primary threats to the identification strategy. First, principals frequently initiate switches in instructional models (Parker et al., 2017). Prior research has shown that principal turnover affects schoolwide achievement (Bartanen et al., 2019; Miller, 2013); we show in Section 7 that results are not sensitive to controlling for principal mobility patterns or omitting schools with principal turnover entirely. Second, departmentalization is sometimes part of a districtwide initiative; we also show that results are robust to controlling for district-specific time trends.

5. Results

A. School Organization and Teaching Assignments

Before moving to the main results, we first demonstrate that switches to DI entail meaningful changes in school structure and classroom assignments. In Table 3, we show results from regressions of classroom and teacher characteristics on indicators for the school's DI status, school-by-grade fixed effects, and year fixed effects. These specifications mirror our primary research design described in Eq. (2). The sample mean of each of these variables for the set of non-DI schools is indicated in column 1. We then present results from two sets of regressions that separately consider effects on school organization by DI type. In the first regression (column 2), we include the DI indicator. In the second regression (columns 3 and 4), we include the mutually exclusive partial and full DI indicators. Each observation in the table represents the coefficient on the specified DI indicator for the outcome specified in the given row.

We first assess the effects of DI switches on characteristics of students' classroom assignments. On average, switches to DI reduce the rate of self-contained assignments by 86 percentage points and increase the number of teachers by about 1 percentage point, which indicates the treatment assignment captures true differences in school organizational models. Results are generally consistent across the various types of DI we consider, although the full DI model increases the number of unique teachers by about an additional 0.2 teachers on average relative to partial DI.

The next set of outcomes examines changes in teacher characteristics. The average proportion of teachers who are specialists across subjects is about 30 percentage points higher in DI than in non-DI schools. This increase is driven mostly by full DI given that partial DI by definition includes only a single specialist. Nonetheless, even in full DI models, schools tend not to completely specialize instruction; teachers in DI models frequently teach three subjects in settings where subjects are not easily divisible across teachers. We find that schools changing to DI increase the rate of in-field licenses by about 3 percentage points on average, about a 300% increase relative to the non-DI mean. Finally, we find an increase in teacher comparative advantage. Recall that this is the difference between the same-subject and other-subject VA of a student's math and ELA teachers. The comparative advantage measure increases by about 0.05 teacher-level SDs in DI models, which suggests that schools are reallocating teachers toward their stronger subjects when they switch to DI. Overall, the results in Table 3 suggest that our DI measure captures meaningful variation in school structure and patterns of teacher assignments.

B. Departmentalization and Student Achievement

Our main results are shown in Figures 2 and 3 and in Table 4. In Panel A, we show the results using all within-school variation in DI status. The preferred DID estimates are in row 1. We estimate DI to improve achievement by about 0.02 SDs in math and about 0.04 SDs in both ELA and science. The average effect in math and ELA is about 0.03 SDs. In row 2, we use the procedure developed in de Chaisemartin and D'Haultfœuille (2022) instead of the TWFE estimator; results are similar for ELA and science. In math, however, we estimate an effect of close to zero. The results are broadly consistent with the dynamic treatment effects shown in Figure 2. The estimated placebo effects prior to treatment are not statistically significant. Consistent with the point estimates, there are apparent effects in ELA and science following the switch to DI, and little evidence that the effects of DI differ significantly over time.

In Panel B, we show the results using switches into DI only. As discussed in section 4, our baseline approach instruments current DI status with an indicator for whether a school has previously switched into DI as of the given school year.[11] The estimates are consistently more positive. Taking the DID estimates at face value, we estimate that switches into DI improve math achievement by about 0.04 SDs, ELA achievement by about 0.07 SDs, and science achievement by about 0.05 SDs. The same patterns are reflected in the dynamic estimates in Figure 3: We again see an upward effect of DI in the first year of implementation. The de Chaisemartin and D'Haultfœuille (2022) results are in the next row. The estimates are qualitatively similar but tend to be slightly smaller, and as before, the results for math are no longer significant. Because we are only using schools that switch into DI in these regressions, we additionally estimate results using the Borusyak et al. (2014) imputation approach. The results tend to be more similar to the

[11] The first-stage coefficient on the excluded instrument is 0.85 with a *t*-statistic of 48.02.

TWFE results: We estimate effects of about 0.05 SDs on the combined math and ELA measure and 0.04 SDs in science.[12]

In Panel C, we assess the effects of DI by the intensity of specialization within the school-grade-year cell. We first divide DI into full and partial DI. Recall that partial DI typically includes schools with a single specialist teacher, often in ELA or science. We find that the positive relationship between DI and test scores in math and ELA is driven by fully departmentalized schools. We find little evidence of effects on test scores in partially departmentalized models.

In Table 5, we estimate effects of DI separately by grade and for distinct student populations. In Panel A, we find little evidence that the effects of DI differ by grade. We include only special education students in Panel B and limited English proficient students in Panel C. We find no evidence that these student groups have weaker effects of DI; if anything, the point estimates are higher than the corresponding estimates for the full sample.

Overall, the findings suggest that switching to DI has positive effects on student achievement. These estimates are large in practical terms: The combined estimate of math and ELA of 0.053 for switchers in Panel B is larger than the average difference in effectiveness between teachers with 2 and 10 years of experience (e.g., Clotfelter et al., 2007). In all specifications, we can rule out effects of the magnitude of those estimated in Fryer (2018). With the caveat that our estimates are non-experimental, our findings suggest that DI as implemented in routine conditions has more beneficial consequences for student achievement than has been suggested by prior research.

[12] We could use analogous methods to isolate variation in school organization arising from switches from DI to self-contained classrooms. We find null effects of these switches, although the number of schools departmentalized in 2012 is considerably smaller than the set of schools in self-contained classes and our estimates are less precise. The null effect is consistent with schools selecting into treatment based on heterogeneity in the effects of DI.

C. School Climate and Student Engagement

Reductions in teacher effectiveness when specializing may be driven in part by weaker teacher-student relationships or inefficient personalization of instruction under DI models (Fryer, 2018). We use administrative data on student behaviors and student surveys of school climate offered to fifth graders (2018 and 2019) and fourth graders (2019) to help assess this possibility. Because we have only 2 years of survey data and only for a subset of elementary grades, we have limited ability to implement our standard DID approach; we can only compare changes in survey responses in schools that departmentalized fifth grade between 2018 and 2019. We therefore combine this research design with two alternative strategies. The first compares differences in 2019 responses in the fourth and fifth grades within schools where one grade offers DI and the other does not. That is, we estimate a model with grade and school-by-year fixed effects:

$$\overline{Y_{gst}} = D_{gst}\delta + \alpha_g + \alpha_{st} + \epsilon_{gst} \tag{4}$$

The identifying variation comes from comparing differences across grades in student responses among schools with one grade offering DI and those with both grades offering self-contained classes. The second approach combines both the cross-sectional variation in outcomes across grades within the same school and the time-series variation in outcomes within the same school-grade cell. That is, we estimate a model with school, grade, and year fixed effects:

$$\overline{Y_{gst}} = D_{gst}\delta + \alpha_g + \alpha_s + \alpha_t + \epsilon_{gst}. \tag{5}$$

Results for student behaviors and perceptions of school climate are shown in Table 6. The results in the first column replicate the DID design shown in Tables 5 and 6. We find little evidence that DI affects student behaviors on the nontest index that includes attendance, suspensions, and grade promotion. The coefficient on the overall survey factor is negative but not statistically significant. We do, however, find some evidence that DI weakens some aspects of the student-teacher relationships. For the participation, instructional environment, and discipline environment factors on the student surveys, the effects of DI are approximately -0.10 SDs and statistically significant. The results tend to be somewhat smaller, but more precisely estimated, when we additionally use cross-sectional variation in student outcomes in columns 2 and 3.

6. Mechanisms

Our results indicate that DI has positive effects on student achievement. At first glance, these results may appear to be inconsistent with the negative effects of DI in Fryer (2018) and of specialization on individual teachers' productivity in Bastian and Fortner (2020) and Hwang and Kisida (2022). In this section, we further explore potential explanations for the DI findings. We first examine the extent to which DI schools strategically assign teachers based on their comparative advantage in math and ELA instruction. We then replicate findings on teacher specialization and student achievement from Bastian and Fortner (2020) and Hwang and Kisida (2022). Finally, we assess the extent to which specialization or teacher comparative advantage (i.e., differential effectiveness across subjects) can explain our main results. The results confirm prior findings that DI involves a trade-off between gains from teacher comparative advantage and reduced individual effectiveness.

A. Comparative Advantage and Teaching Assignments

We assess strategic teaching assignments in DI and non-DI schools using the full set of teachers working in each school, grade, and year in our sample of elementary schools. We link the annual teacher data to the measures of teaching skill described in Section 3.B and regress indicators for whether a teacher is assigned to teach ELA or math on subject-specific skill measures and school-grade-year fixed effects separately for non-DI and DI schools. That is, we estimate

$$Y_{jgst} = T_{jt}\beta + \theta_{gst} + \epsilon_{ijt} \quad (6)$$

separately in DI and non-DI schools, where T_{jt} is a set of teacher characteristics and θ_{gst} is a school-grade-year fixed effect. The coefficients in β indicate the extent to which teachers with the indicated attributes are more likely to teach math or ELA compared to other teachers in the same school. The two sets of skill measures are teachers' VA in math and ELA and indicators for subject-area licenses. Following Bastian and Fortner (2020), we use prior-year teacher VA measures to predict current-year teaching assignments. We also include skill measures for both subjects in each regression. Principals seeking to maximize student achievement should assign teachers based on the difference between math and ELA VA (Condie et al., 2014; Goldhaber et al., 2013). Note that including both measures simultaneously ensures that identification comes from the portion of teacher VA (licensure) that is orthogonal to the corresponding measurement in the other subject.

The results are shown in Table 7 for ELA assignments (columns 1–2) and math assignments (columns 3–4). Our results suggest that schools do engage in strategic staffing when they switch to DI. In DI schools, teacher VA in each subject is positively predictive of assignment in that subject. A one SD increase in ELA VA increases the probability of assignment to ELA classes by about 2 percentage points; A one SD increase in math VA increases the probability of assignment to math classes by about 4 percentage points. Similar results hold for teacher license areas.[13] The relationships for non-DI schools are significantly weaker, consistent with the fact that most teachers in these schools instruct both subject areas.

[13] This finding is consistent with evidence on teacher specialization from Bastian and Fortner (2020) but contrasts with Hwang and Kisida (2022), who find that more effective teachers in a given subject are *less* likely to specialize in that subject. One difference between our results and theirs is that we consider how subject assignments vary *conditional* on corresponding skill measures in the other subject and how they control only for subject-specific skill measures. These regressions isolate only the portion of the skill measure that is specific to the subject under consideration.

The results are also consistent with the increase in the comparative advantage measure in DI schools noted in Table 3.

B. Teacher Specialization and Productivity

Bastian and Fortner (2020) and Hwang and Kisida (2022) both find that specialization reduces teacher effectiveness. These effects should at least partially offset the potential gains from trading instructional tasks in DI models. In this section, we replicate their findings in our sample. We estimate the impact of specializing on *individual teacher productivity* by regressing student outcomes on the number of subjects that a teacher instructs in a given year:

$$Y_{ijt} = \sum_{k=1}^{3} 1(Subjects_{jt} = k) * \delta^k + X_{it} + \alpha_{gs} + \gamma_{gt} + \mu_j + \epsilon_{ijt}. \quad (7)$$

For comparison with our main findings, we also estimate the effects of DI on individual productivity using the regression

$$Y_{ijt} = DI_{gst}\delta + X_{it} + \alpha_{gs} + \gamma_{gt} + \mu_j + \epsilon_{ijt}. \quad (8)$$

In Eqs. (7) and (8), each subject is estimated with a separate regression, and $Subjects_{jt}$ is the number of subjects taught by teacher j in a given subject in year *t*. The vector of student-level controls X_{it} includes cubic polynomials in prior achievement in math and ELA, student race/gender, limited English proficiency status, special education, test type and mode of delivery, classroom and school means of these characteristics, and grade-year effects.[14]

The results are shown in Table 8. Despite using data from a different state, our results largely match prior research on teacher specialization. We estimate that teaching math only reduces teacher VA by about 0.05–0.06 student SDs relative to teaching all subjects. The effects for teaching math and one or two other subjects are smaller, and not statistically significant, but

[14] Some schools in Massachusetts switched to the PARCC assessment in 2015, which was administered in both online and paper formats. Backes and Cowan (2019) find large test-mode penalties associated with the paper version.

still negative. The results for ELA are not significant; however, we cannot rule out effects of the magnitude of those found in Bastian and Fortner (2020) or Hwang and Kisida (2022). We find little evidence of any effect of specialization on science teachers' productivity, which is again consistent with Bastian and Fortner (2020).

In Panel B, we estimate the effects of switching to DI on individual teacher productivity. Because many teachers in DI models instruct multiple subjects, the negative results of DI for math are somewhat attenuated and no longer statistically significant. For ELA, the effects of DI are positive and marginally significant. The primary explanation for this divergence is that single subject specialization is relatively rare in DI models. In our sample, only 19% of teachers in DI schools teach one subject and 61% teach two or three subjects. Thus, although specialization does tend to reduce individual teacher productivity in math, the effects for DI specifically are at the low end of what has been found previously.

It is worth investigating whether these findings are consistent with the primary results in Section 5. In Table 3, we showed that DI increases the comparative advantage of a student's teachers by 0.05 teacher-level SDs. Assuming this captures the total effect of reassignments on teacher quality, this corresponds to an increase in student test scores of about 0.01 student-level SDs.[15] The results in Table 8 suggest that this reassignment effect would be fully offset by negative specialization effects in math, and that the combined effect of specialization and reassignments is about 0.025 SDs in ELA. These back-of-the-envelope calculations both correspond quite closely to the CH22 estimates in Table 4.

C. Specialization, Comparative Advantage, and the Effects of Departmentalization

[15] One SD in teacher VA corresponds to about 0.19 SDs in student achievement in math and 0.18 SDs in ELA. The estimated effects of DI on the average VA of a student's observed math and ELA teachers (using data from self-contained classrooms to estimate teacher VA, not shown) is slightly higher and corresponds to an increase in student achievement of about 0.015 student-level SDs based on teacher reassignments.

The results in this section suggest a trade-off associated with teacher specialization. Although specializing assignments is necessary for exploiting teacher comparative advantage, it appears to come at a cost for student achievement through reduced teacher productivity. We next test the association between these two factors and the effects of departmentalization. We estimate DI treatment effects at the school-grade-year level using the imputation approach of Borusyak et al. (2024). We then project the treatment effects on measures of school specialization, the scope for changes in teaching assignments, and teacher comparative advantage. To ensure the comparability of findings across covariates, we standardize each of the specialization measures within the distribution of departmentalized schools.

We present the results in Table 9. The first two rows show the relationship between two measures of actual teacher specialization, the average number of teachers assigned to each student and the proportion of specialist teachers, and the DI treatment effects. We find little evidence that either measure predicts the effects of DI. The lone exception is ELA, where the number of unique ELA teachers is marginally significant and positive. These results are generally consistent with Bastian and Fortner (2020) and Hwang and Kisida (2022), who find no relationship between specialization and student achievement, and with Fryer (2018), who finds that specialization is inversely related to the effects of DI.

The second two rows show the same relationships for our measures of comparative advantage, the proportion of teachers with an in-field content license, and the average difference between teachers' math and ELA VA. In contrast to the results on specialization, we do find that the extent of comparative advantage predicts the effectiveness of DI. Our estimated DI effects are larger when teachers have greater comparative advantage in the subjects they are teaching. The latter result is also consistent with Fryer (2018). Overall, we find that the gains from DI appear to be driven by schools exploiting the comparative advantage of their teachers, rather than from specialization per se. This may explain the apparent contradiction of our findings and the school-level analyses of specialization by Bastian and Fortner (2020) and Hwang and Kisida (2022).

Finally, we examine the relationship between the scope for reassigning teachers across grades within schools and the effects of DI. In the final row, we consider the proportion of the grades in each school that are departmentalized. Schools with more departmentalized grades tend to have larger effects of DI in both subjects. This last result may explain some of the discrepancies with the results in Fryer (2008), in which schools were unable to reassign teachers across grade levels.

7. Robustness Checks

A. Tests for Endogenous Instructional Model Switches

The key identifying assumption for both DID designs is that changes in school instructional models are not associated with changes in other factors affecting student achievement. Although we find little evidence that schools switching models are on different trajectories prior to changing models, one might be concerned that switches to DI correlate with other changes in school enrollment or policy. There are at least two distinct sources of potential bias. First, changes in instructional models might be associated with changes in the composition of the student body. Research designs that use variation in DI exposure across cohorts within a school might therefore conflate compositional differences with the effects of instructional models. Second, other school policy changes may coincide with switches to DI. In particular, prior research has identified principals as important drivers of changes in instructional models (Strohl et al., 2014). Research also indicates that principal turnover has independent effects on student outcomes (Bartanen et al., 2019; Miller, 2013). To the extent that new principals and new instructional models coincide, estimates of DI effects may partially capture effects of changes in school leadership.

To assess the potential confounding effect of compositional or other school changes, we regress student demographic and principal mobility information on DI indicators in Table 10. We conduct this analysis using all within-panel variation in DI status (columns 1 and 2) and only schools switching into DI models during the sample (columns 3 and 4). We estimate two versions of each regression: a panel fixed-effects specification that exactly mirrors the primary research design and a panel first-differences specification that focuses only on the first year of implementation and facilitates a cleaner assessment of the dynamics of principal turnover. Because the fixed-effects and first-differences estimates of the effects of student demographics are similar, we focus the discussion on the first-differences specifications.

We find limited evidence that DI is correlated with changes in student demographics. Switching to DI is associated with a 0.5 percentage-point reduction in the proportion of special education students but is associated neither with changes in participation in other special programming nor with student race/ethnicity. Consistent with the qualitative research on schools' instructional choices, we find more consistent evidence that switches are associated with principal turnover. Focusing on the schools switching into DI (column 4), we find that newly departmentalized schools are about 9 percentage points less likely to have switched principals in the same year. The coefficient on the DI indicator for the lagged principal transition outcome in the next row indicates that most of this effect can be explained by higher principal mobility in the year *prior* to switching to DI. It thus appears that new principals are disproportionately likely to change models in their second year of tenure. We show below that the main results are robust to various methods for controlling for the dynamics of principal mobility.

B. Alternative Specifications

Our main results in Table 4 suggest that switches to DI improve student achievement. As discussed in Section 4, the primary threat to identification is the possibility of switches to DI being correlated with other school-level shocks that affect student achievement. In this section, we investigate several alternate explanations for the positive effect of DI on student achievement. Results are shown in Table 11.

In columns 1 (TWFE) and 2 (de Chaisemartin & D'Haultfœuille, 2022), we investigate whether results could be driven by changes in the composition of the student body by adding controls for average race, free and reduced-price lunch, English language learner status, special education, whether students took a test online, and gender. Results are nearly identical to the columns 1 and 2 equivalents in Table 4. Column 3 adds grade-by-year fixed effects, again with nearly identical results. Consistent with the tests for selection on observables in Table 10, it does not appear that compositional changes pose a significant threat to identification.

As discussed above, another threat to identification is the possibility that DI is associated with other district policy changes. Although our interviews with school principals suggested that DI decisions are largely at the school (rather than district) level, it is possible that that some switches to DI were correlated with other measures intended to boost the performance of schools in the district. Column 4 adds district-by-year fixed effects to remove any variation in student outcomes associated with such policy changes. Results are again similar to the base specification in column 1.

Another concern raised in Section 4 is the role of principal mobility in explaining changes in instructional models and student achievement. Principal turnover could generate two sources of bias in our estimates. First, research on principal turnover suggests that they are related to dynamic changes in school effectiveness. Principal turnover has negative effects on student achievement in the short run and may tend to follow secular declines in school achievement (Bartannen et al., 2019; Miller, 2013). Thus, our estimates of DI effects could conflate changes in instructional models with the mean reversion that follows changes in school leadership. Second, changes in instructional models may be correlated with the quality of school leadership if principals who switch to DI are more effective in other respects. If switches to DI are correlated both with principal VA and principal switches, then we might conflate the effects of instructional models with principal effects.[16] We test for these possibilities in the remaining columns.

If biases from principal mobility are caused solely by the dynamic effects of leadership changes on student outcomes, rather than by differences in time-invariant principal quality, then allowing time trends to differ with the patterns of principal mobility should ameliorate these biases. In column 5, we interact the year fixed effects with identifiers for *principal mobility groups*, which are groups of schools defined by the set of school years in which they had a new principal. These regressions compare DI and non-DI schools that have had identical evolutions of school leadership. Results are again nearly identical to those in column 1. In column 6, we implement de Chaisemartin & D'Haultfœuille's (2022) approach while allowing flexible time trends for the principal mobility groups. Estimates are less precise and not statistically significant but are similar in magnitude to the corresponding estimates in Table 5.

Finally, if biases arise from changes in principal effectiveness, then focusing on a subset of schools that had no changes in principal leadership should provide unbiased estimates of the

[16] There is some disagreement about whether principal effects on student test scores are large enough to warrant concern in this setting. Bartanen et al. (2022) find that the variance in principal effects on student outcomes in the short run is quite small and statistically indistinguishable from zero in many cases. On the other hand, Austin et al. (2023) estimate that one SD in principal test VA corresponds to about 0.06 SDs in student achievement terms.

effects of DI on student outcomes. The logic of this test is that bias from principal quality results only when school principals (and their VA) change and this change leads to a change in instructional model. Focusing on schools with stable leadership precludes changes in schools' principal VA. We pursue this approach in column 7, retaining only schools that had the same principal throughout the sample period. Results are again similar.

C. Triple-Differences Estimates

Perhaps the simplest way of accounting for schoolwide policy or leadership changes is to compare grades that change models to other grades in the same school that retain the same organizational structure. In this section, we pursue this strategy using a triple-differences (TD) design that relies on the differential timing of changes to DI across grades. We estimate the regression

$$\overline{Y_{gst}} = D_{gst}\delta + \alpha_{gs} + \lambda_{gt} + \theta_{st} + \epsilon_{gst}, \tag{5}$$

adding a school-by-year effect θ_{st} to Eq. (2). The TD design compares the evolution of student outcomes in newly departmentalized grades in a school relative to grades that retain their existing instructional model. This design therefore uses only two sources of variation in instructional models: (1) schools that switch some grades' instructional models while keeping others constant and (2) schools that roll out changes to DI across grades in separate years. The excluded source of variation is schools that fully implement DI across grades, which comprises about half the departmentalized school-grade-year cells in our sample.

We present the TD estimates in Table 12. The estimated effect of DI on average math and ELA test scores is close to zero and statistically insignificant; estimates are similar for math and ELA scores separately. There are two potential explanations for the discrepancy between the DID and TD results. First, switches to DI may be correlated with other schoolwide factors that tend to increase achievement and the DID estimates are biased. Second, as we note above, the TD design relies only on the roughly 50% of DI schools that implement DI only in some grades. It may be the case that these schools have weaker implementation and smaller effects of DI on student outcomes. We present two sources of evidence suggesting the latter interpretation.

First, in the next panel of Table 12, we compare the TD results to the DID results by intensity of school adoption. For schools that only partially implement DI across grades, the results are quite similar to the baseline estimates for the TD design in the first panel. The coefficient on partial school adoption is 0.007 in the DID design for the average test-score measure. Results for math and ELA scores are also similar. Hence, the DID regression yields estimated effects for the partial adopters that match those from the TD regression. Second, when we add the fully interacted fixed effects to this regression in the third panel, there is little movement in the coefficient on partial adopters.[17] This suggests that unobserved shocks to school performance are not strongly correlated with switches to DI, at least among schools that switch only a subset of grades at once. Thus, the difference between the TD and DID models appears to be driven by the exclusion of full-DI schools rather than school shocks.

We conclude that our results are likely not driven by changes in school composition or policy. Although the TD designs are close to zero and statistically insignificant, this appears to be driven by treatment-effect heterogeneity rather than by time-varying school unobservables. Nonetheless, it is worth noting that even with the most conservative estimates of the effects of DI on student achievement, we can rule out negative effects of more than about 0.016 SDs, an effect size that is smaller than the magnitude found by Fryer (2018).

[17] There are no estimates displayed for entirely departmentalized schools with school-by-year fixed effects in Table 11 because these estimates require within-school variation in DI. About half of the DI sample comes from fully departmentalized schools.

8. Discussion

Prior research has shown that teacher specialization has negative effects on teacher productivity but has reached mixed conclusions about the effectiveness of DI as a school organizational model (Bastian & Fortner, 2020; Fryer, 2018; Hwang & Kisida, 2021). Our findings on individual teacher productivity are generally consistent with prior research on teacher specialization from Bastian and Fortner (2020) and Hwang and Kisida (2021) in that individual teachers are no more effective in specialized settings than in self-contained classrooms. But we still find that switches to DI can lead to net increases in student achievement due to an improvement in the match between teachers and subject assignments in DI. The overall effects of DI are most positive in ELA and science, where the negative effects of working in specialized assignments are smallest, and mixed in math, where a more significant penalty is associated with specialization.

The benefits to DI are concentrated in schools with greater scope for reassigning teachers to stronger subjects. This includes schools where teachers differ in their effectiveness in teaching math and English and in schools that switch all grades to DI. One potential implication of this finding is that voluntary switches made by principals might be expected to generate stronger results than in cases in which DI is externally imposed. In other words, the schools in our sample observed implementing DI may have more positive treatment effects than the state as a whole. This is one possible explanation for the divergence between the results here and in Fryer (2018). Thus, while it may be tempting to argue that taking the results of this paper at face value suggests that more schools should switch to DI, these findings may not hold for schools with less scope to strategically reassign teachers.

Nonetheless, these findings are notable for three reasons. First, changes in assignment patterns following switches to DI suggest that principals are aware of which teachers are stronger in which subjects and use this knowledge to inform assignment patterns in departmentalized schools. These changes in teacher assignments appear to offset the possible negative effects of teacher specialization on individual teacher productivity. Second, the actual benefits of DI estimated here are remarkably similar to simulations of the expected benefits associated with reassigning teachers to their stronger subjects. In particular, Goldhaber et al. (2013) estimate gains in average math and ELA achievement of 0.049–0.068 SDs when principals make decisions based on a 3-year average of teacher VA and operate to maximize total teacher VA standardized by subject. And third, with the caveat that these results are non-experimental, the difference between these results and those in Fryer (2018) suggests that DI implemented voluntarily may be beneficial when schools have the opportunity to coordinate changes in instructional models and where the scope for exploiting comparative advantage is significant.

References

Anderman, E. M., & Maehr, M. L. (1994). Motivation and schooling in the middle grades. *Review of Educational Research*, *64*(2), 287–309. https://doi.org/10.3102/00346543064002287

Anderman, L. H. (2003). Academic and social perceptions as predictors of change in middle school students' sense of school belonging. *The Journal of Experimental Education*, *72*(1), 5–22. https://doi.org/10.1080/00220970309600877

Anderson, R. C. (1962). The case for teacher specialization in the elementary school. *The Elementary School Journal, 62*(5), 253–260.

Austin, W., Chen, B., Goldhaber, D., Hanushek, E., Holden, K., Koedel, C., Ladd, H., Luo, J., Parsons, E., Phelan, G., Rivkin, S., Sass, T., & Turaeva, M. (2023). *Does regulating entry requirements lead to more effective principals?* (CALDER Working Paper No. 213-0323-2). National Center for Analysis of Longitudinal Data in Education Research, American Institutes for Research.

Backes, B., & Cowan, J. (2019). Is the pen mightier than the keyboard? The effect of online testing on measured student achievement. *Economics of Education Review*, *68*, 89–103. https://doi.org/10.1016/j.econedurev.2018.12.007

Backes, B., Cowan, J., Goldhaber, D., & Theobald, R. (2022). *Teachers and school climate: Effects on student outcomes and academic disparities* (CALDER Working Paper No. 274-1022). National Center for the Analysis of Longitudinal Data in Education Research, American Institutes for Research. https://caldercenter.org/sites/default/files/CALDER%20Working%20Paper%20274-1022_0.pdf

Backes, B., Cowan, J., Goldhaber, D., & Theobald, R. (2023). *How to measure a teacher: The influence of test and nontest value-added on long-run student outcomes* (CALDER Working Paper No. 270-0423-2). National Center for Analysis of Longitudinal Data in Education Research, American Institutes for Research. https://caldercenter.org/sites/default/files/CALDER%20WP%20270-0423-2.pdf

Bartanen, B., Grissom, J. A., & Rogers, L. K. (2019). The impacts of principal turnover. *Educational Evaluation and Policy Analysis*, *41*(3), 350–374. https://doi.org/10.3102/0162373719855044

Bartanen, B., Husain, A. N., & Liebowitz, D. D. (2022). *Rethinking principal effects on student outcomes* (No. 22–621). Annenberg Institute at Brown University. https://doi.org/10.26300/R5SF-3918

Bastian, K. C., & Fortner, C. K. (2020). Is less more? Subject-area specialization and outcomes in elementary schools. *Education Finance and Policy*, *15*(2), 357–382. https://doi.org/10.1162/edfp_a_00278

Becker, G. S., & Murphy, K. M. (1992). The division of labor, coordination costs, and knowledge. *The Quarterly Journal of Economics, 107*(4), 1137–1160. https://doi.org/10.2307/2118383

Blazar, D. (2015). Grade assignments and the teacher pipeline: A low-cost lever to improve student achievement? *Educational Researcher*, *44*(4), 213–227. https://doi.org/10.3102/0013189X15580944

Borusyak, K., Jaravel, X., & Spiess, J. (2024). *Revisiting event study designs: Robust and efficient estimation* (arXiv:2108.12419). arXiv. https://doi.org/10.48550/arXiv.2108.12419

Brobst, J. A., & Markworth, K. A. (2019). Elementary content specialization: Perspectives on perils and promise. *School Science and Mathematics, 119*(7), 369–381. https://doi.org/10.1111/ssm.12362

Chan, T.C., & Jarman, D. (2004). Departmentalize elementary schools. *Principal, 84*(1), 70–72.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *American Economic Review, 104*(9), 2633–2679

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2007). Teacher credentials and student achievement: Longitudinal analysis with student fixed effects. Economics of Education Review, 26(6), 673–682. https://doi.org/10.1016/j.econedurev.2007.10.002

Condie, S., Lefgren, L., & Sims, D. (2014). Teacher heterogeneity, value-added and education policy. *Economics of Education Review*, *40*, 76–92. https://doi.org/10.1016/j.econedurev.2013.11.009

Davis, H. A. (2003). Conceptualizing the role and influence of student-teacher relationships on children's social and cognitive development. *Educational Psychologist*, *38*(4), 207–234. https://doi.org/10.1207/S15326985EP3804_2

de Chaisemartin, C., & D'Haultfœuille, X. (2018). Fuzzy differences-in-differences. *The Review of Economic Studies*, *85*(2), 999–1028. https://doi.org/10.1093/restud/rdx049

de Chaisemartin, C., & D'Haultfœuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, *110*(9), 2964–2996. https://doi.org/10.1257/aer.20181169

de Chaisemartin, C., & D'Haultfœuille, X. (2022). *Difference-in-differences estimators of intertemporal treatment effects* (NBER Working Paper No. 29873). National Bureau of Economic Research. https://doi.org/10.3386/w29873

Dessein, W., & Santos, T. (2006). Adaptive organizations. *Journal of Political Economy, 114*(5), 956–995. https://doi.org/10.1086/508031

Fryer, R. G. (2018). The “pupil” factory: Specialization and the production of human capital in schools. *American Economic Review*, *108*(3), 616–656. https://doi.org/10.1257/aer.20161495

Goldhaber, D., Cowan, J., & Walch, J. (2013). Is a good elementary teacher always good? Assessing teacher performance estimates across subjects. *Economics of Education Review*, *36*, 216–228. https://doi.org/10.1016/j.econedurev.2013.06.010

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, *225*(2), 254–277. https://doi.org/10.1016/j.jeconom.2021.03.014

Haley, III, R. T. (2018). *A case study on the perceived impact of elementary school departmentalization on teacher math self-efficacy* [Doctoral dissertation, University of the Pacific]. https://scholarlycommons.pacific.edu/uop_etds/3548

Harter, S., Whitesell, N. R., & Kowalski, P. (1992). Individual differences in the effects of educational transitions on young adolescent’s perceptions of competence and motivational orientation. *American Educational Research Journal*, *29*(4), 777–807. https://doi.org/10.3102/00028312029004777

Hill, A. J., & Jones, D. B. (2018). A teacher who knows me: The academic benefits of repeat student-teacher matches. *Economics of Education Review*, *64*, 1–12. https://doi.org/10.1016/j.econedurev.2018.03.004

Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers’ mathematical knowledge for teaching on student achievement. *American Educational Research Journal, 42*, 371–406.

Hwang, N., & Kisida, B. (2022). Spread too thin: The effect of specialization on teaching effectiveness. *Educational Evaluation and Policy Analysis, 44*(4), 593–607.

Jackson, C. K. (2018). What do test scores miss? The importance of teacher effects on non-test score outcomes. *Journal of Political Economy, 126*(5), 2072–2107.

Jacob, B. A., & Rockoff, J. E. (2011). *Organizing schools to improve student achievement: Start times, grade configurations, and teacher assignments*. Brookings Institution, Hamilton Project.

Kane, T., & Staiger, D. (2008). *Estimating teacher impacts on student achievement: An experimental evaluation* (Working Paper No. 14607). National Bureau of Economic Research. https://doi.org/10.3386/w14607

Ladd, H., & Sorensen, L. C. (2017). Returns to teacher experience: Student achievement and motivation in middle school. *Education Finance and Policy, 12*(2), 241–279.

Lobdell, L. O., & Van Ness, W. J. (1963). The self-contained classroom in the elementary school. *The Elementary School Journal, 63*(4), 212–217.

Marks, H. M. (2000). Student engagement in instructional activity: Patterns in the elementary, middle, and high school years. *American Educational Research Journal*, *37*(1), 153–184. https://doi.org/10.3102/00028312037001153

Massachusetts Department of Elementary and Secondary Education [DESE]. (2019). *2018 Views of Climate and Learning (VOCAL) validity study*. Malden, MA: Department of Elementary and Secondary Education.

Miller, A. (2013). Principal turnover and student achievement. *Economics of Education Review*, *36*, 60–72. https://doi.org/10.1016/j.econedurev.2013.05.004

National Center for Education Statistics [NCES]. (2007). Schools and Staffing Survey, 2003-2004. United States Department of Education, National Center for Education Statistics.

National Center for Education Statistics [NCES]. (2017). National Teacher and Principal Survey, 2015-2016. United States Department of Education, National Center for Education Statistics.

National Center for Education Statistics [NCES]. (2023). National Teacher and Principal Survey, 2020-2021. United States Department of Education, National Center for Education Statistics.

National Forum on Education Statistics. (2011). *Prior-to-secondary school course classification system: School codes for the exchange of data (SCED)* (NFES 2011–801). National Center for Education Statistics, U.S. Department of Education.

Ost, B. (2014). How do teachers improve? The relative importance of specific and general human capital. *American Economic Journal: Applied Economics*, *6*(2), 127–151. https://doi.org/10.1257/app.6.2.127

Otto, H. J. (1931). Specialization in teaching in the elementary school. *The Elementary School Journal, 32*(1), 17–21.

Parker, A., Rakes, L., & Arndt, K. (2017). Departmentalized, self-contained, or somewhere in between: Understanding elementary grade-level organizational decision-making. *The Educational Forum*, *81*(3), 236–255. https://doi.org/10.1080/00131725.2017.1314569

Pellegrini, A. D., & Bartini, M. (2000). A longitudinal study of bullying, victimization, and peer affiliation during the transition from primary school to middle school. *American Educational Research Journal*, *37*(3), 699–725. https://doi.org/10.3102/00028312037003699

Reeve, J. (2006). Teachers as facilitators: What autonomy-supportive teachers do and why their students benefit. *The Elementary School Journal*, *106*(3), 225–236. https://doi.org/10.1086/501484

Shavelson, R. J., & Stern, P. (1981). Research on teachers' pedagogical thoughts, judgments, decisions, and behavior. *Review of Educational Research, 51*(4), 455–498. https://doi.org/10.2307/1170362

Strohl, A., Schmertzing, L., & Schmertzing, R. (2014). Elementary teachers' experiences and perceptions of departmentalized instruction: A case study. *Journal of Case Studies in Education, 6*, 1–17.

Thomas, S., & Oldfather, P. (1997). Intrinsic motivations, literacy, and assessment practices: "That's my grade. That's me." *Educational Psychologist*, *32*(2), 107–123. https://doi.org/10.1207/s15326985ep3202_5

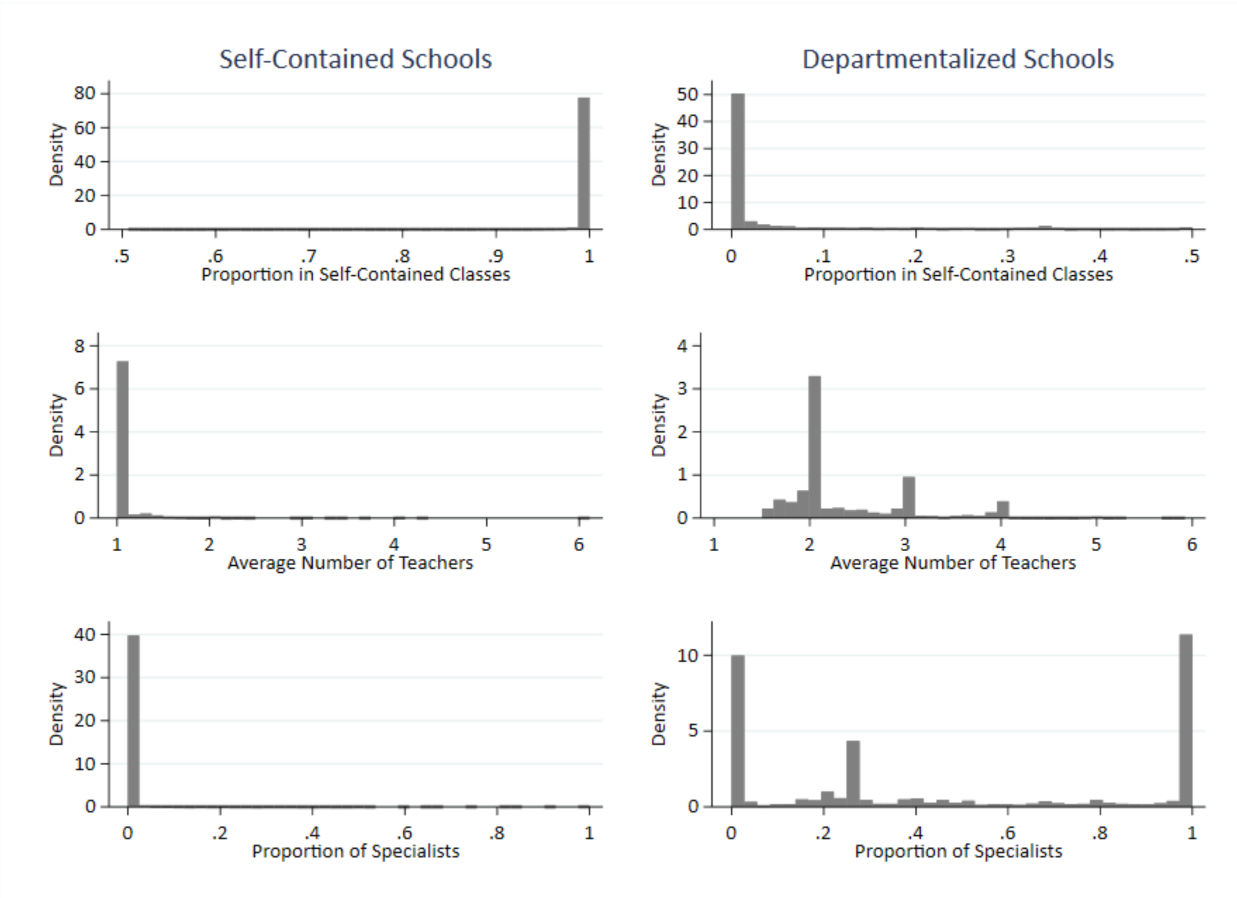
Wedenoja, L., Papay, J., & Kraft, M. A. (2022). *Second time's the charm? How repeat student-teacher matches build academic and behavioral skills* (EdWorkingPaper No. 22-590). Annenberg Institute at Brown University. https://doi.org/10.26300/SDDW-AG22

Wentzel, K. R. (1997). Student motivation in middle school: The role of perceived pedagogical caring. *Journal of Educational Psychology*, *89*(3), 411–419. https://doi.org/10.1037/0022-0663.89.3.411

Zimmerman, A. (2018). As NYC encourages more elementary teachers to specialize in math, new research shows the strategy could hurt student learning. *Chalkbeat.* Retrieved from https://www.chalkbeat.org/posts/ny/2018/05/17/as-nyc-encourages-more-elementaryteachers-to-specialize-in-math-new-research-shows-the-strategy-could-hurt-studentlearning/

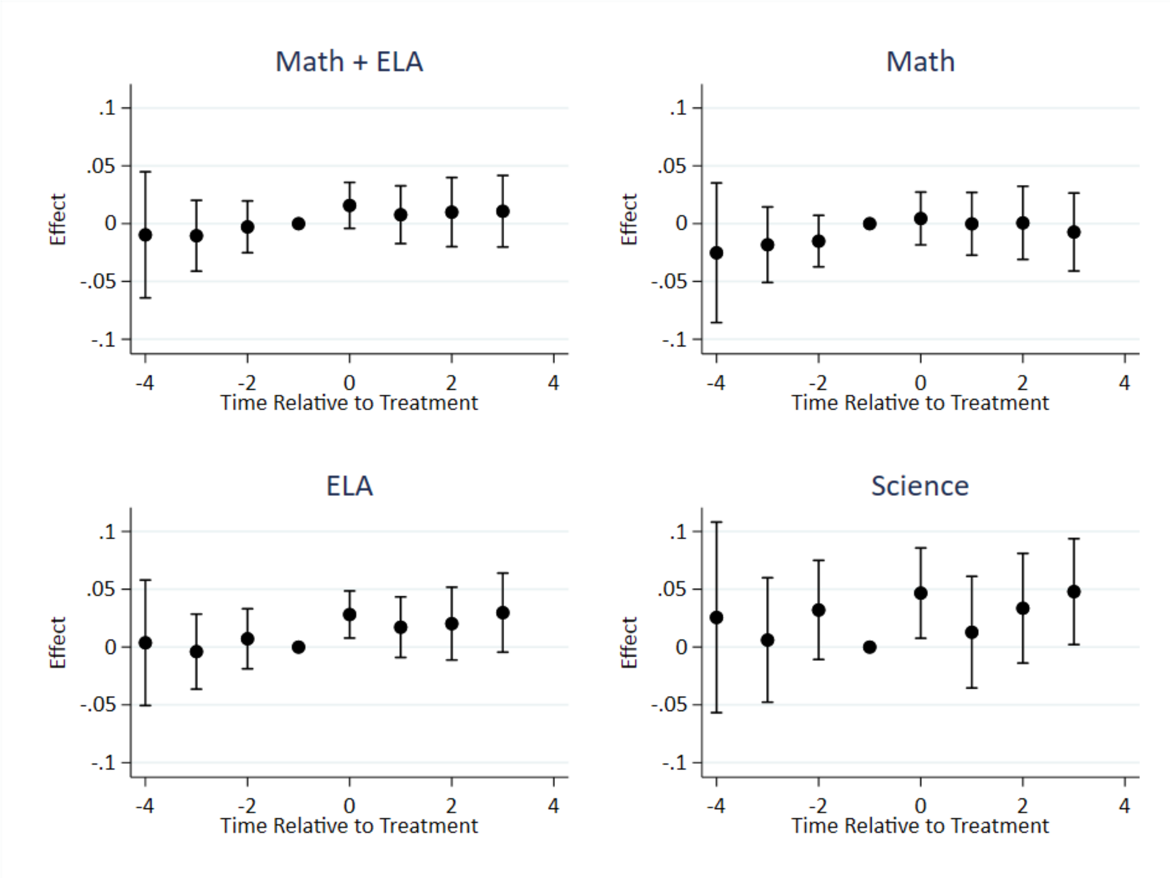
Figures and Tables

Figure 1. Class and Teacher Assignments by Instructional Model



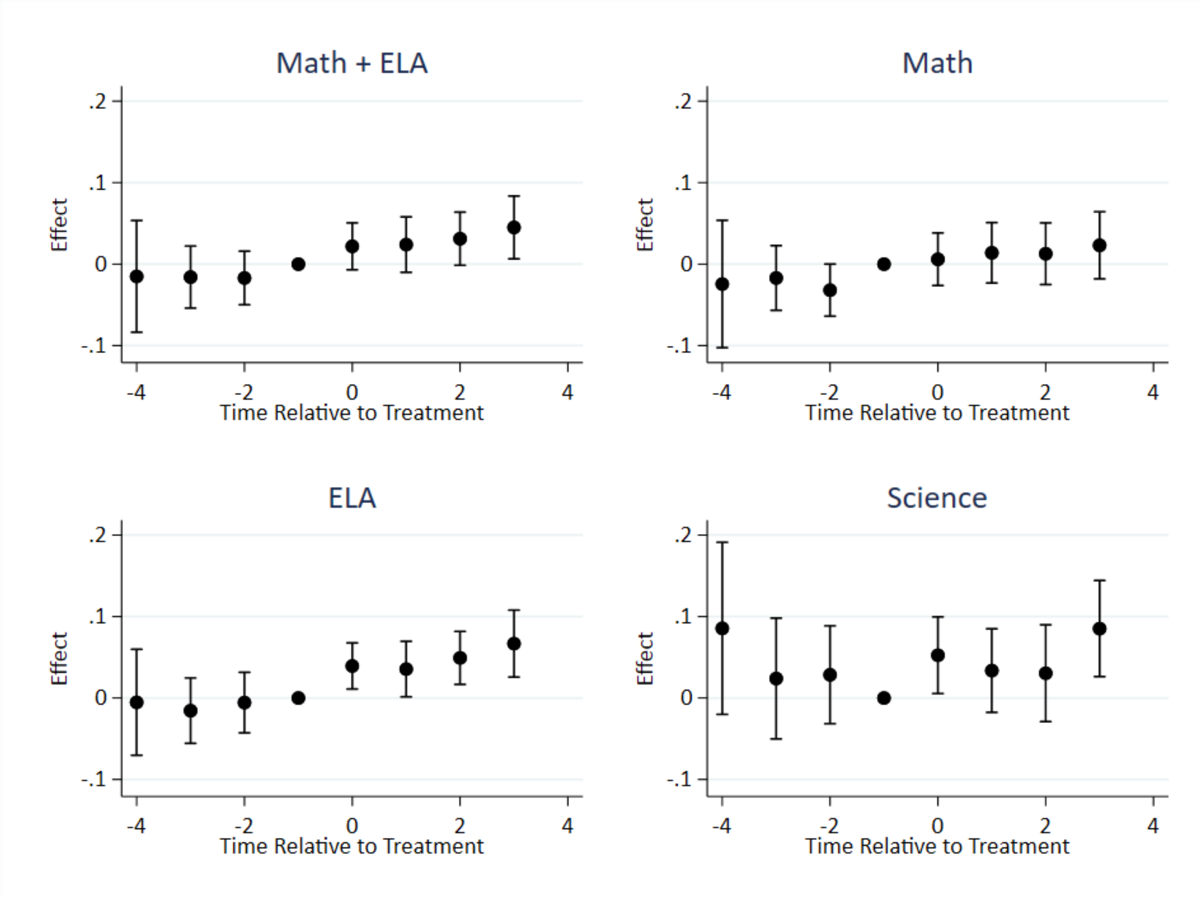
Notes: Histograms of the proportion of students in self-contained classrooms (row 1) and average number of teachers (row 2) by school, grade, and year for each organizational model (self-contained or departmentalized). Self-contained class assignments are identified as those in which a student is assigned to one teacher for all core subject classes. Number of teachers indicates average number of teachers assigned to a student across the four core subjects, including co-teachers. Specialists indicates proportion of a student's teachers who instruct two or fewer subjects. Departmentalized schools are defined as those in which no more than 50% of students are assigned to self-contained classes.

Figure 2. Dynamic Effects of Departmentalization on Student Achievement



Notes. Dynamic effects of DI switches on student outcomes estimated using method of de Chaisemartin and D'Haultfœuille (2022). Models include school-grade and year fixed effects. Observations weighted by student enrollment. Point estimates and 95% confidence intervals included. Standard errors clustered by school.

Figure 3. Event Study Estimates of Departmentalization Switches on Student Achievement



Notes. Dynamic effects of switches to DI on student outcomes estimated using method of de Chaisemartin and D'Haultfœuille (2022). Models include school-grade and year fixed effects. Observations weighted by student enrollment. Point estimates and 95% confidence intervals included. Standard errors clustered by school.

Table 1. Descriptive Statistics by School Instructional Model

	(1) Self-Contained		(2) Departmentalized	
	Mean	SD	Mean	SD
<u>Student Characteristics</u>				
English learner	0.093	0.125	0.084	0.114
Male student	0.512	0.059	0.510	0.057
Student eligible for FRPL	0.359	0.306	0.410	0.309
Full-inclusion SPED students	0.129	0.054	0.120	0.056
Partial-inclusion SPED students	0.026	0.033	0.029	0.034
Substantially separate SPED students	0.022	0.037	0.027	0.042
Asian students	0.071	0.092	0.059	0.085
Black students	0.072	0.113	0.083	0.134
Hawaiian/Pacific Islander students	0.001	0.004	0.001	0.004
Hispanic students	0.175	0.212	0.204	0.247
American Indian students	0.002	0.006	0.003	0.009
Multiple race/ethnic groups	0.037	0.030	0.033	0.032
<u>Student Outcomes</u>				
Standardized ELA score	0.057	0.435	-0.014	0.417
Standardized math score	0.061	0.432	-0.013	0.413
Standardized science score	0.101	0.486	0.035	0.449
Average math and ELA tests	0.057	0.423	-0.016	0.402
Standardized behavioral index	0.034	0.225	-0.027	0.269
<u>Organizational Characteristics</u>				
Number of teachers	1.059	0.211	2.380	0.723
Self-contained	0.986	0.064	0.069	0.125
Subject specialists	0.005	0.039	0.478	0.418
In-field license	0.011	0.035	0.108	0.190
Comparative advantage	0.001	0.036	0.079	0.515
Observations	12678		4170	

Notes. Summary statistics by organizational model used in school-grade-year cell. Sample includes all elementary school grades 3–6 between 2012 and 2018. Observations weighted by student enrollment. Nontest index includes a factor constructed from log absences, log days suspended, and an indicator for grade promotion. Subject specialists indicates average proportion of students' teachers who teach two or fewer subjects. In-field licensed teachers indicates average proportion of students' teachers with a subject-specific (as opposed to generalist) license. Comparative advantage is the average difference between the VA of students' math or ELA teachers in the given subject and their VA in the other subject. SD = standard deviation. FRPL = free and reduced-price lunch. SPED = special education. ELA = English language arts.

Table 2. Timing of Instructional Model Switches

Year	Self-Contained	Departmentalized	Percent DI	Switch to DI	Switch from DI
2012	1657	449	21.3		
2013	1639	467	22.2	79	61
2014	1637	469	22.3	64	62
2015	1599	507	24.1	101	63
2016	1566	540	25.6	86	53
2017	1563	543	25.8	70	67
2018	1541	565	26.8	97	75
2019	1476	630	29.9	116	51

Notes. Instructional model and switches by school year. Observations are at the school-grade-year level. Switches indicate number of school-grade cells that operated the other model in the previous school year.

Table 3. Effects of School Model on Teacher Assignments

	Non-DI Mean	DI	By DI Intensity	
			Full DI	Partial DI
Number of teachers	1.059	1.040***	1.075***	0.876***
		(0.030)	(0.034)	(0.069)
Self-contained	0.986	-0.856***	-0.845***	-0.908***
		(0.012)	(0.013)	(0.013)
Subject specialists	0.005	0.320***	0.368***	0.093***
		(0.019)	(0.022)	(0.021)
In-field license	0.011	0.032***	0.025***	0.064***
		(0.004)	(0.004)	(0.011)
Comparative advantage	0.001	0.051***	0.051**	0.047*
		(0.019)	(0.023)	(0.028)

Notes. Coefficients on school DI models from regressions of class characteristics or teacher specialization measures on DI status, school-grade, and grade-year fixed effects. Departmentalization measures constructed as described in the text. Number of teachers indicates the number of unique teachers in core subject classes. Self-contained class is an indicator for whether a student is assigned to a single teacher for all classes. Specialists indicates proportion of a student's teachers in each subject who specialize (teaching two or fewer subjects) in each of the core subject areas. In-field license indicates the proportion of a student's teachers in each subject who have an active in-field teaching license. Comparative advantage is the average difference between the VA of students' math or ELA teachers in the given subject and their VA in the other subject. Observations weighted by student enrollment. Standard errors clustered by school in parentheses. *$p < 0.10$; **$p < 0.05$; ***$p < 0.01$.

Table 4. Effects of Departmentalization on Student Achievement

Estimator	*Treatment*	(1) Tests	(2) Math	(3) ELA	(4) Science
Panel A. All Switches					
TWFE	DI	0.027***	0.020*	0.036***	0.039***
		(0.009)	(0.011)	(0.009)	(0.014)
CH22	DI	0.014	-0.000	0.030**	0.043**
		(0.013)	(0.014)	(0.013)	(0.022)
Panel B. Switches to DI					
TWFE-IV	DI	0.053***	0.040**	0.069***	0.047**
		(0.016)	(0.019)	(0.016)	(0.021)
CH22	DI	0.036**	0.016	0.057***	0.059**
		(0.017)	(0.019)	(0.016)	(0.026)
BJS24	DI	0.047***	0.031*	0.065***	0.036*
		(0.016)	(0.018)	(0.016)	(0.020)
Panel C. Effects by DI Intensity					
TWFE	Partial DI	0.011	0.000	0.023	-0.004
		(0.017)	(0.020)	(0.018)	(0.027)
	Full DI	0.030***	0.024**	0.038***	0.044***
		(0.010)	(0.011)	(0.010)	(0.014)
Observations		16,848	16,848	16,848	4,448

Notes. Coefficients on school DI models from regressions of mean student outcomes on DI status and specified fixed effects. Sample includes students in grades 3–6 in elementary schools. DI treatment indicators constructed as described in the text. Tests is the average standardized test score in both math and ELA tests. TWFE = two-way fixed effects estimator. CH22 = de Chaisemartin & D'Haultfœuille (2022). BJS24 = Borusyak et al. (2024). The estimates in Panel B use only identifying variation from schools newly switching from self-contained to DI between 2013 and 2018. The TWFE-IV estimator instruments current DI status using an indicator for a year after a school's initial switch to DI. The CH22 and BJS24 estimators construct difference in differences for newly switching schools only. Observations weighted by student enrollment. Standard errors clustered by school in parentheses. *$p < 0.10$; **$p < 0.05$; ***$p < 0.01$.

Table 5. Effects of Departmentalization by Student Characteristics

	Tests	ELA	Math	Science
Panel A. Effects by Grade				
3rd Grade	0.028**	0.040***	0.016	
	(0.014)	(0.014)	(0.016)	
4th Grade	0.027*	0.035**	0.020	
	(0.014)	(0.015)	(0.015)	
5th Grade	0.029**	0.043***	0.021	
	(0.012)	(0.013)	(0.014)	
6th Grade	0.027	0.015	0.040	
	(0.024)	(0.028)	(0.027)	
Observations	16,680	16,680	16,680	
Panel B. Special Education Students				
DI	0.030***	0.023*	0.013	0.010
	(0.010)	(0.013)	(0.015)	(0.017)
Observations	232,585	206,229	206,395	52,908
Panel C. English Language Learners				
DI	0.052***	0.067***	0.051**	0.040
	(0.017)	(0.020)	(0.023)	(0.028)
Observations	121,695	97,317	99,467	17,842

Notes. Coefficients on school DI models from regressions of mean student outcomes on DI status and specified fixed effects. Sample includes students in grades 3–6 in elementary schools. Tests is the average standardized test score in both end-of-grade tests. In Panel A, observations are weighted by student enrollment. In Panels B and C, estimates are derived from student-level data using only the indicated samples of students. Standard errors clustered by school in parentheses. *$p < 0.10$; **$p < 0.05$; ***$p < 0.01$.

Table 6. Engagement and Perceptions of School Climate

	(1)	(2)	(3)
Nontest Index	0.001	-0.017	0.004
	(0.007)	(0.017)	(0.022)
Survey: Overall	-0.060	-0.042	-0.061
	(0.066)	(0.028)	(0.038)
Survey: Relationships	-0.073	-0.033	-0.045
	(0.061)	(0.024)	(0.034)
Survey: Participation	-0.099*	-0.037	-0.046
	(0.056)	(0.024)	(0.033)
Survey: Instructional Environment	-0.131**	-0.054**	-0.052
	(0.060)	(0.024)	(0.035)
Survey: Discipline Environment	-0.120**	-0.042*	-0.037
	(0.053)	(0.025)	(0.036)
School FE		X	
Grade FE		X	X
Year FE	X	X	
School-Grade FE	X		X
School-Year FE			X

Notes. Coefficients on school DI models from regressions of student behavioral measures and average student survey responses on DI status and specified fixed effects. Nontest index is a factor constructed from log absences, log days suspended, and an indicator for grade promotion. VOCAL survey data are available in 2018 and 2019 for grade 5 and in 2019 for grade 4. Observations are at the school-grade-year level. Departmentalization measures constructed as described in the text. Observations weighted by number of responses. Standard errors clustered by school in parentheses. *$p < 0.10$; **$p < 0.05$; ***$p < 0.01$.

Table 7. Subject-Specific Skill Measures and Subject Assignments

	Teach ELA		Teach Math	
	Non-DI	DI	Non-DI	DI
Panel A. Teacher VA				
ELA VA	0.001	0.020*	0.001	-0.033**
	(0.001)	(0.011)	(0.001)	(0.013)
Math VA	-0.001	-0.033***	-0.000	0.043***
	(0.001)	(0.010)	(0.001)	(0.011)
Observations	15,377	3,424	15,377	3,424
Panel B. Teacher Licenses				
English/History License	0.010**	0.294***	-0.005	-0.355***
	(0.005)	(0.032)	(0.006)	(0.032)
Math/Science License	-0.008	-0.420***	0.004	0.429***
	(0.006)	(0.034)	(0.005)	(0.028)
Observations	43,165	15,858	43,165	15,858

Notes. Regressions of indicators for assignment to teach ELA (Panel A) or math (Panel B) on teacher subject-specific skill measures. ELA/Math VA measure indicates empirical Bayes teacher VA prediction from prior year including student controls, classroom, and school characteristics. Non-DI schools includes all schools that do not currently implement DI. DI schools includes all schools currently implementing DI. All models includes experience indicators and school-grade-year FE. Standard errors clustered by school in parentheses.
**p* < 0.10; ***p* < 0.05; ****p* < 0.01.

Table 8. Teacher Specialization and Productivity

	Math		ELA		Science	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Number of Subjects Taught						
1 Subject	-0.058***	-0.054***	0.017	0.015	-0.002	-0.002
	(0.014)	(0.013)	(0.014)	(0.015)	(0.024)	(0.024)
2 Subjects	-0.010	-0.013	0.001	-0.001	-0.003	-0.003
	(0.009)	(0.009)	(0.010)	(0.010)	(0.016)	(0.016)
3 Subjects	-0.017**	-0.017**	0.008	0.008	-0.008	-0.008
	(0.008)	(0.008)	(0.007)	(0.007)	(0.014)	(0.014)
Observations	944,628	944,628	986,695	992,557	377,983	377,983
Panel B. DI Status						
DI	-0.012	-0.012	0.015**	0.013*	0.012	0.013
	(0.007)	(0.007)	(0.007)	(0.007)	(0.012)	(0.012)
Observations	950,113	950,113	992,557	992,557	380,512	380,512
Fixed-Effects:						
Teacher	Yes		Yes		Yes	
Teacher-School		Yes		Yes		Yes

Notes. Regressions of student achievement on teacher specialization (Panel A) or departmentalization (Panel B) status in elementary school grades 4–6 (math and ELA) or grade 5 (science). Number of subjects taught indicates the subjects taught by the teacher in the given school year. Each observation is a separate student-teacher link. Controls include cubic polynomials in prior achievement in math and ELA, gender, race/ethnicity, limited English proficiency status, special education status, participation in subsidized lunches, and classroom means of the student variables in addition to the specified fixed effects. Standard errors clustered by school in parentheses.
*$p < 0.10$; **$p < 0.05$; ***$p < 0.01$.

Table 9. Effects of DI by Comparative Advantage and Teacher Specialization

	(1) Tests	(2) Math	(3) ELA
Number of teachers	0.012	0.007	0.024*
	(0.012)	(0.014)	(0.012)
Specialists	0.003	0.001	0.007
	(0.012)	(0.013)	(0.012)
In-field license	0.015	0.013	0.016
	(0.016)	(0.018)	(0.016)
Comparative advantage	0.031**	0.025	0.038***
	(0.014)	(0.015)	(0.013)
Proportion grades departmentalized	0.027**	0.029**	0.021*
	(0.012)	(0.014)	(0.011)

Notes. Projection of school-grade-year DI effects on school characteristics. The school-grade-year DI effects are computed for departmentalized schools using the imputation procedure of Borusyak et al. (2024). The sample includes schools that were not departmentalized in 2012 and omits years following a switch back to self-contained from DI. The coefficients represent the projection of the school-grade-year DI effects on school-grade-year characteristics. Observations weighted by student enrollment. Standard errors clustered by school calculated using the method discussed in Borusyak et al. (2024) in parentheses. *$p < 0.10$; **$p < 0.05$; ***$p < 0.01$.

Table 10. Placebo Effects of Instructional Switches on Observable Characteristics

	(1)	(2)	(3)	(4)
	All Changes		Switchers	
English language learner	0.003	-0.002	0.006	-0.001
	(0.003)	(0.002)	(0.005)	(0.002)
Special education	-0.004*	-0.005**	-0.006*	-0.005*
	(0.002)	(0.002)	(0.003)	(0.003)
Male	-0.000	-0.002	0.003	-0.004
	(0.002)	(0.003)	(0.003)	(0.004)
Free/Reduced-price lunch	0.009	-0.000	0.014	-0.001
	(0.007)	(0.003)	(0.010)	(0.005)
Asian	-0.002	-0.002	-0.003*	-0.003
	(0.001)	(0.001)	(0.002)	(0.002)
Black	0.001	0.001	-0.001	0.001
	(0.001)	(0.001)	(0.002)	(0.002)
Hispanic	0.002	0.003	0.005	0.004
	(0.002)	(0.002)	(0.003)	(0.003)
New principal	-0.034*	-0.050**	-0.027	-0.093**
	(0.019)	(0.025)	(0.025)	(0.041)
New principal (Year *t*-1)	-0.004	0.006	0.025	0.058
	(0.021)	(0.028)	(0.024)	(0.045)
New principal (Year *t*-2)	0.027	0.027	0.013	-0.007
	(0.018)	(0.028)	(0.023)	(0.040)
New principal (Year *t*-3)	-0.009	-0.032	0.005	-0.013
	(0.021)	(0.034)	(0.024)	(0.045)
N	16636	14543	16636	14543
FE/FD	FE	FD	FE	FD

Notes. Regressions of time-variant characteristics on DI indicators. Regressions in columns (1) and (2) use the DI treatment indicator described in the text. Regressions in columns (3) and (4) use an indicator for schools that switch from self-contained to DI during the sample period. Models in odd-numbered columns use panel fixed effects (FE); models in even-numbered columns use panel first differences (FD). Observations weighted by student enrollment. Standard errors clustered by school in parentheses. *$p < 0.10$; **$p < 0.05$; ***$p < 0.01$.

Table 11. Estimated Effects of Departmentalization under Alternative Specifications

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A. All Switches							
Math + ELA	0.026***	0.016	0.027***	0.017*	0.026***	0.026	0.029*
	(0.009)	(0.012)	(0.009)	(0.010)	(0.008)	(0.016)	(0.016)
Math	0.020*	0.001	0.020*	0.010	0.019**	0.014	0.015
	(0.010)	(0.014)	(0.011)	(0.012)	(0.009)	(0.018)	(0.016)
ELA	0.035***	0.033***	0.035***	0.028***	0.035***	0.039**	0.043**
	(0.009)	(0.012)	(0.009)	(0.009)	(0.009)	(0.017)	(0.018)
Science	0.039***	0.049**	0.039***	0.040**	0.034**	0.053**	0.004
	(0.013)	(0.021)	(0.014)	(0.017)	(0.015)	(0.024)	(0.035)
Panel B. DI Switches							
Math + ELA	0.052***	0.037**	0.053***	0.045**	0.052***	0.041**	0.053*
	(0.016)	(0.017)	(0.016)	(0.018)	(0.015)	(0.019)	(0.029)
Math	0.038**	0.017	0.039**	0.036*	0.038**	0.022	0.026
	(0.018)	(0.019)	(0.018)	(0.020)	(0.016)	(0.022)	(0.029)
ELA	0.067***	0.059***	0.069***	0.058***	0.069***	0.062***	0.079**
	(0.015)	(0.017)	(0.016)	(0.017)	(0.015)	(0.020)	(0.031)
Science	0.046**	0.066***	0.047**	0.083**	0.031	0.068**	-0.001
	(0.020)	(0.025)	(0.021)	(0.035)	(0.024)	(0.027)	(0.049)
Controls	X	X					
Grade Trends			X				
District Trends				X			
Principal Trends					X	X	
Stable Principal							X
CH22		X				X	

Notes. Coefficients on school DI models from regressions of mean student outcomes on DI status. Sample includes students in grades 3–6 in elementary schools. DI treatment indicators constructed as described in the text. Controls include student gender, race/ethnicity, free and reduced-price lunch status, limited English proficiency status, special education status, and test mode (paper/online) and type (MCAS/PARCC) indicators. Grade trends indicate that year fixed effects (FE) have been replaced by grade-by-year FE. District trends indicate the inclusion of district-by-year FE. Principal trends indicate the presence of principal mobility group by year FE. Principal mobility groups are defined as a set of schools with principal changes in the same school year(s). Stable principal indicates the sample consists of a set of schools without principal changes between 2012 and 2019. CH22 = de Chaisemartin & D'Haultfœuille (2022). Observations weighted by student enrollment. Standard errors clustered by school in parentheses. *$p < 0.10$; **$p < 0.05$; ***$p < 0.01$.

Table 12. Triple-Differences (TD) Estimates of Effects of DI

	(1) Tests	(2) Math	(3) ELA
Baseline TD	-0.000	-0.006	0.010
	(0.008)	(0.009)	(0.009)
Full/Partial School DI			
Partial School DI (DID)	0.007	0.003	0.015
	(0.009)	(0.010)	(0.010)
Full School DI (DID)	0.046***	0.036***	0.056***
	(0.012)	(0.014)	(0.012)
Partial School (TD)	-0.000	-0.006	0.010
	(0.008)	(0.009)	(0.009)

Notes. Coefficients on school DI models from regressions of mean student outcomes on DI status and specified fixed effects. Sample includes students in grades 3–6 in elementary schools. DI treatment indicators constructed as described in the text. Average math and ELA is the average standardized test score in both end-of-grade tests. Nontest index is a factor constructed from log absences, log days suspended, and an indicator for grade promotion. CH22 = de Chaisemartin & Haultfœuille (2022). Observations weighted by student enrollment. Standard errors clustered by school in parentheses. *$p < 0.10$; **$p < 0.05$; ***$p < 0.01$.

Appendix A. Data Appendix

A.1. Characterizing School Instructional Models

We identify school instructional models using data on student schedules and teacher assignments. The two datasets are linked by a common set of unique course identifiers. We first limit the sample to classes in elementary schools in Massachusetts identified in the NCES Common Core of Data. Because we only have achievement outcomes for students in grades 3 and higher, we drop schools with highest grade of 3 or lower.[18]

We then restrict the set of student-teacher course matches to keep records associated with core-subject teachers using data on teaching assignments. We first keep records associated with teaching, co-teaching, and long-term substitute roles. We then restrict the sample based on the nature of the teaching assignment. The administrative data identify instructors of core subject content, teachers of core content for students with disabilities or English language learners, and resource or consultative teachers. We keep all teaching assignments associated with a primary school core academic subject assignment.

Finally, we restrict the sample to include only course codes associated with the primary core subjects in ELA, math, science, and social studies in grades 3–6 (National Forum on Educational Statistics [NFES], 2011). The course codes include a set of codes for non-differentiated instruction: "[C]ourses that are not differentiated by subject area—that is, instances in which students are enrolled in a grade-specified course and are taught various subjects throughout the day, rather than being enrolled in subject-specific courses" (NFES, 2011). These codes are typically, but not always, used for self-contained classrooms. Because they are

[18] There are very few schools that departmentalize in grade 3, so we drop schools with a terminal grade of 3 for this analysis.

sometimes used for other class assignments, we drop these courses if a student is assigned to a separate course in each of the four core subject areas.

We use this dataset to categorize school models. We first construct two variables: (1) the total number of classes in which each student is enrolled and (2) the total number of classes in which each student is linked to each of their teachers. We consider a student as being assigned to a self-contained classroom if the number of classroom assignments to any of their teachers is equal to the number of total classroom assignments. We take the average proportion of self-contained assignments at the school-grade-year level and categorize any cell with 50% or fewer self-contained assignments as departmentalized.

We then further categorize departmentalized schools into two submodels based on classroom assignment patterns. The common conceptualization of DI is schools in which teachers share responsibility for the main core subject areas (e.g., one teacher takes math and science; another teacher takes ELA and social studies). Some schools alternatively have a single teacher who takes responsibility for several classes in one subject only. In our sample, these teachers typically instruct writing or science. We therefore identify students who have a single teacher for at least three of the four core subjects and one or more teachers who teach only a single subject. We consider these students to be in *partial DI* models. We average the incidence of these assignments to the school-grade-year level and assign any school with more than 50% of students in a partial DI model to be partially departmentalized. The remaining schools, which comprise 86% of the departmentalized sample, we consider to be *fully DI*.

Although we use this student-teacher linked dataset to identify instructional models, we make no sample restrictions when constructing datasets with student outcomes.[19] That is, we do

[19] As described in the text, we limit the sample to include only schools with balanced panels (i.e., those operating and reporting student assignments in all years).

not limit the sample to students who can be matched to core subject teachers in mainstream educational settings.

A.2. Teacher Specialization Sample

We construct a separate teacher-level dataset to consider the effects of departmentalization and specialization on individual teacher productivity in Section 7. We start with the linked student-teacher dataset described in Section A.1 and make one revision to account for self-contained classroom assignments. If a student is assigned to a non-differentiated classroom in this dataset, we assume that the teachers of this class are responsible for instruction in all subjects in which the student does not have another classroom assignment. For instance, if a student has a non-differentiated classroom assignment with teacher A in classroom X and a science assignment with teacher B in classroom Y, we assume that teacher A instructs the student in ELA, math, and social studies in classroom X.

Using these data, we construct measures of teacher specialization following Bastian and Fortner (2020). We identify generalist teachers as those teaching three or four subjects in a particular school year and specialist teachers as those teaching one or two subjects in a particular school year. Note that these definitions refer to teaching *assignments* and not teaching *qualifications*, which we consider elsewhere in the paper.

Appendix B. Alternative Definitions of School Switches

In the full dataset, we observe 528 school-grade panels that switch between DI and self-contained models. Of these switches, 170 (32%) switch back to the original model in the following school year. These switches may be true switches; for instance, it could be the case that these schools had poor experiences with DI (or self-contained classrooms) and reverted to the original model after a year. Qualitative studies also report that schools sometimes engage in pilots of DI before deciding on an instructional model (Haley, III, 2018; Strohl et al., 2014; Parker et al., 2017). Alternatively, these may be data errors. We dichotomize a continuous variable (the proportion of students in self-contained classrooms) to construct the DI treatment indicator, and the temporary switchers might be schools with nearly half of their students in self-contained classrooms that experience small year-to-year fluctuations in their student assignment patterns. These switches may also reflect errors in the student- or-teacher scheduling data reported to Massachusetts. In either case, the experiences of these schools may not be representative of schools making more permanent changes to or from DI.

In Table B.1, we explore the observed staffing changes in these schools following changes in instructional models. Specifically, we recreate the results from Table 3 and allow the effect of DI on student assignment patterns to differ by switch type. The results do not suggest that the short switches are a result of dichotomizing the proportion of self-contained assignments: The coefficient on DI for the 1-year switches (-0.81) is nearly as large as that for more permanent switches (-0.88). We also find similar, albeit somewhat attenuated, effects on the number of teachers and the proportion of specialists. Although we cannot rule out that these switches are the result of reporting errors in the scheduling data, they do not appear to be an artifact of how the treatment variable is constructed.

Table B.1. Effects of Departmentalization on Student Assignments by Switch Duration

	(1) Self-Contained	(2) Teachers	(3) ELA Specialist	(4) Math Specialist	(5) Science Specialist	(6) Social Studies Specialist
DI > 1 Year	-0.88***	1.10***	0.39***	0.34***	0.35***	0.30***
	(0.01)	(0.03)	(0.02)	(0.03)	(0.03)	(0.02)
DI 1 Year	-0.81***	0.93***	0.24***	0.22***	0.34***	0.23***
	(0.02)	(0.07)	(0.03)	(0.03)	(0.04)	(0.03)
N	16680	16680	16680	16680	16680	16680

Notes. Estimates of the effect of departmentalization on student assignment patterns based on duration of switch. DI > 1 Year indicates the interaction between current DI status and an indicator for school-grade cells that do not have two transitions in 2 consecutive school years. DI 1 Year indicates the interaction between current DI status and an indicator for school-grade cells observed with two transitions in 2 consecutive school years. All models include school-grade and grade-year fixed effects. Standard errors clustered by school in parentheses.
*$p < 0.10$; **$p < 0.05$; ***$p < 0.01$.

In Table B.2, we re-estimate the baseline models after dropping any school-grade cell that changes models in 2 consecutive school years. The results are all quite similar to the baseline results shown in Table 4 and indicate that the results are not likely driven by measurement error in the school switches.

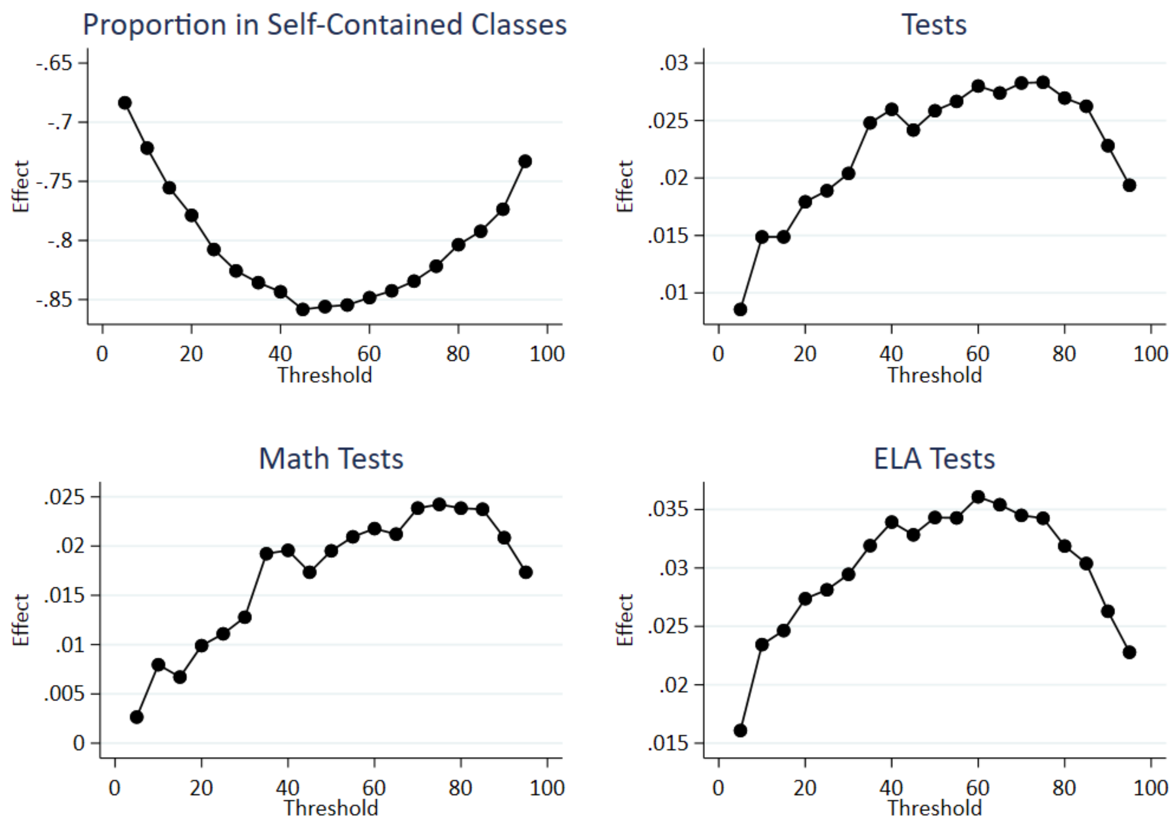
Table B.2. Estimates of the Effects of Departmentalization Omitting 1-Year Switches

Panel A: Average Math and ELA						
DI	0.010	0.031***	0.031***	0.031***	0.007	0.010
	(0.009)	(0.011)	(0.011)	(0.010)	(0.009)	(0.013)
Observations	15320	15320	15320	15320	14856	
Panel B: Math						
DI	-0.005	0.025**	0.025**	0.025**	0.007	-0.002
	(0.011)	(0.012)	(0.012)	(0.012)	(0.010)	(0.013)
Observations	15320	15320	15320	15320	14856	
Panel C: ELA						
DI	0.027***	0.041***	0.041***	0.041***	0.013	0.024
	(0.009)	(0.011)	(0.011)	(0.011)	(0.011)	(0.015)
Observations	15320	15320	15320	15320	14856	
Panel D: Science						
DI		0.052***		0.052***		0.043*
		(0.016)		(0.016)		(0.024)
Observations		3920		3920		
Panel D: Nontest Index						
School FE	X					
Year FE		X				X
School-Grade FE		X	X	X	X	X
Grade-Year FE	X		X	X	X	
School-Year FE					X	
Controls				X		

Notes. Coefficients on school DI models from regressions of mean student outcomes on DI status and specified fixed effects. Sample includes students in grades 3–6 in elementary schools and excludes all schools that switch models in 2 consecutive school years. DI treatment indicators constructed as described in the text. Average math and ELA is the average standardized test score in both end-of-grade tests. Nontest index is a factor constructed from log absences, log days suspended, and an indicator for grade promotion. CH22 = de Chaisemartin & Haultfœuille (2022). Observations weighted by student enrollment. Standard errors clustered by school in parentheses.
$^*p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$.

In the empirical analyses, we classify schools based on the proportion of students in self-contained classes. We characterize schools as departmentalized if at least half of students in the schedule data are not assigned to self-contained classes. As shown in Figure 1, the distribution of the proportion in self-contained classes is bimodal with peaks at zero and one (86% of observations take one of these two values). However, the treatment of the remaining 14% of observations is necessarily somewhat arbitrary. We use the modal instructional assignment in each cell; in this section, we show that the results are not sensitive to reasonable departures from this rule. In Figure B.1, we plot results using alternative thresholds. The x-axis depicts the percentage of students in self-contained classes used to assign schools to DI. The y-axis depicts the estimated DID effect using our preferred specification.

Figure B.1. Difference-in-Differences Estimates Using Alternative Thresholds



Notes. Estimated difference-in-differences effects on specified outcomes using alternative thresholds for identifying departmentalized schools. Threshold indicates proportion of students in self-contained classes used to differentiate instructional model (DI schools are defined as those with fewer students in self-contained classes than the threshold number).

In the first panel, we show the effects of using different thresholds on the proportion of students in self-contained classes (the continuous measure). The magnitude of this coefficient peaks at 45% and declines as the threshold approaches 0% or 100%. This pattern reflects the fact that using different thresholds tends to capture more idiosyncratic variation in student assignments across classrooms, and suggests that estimated treatment effects should tend to be somewhat attenuated using different thresholds.

This is indeed what we see in the remaining columns. Although thresholds near 50% tend to yield similar point estimates, the estimated effects on test scores tend to be somewhat smaller using thresholds below 30% or above 80%. This likely reflects the fact that more or less stringent thresholds incorporate many false switches resulting from minor changes in student assignment patterns (as in the first panel).