Working Paper 2020–1
FALL 2020

UPDATED WINTER 2022

Are Two Teachers Better Than One?

The Effect of Co-Teaching on Students With and Without Disabilities

Nathan Jones, PhD Associate Professor Boston University

Marcus A. Winters, PhD Associate Professor Boston University

January 2022

Boston University Wheelock College of Education & Human Development
Wheelock Educational Policy Center

WHEELOCK EDUCATIONAL POLICY CENTER

Are Two Teachers Better Than One?

The Effect of Co-Teaching on Students With and Without Disabilities

Nathan Jones, PhD* Marcus A. Winters, PhD†

Abstract

Co-teaching, in which a general education teacher and special education teacher collaboratively provide instruction to students with and without disabilities in the same classroom, is widely endorsed as a strategy to give instructional support to students with disabilities within inclusive environments. We leverage longitudinal administrative data in Massachusetts to provide the first causal estimate for the effect of co-teaching across a large public school system. We find evidence that co-teaching leads to statistically significant test score improvements for both students with and without disabilities. However, the benefits for students with disabilities are much smaller than reported in prior studies.

Acknowledgements: The data can be obtained by filling a request directly with the Massachusetts Department of Elementary and Secondary Education (MADESE). The authors are willing to assist. Neither author received financial support or has a conflicting interest. We extend our gratitude to the Massachusetts Department of Elementary and Secondary Education, and especially Carrie Conaway, Matthew Deninger, and Kendra Winner for their cooperation and support for this project. We also would like to thank Allison Gilmour, Tim Sass, Jesse Bruhn, Thomas Pearson, participants at the Association for Public Policy Analysis and Management annual conference, and the Half Baked Seminar at The Annenberg Institute at Brown University for their helpful comments. Cheonghum Park provided excellent research assistance. All remaining errors are our own.

*Nathan Jones is an associate professor of special education at the Boston University Wheelock College of Education & Human Development.

†Marcus A. Winters is an associate professor and chair of the Educational Leadership & Policy Studies department at the Boston University Wheelock College of Education & Human Development. He is also faculty director of the Wheelock Educational Policy Center. marcusw@bu.edu

1 Introduction

About 13.2% of public school students in the United States receive special education services.[1] Federal special education law prioritizes two concurrent goals for students with disabilities (SWDs). Instruction should be individualized – specially designed to meet students' individual academic and behavioral needs. At the same time, instruction should be inclusive – ensuring that SWDs have access to the general education curriculum and a highly-qualified content area teacher. At $13 billion annually, providing special education services designed to achieve these goals is the second largest category of federal funding for K12 schools. That SWDs nonetheless achieve substantially worse educational outcomes than students without disabilities on average has led policymakers to consider alternative strategies.

Co-teaching, in which a general education teacher and special education teacher share teaching responsibilities for students with and without disabilities in the same classroom (Friend, 2008), has gained wide support in the special education literature (e.g., Friend et al., 2010; Friend, 2015) and is recommended by several state departments of education. However, though the logic behind co-teaching is intuitively appealing, prior research suggests that it is rarely implemented in the idealized form envisioned by its advocates (Cook et al., 2011; Scruggs et al., 2007; Wexler et al., 2018). We also currently know little about the impact of co-teaching on student learning, especially as implemented within large public school systems. Each of the handful of quantitative studies examining the effectiveness of co-teaching suffer from severe methodological limitations and/or are limited to only one or a few schools or classrooms (Murawski and Swanson, 2001; Cook et al., 2011).

In this paper, we employ a student fixed-effect approach in order to provide the first causal estimate for the effect of co-teaching on student test scores across a large public school system. Leveraging longitudinal administrative data on the universe of public school students and teachers in Massachusetts, we find that attending a co-taught classroom on average leads to test score improvements of about 0.016σ in English Language Arts (ELA) and 0.026σ in math for

[1] National Center for Education Statistics, Digest of Education Statistics, 2017, Table 204.30

SWDs. For SWDs, the effect of co-teaching on math scores is significantly larger in middle school grades (0.036σ) than in elementary grades (0.014σ). For students without disabilities, attending a co-taught classroom improves test scores by about 0.012σ in math but has no significant effect in ELA.

In addition to expanding upon the limited prior research on the effect of co-teaching on student outcomes, our study contributes to several lines of research.

First, our results add to a surprisingly small but emerging literature measuring the impact of resource allocation and changes in settings to educate SWDs. Recent studies find that receiving special education services causes an increase in a student's test scores (Hanushek et al., 2002; Ballis and Heath, 2021a; Schwartz et al., 2019) and educational attainment (Ballis and Heath, 2021b). In addition, Setren (2019) finds that SWDs benefit from receiving more general educational services within high-performing charter schools.

Since co-teaching undoubtedly reduces the student-teacher ratio within a classroom, our results also relate to the expansive literature on the effect of reducing class size on student educational outcomes. Prior studies that exploit exogenous variation in class size tend to find short-run improvements in student test scores (Krueger, 2003; Angrist and Lavy, 1999; Hoxby, 2000; Leuven et al., 2008) and non-cognitive outcomes (Blatchford et al., 2011; Dee and West, 2011; Finn et al., 2003), though the evidence on the longer run effects is more mixed (Krueger and Whitmore, 2001; Chetty et al., 2011; Fredriksson et al., 2013; Falch et al., 2017; Leuven and Løkken, 2020; Woessmann and West, 2006). However, it is important to note that co-teaching is unlikely to be equivalent to simply dividing a classroom in half. Though in theory co-teachers are meant to take equal responsibility for all students in the classroom, in practice the general education teacher often assumes primary instructional responsibilities and the special educator provides individual or small-group support specifically to the SWDs (Cook et al., 2011; Scruggs et al., 2007; Wexler et al., 2018). Nonetheless, even if co-teaching relationships are not implemented as intended, the support of another trained adult in the classroom could yield benefits (Hemelt et al., 2021; Andersen et al., 2020).

Finally, our analysis contributes to the wide body of research on policies intended to

improve teacher effectiveness. The teacher quality literature primarily focuses on the impact of individual teachers (Jones et al., 2019) and ways to improve their performance through additional training (for example, Kraft et al., 2018). Co-teaching is among a class of reforms, such as using technology to aid instruction, that alter the allocation of resources in a way that could better leverage current teachers' abilities.

The remainder of this paper proceeds as follows. Section 2 provides a brief introduction to co-teaching and an overview of existing scholarship on the topic. Section 3 describes the Massachusetts administrative data. Section 4 outlines our strategy for estimating the impact of co-teaching on student test scores. Section 5 reports our results and provides some specification tests. Finally, Section 6 summarizes the results and concludes.

2 Co-Teaching

Co-teaching allows schools to simultaneously meet multiple federal mandates related to educating SWDs. First, co-teaching allows schools to provide instruction in the least restrictive environment, as mandated by IDEA. Second, co-teaching allows schools to satisfy federal requirements associated with standards-based reforms. The No Child Left Behind Act (2002) included the requirement that SWDs receive instruction from highly qualified teachers in all academic content areas and make progress in the general education curriculum. If a SWD were educated by a general educator alone, they may not have adequate access to differentiated support. And, if a SWD were educated by a special educator alone, they may not have access to relevant content expertise.

Most prior studies of co-teaching focus on recommendations for its implementation (e.g., Murawski and Dieker, 2004; Sileo, 2011) using information from observations of existing co-teaching programs (e.g., Dieker, 2001; Mastropieri et al., 2005) and interviews and surveys with practicing co-teachers (e.g., Austin, 2001; Keefe and Moore, 2004). As such, this literature is largely concentrated in two main areas: a) co-teaching program logistics[2] and b) co-teaching rela-

[2]This literature has focused on the importance of establishing and maintaining a strong co-teacher relationship (e.g., Bessette, 2008; Gately and Gately, 2001; Keefe and Moore, 2004; Mas-

tionships.[3]

On surveys and interviews teachers who participate in co-teaching report positive perceptions of it (Scruggs et al., 2007; Welch, 2000; McDuffie et al., 2009). In particular, teachers have described greater academic benefits for students related to co-teaching over instruction in general education classes taught by a single general educator (e.g., Rice and Zigmond, 2000; Austin, 2001). They attribute these academic benefits to the increase of individualized attention to students, additional teacher time, and a reduction in the student-teacher ratio (Walther-Thomas, 1997; Trent, 1998; McDuffie et al., 2009). Teachers also report that students without disabilities benefit from the type of instruction offered in co-taught classes (Trent, 1998).

While largely supported by educators, co-teaching's effectiveness for improving student outcomes depends on a key assumption – that the presence of a second adult results in more individualized learning opportunities for SWDs. Unfortunately, the few observational studies on co-teaching suggests that this is unlikely to occur in practice (Scruggs et al., 2007; Wexler et al., 2018; Cook et al., 2011).

The evidence base regarding the effect of co-teaching on measurable student outcomes is sparse. The few studies investigating the effect of co-teaching on student achievement yield positive results, with some reporting that it improves SWDs' test scores by between a quarter and a full standard deviation (Cook et al., 2011). However, the three prior studies that used an experimental

tropieri et al., 2005; Ploessl et al., 2010). Often referred to as a professional marriage, the relationship between co-teachers is described as evolving over time as teachers work together to solve problems as they arise (Gately and Gately, 2001; Sileo, 2011).

[3]Practitioner-focused journals are replete with "how-to" articles and books offering recommendations for developing co-teaching programs (Murawski, 2005; Murawski and Dieker, 2004), planning and implementing instruction in the co-taught classroom (e.g., Vaughn et al., 1997; Wilson, 2008; Conderman and Hedin, 2014) maintaining productive co-teaching relationships (e.g., Kohler-Evans, 2006; Sileo, 2011; Pratt, 2014), supervising co-teachers (e.g., Walther-Thomas et al., 1996; Wilson, 2005; Nierengarten, 2013), and evaluating co-teachers (Murawski and Lochner, 2011).

design to measure co-teaching impacts took place within a single classroom or school (Murawski, 2006; Fontana, 2005), and other studies either compared student outcomes across existing instructional models without leveraging any sort of exogenous variation (McDuffie et al., 2009; Hang and Rabren, 2009; Molnar et al., 1999) or included no comparison group (Klinger et al., 1998).

Existing research provides little data on co-teaching's prevalence overall or variation in its use by subject area or grade level. We also do not know how decisions to assign students to co-taught classrooms get made or the overall costs associated with implementing co-teaching within a school or district. Existing research has not investigated the extent to which co-teaching placements change across a child's years of schooling. Implementing co-teaching has implications for the master schedule, including changes to special educators' workloads and for supporting common planning time (Murawski and Bernhardt, 2015), and it also likely requires professional development for educators, as well as potential hiring of additional personnel to meet the needs of all students.

We anticipate that the effectiveness of co-teaching might differ when applied in elementary or secondary classrooms. Because elementary teachers work in the same classroom with the same group of students each day, they may be more prone to the kinds of deep, collaborative co-teaching relationships recommended in the special education literature.

We also explore the effects of co-teaching by student characteristics, including students' disability subcategories and whether there are differential effects across students with and without disabilities. Existing evidence suggests that we cannot treat student disability categories interchangeably (e.g., Gilmour et al., 2019; Gilmour and Henry, 2018; Schulte and Stevens, 2015). Students from some "high-incidence disabilities" (e.g., specific learning disabilities, speech or language impairment, other health impairment) typically need less individualized support and are more likely to receive instruction in the general education classroom. Others, including students with autism and emotional behavioral disorders (EBDs), may need additional support for disruptive behavior (Crosland and Dunlap, 2012; Harrison et al., 2012; Lane et al., 2006). Importantly, some evidence suggests that students without disabilities may exhibit lower academic performance when they have classmates with disabilities, particularly EBDs (Fletcher, 2010; Gottfried, 2014;

Gottfried et al., 2016; Gottfried and Harven, 2015). By adding another teacher into the classroom, co-teaching could mitigate or perhaps even reverse any negative effect of inclusion on general education students.

2.1 Co-Teaching in Massachusetts

The Massachusetts Department of Elementary and Secondary Education (DESE) encourages co-teaching but has not adopted it as a statewide initiative. In most cases, schools determine whether to adopt co-teaching and the extent to which they implement co-teaching for all or part of the school.[4]

Figure 1 illustrates the growth in co-teaching within Massachusetts from 2011 through 2018. The percentage of students in co-taught classrooms increased substantially over this time, with the largest gains occurring between 2014 and 2018. As we would expect, a higher proportion of SWDs than students without disabilities are enrolled in co-taught classrooms. Table B1 in the Online Appendix shows that the growth in co-teaching over this time period for both students with and without disabilities was similar across grades. Appendix Figure C1 illustrates that the growth in co-teaching occurred throughout the state.

[FIGURE 1 ABOUT HERE]

In most cases, schools do not adopt co-teaching as a uniform policy that applies to all students but instead offer a mixture of co-taught and single-teacher courses. Figure 2 illustrates changes in the proportion of students with and without disabilities taught within co-taught classrooms by school from 2011 through 2018.[5] Many schools that exhibited little or no co-teaching in 2011 taught substantial portions of students within co-taught classrooms in 2018. Figure C3 in the Online Appendix further illustrates the growth of co-teaching across schools by rank-ordering schools by their share of students in co-taught classrooms at three points in time. In 2011, more

[4]We confirmed the basic understanding of how co-teaching is implemented across the state with officials at DESE.

[5]Figure C2 in the Online Appendix presents a similar illustration for all students in the school.

than 90 percent of schools in the state educated no SWDs in co-taught classrooms; in comparison, in 2018, in nearly 30 percent of schools, at least some SWDs were co-taught. There is substantial variation in assignment to co-taught classrooms both across schools over time and within schools during a given year.

[FIGURE 2 ABOUT HERE]

There is also substantial variation in attending a co-taught classroom for particular students over time. Table B2 in the Online Appendix reports the average percentage of years that a student is in a co-taught classroom following the first year that they were co-taught. Most students who enter a co-taught class do not remain in that environment for the rest of their academic career but rather switch between co-taught and single-teacher classrooms over time.

3 Data

We utilize longitudinal administrative data for the universe of students and teachers from 2007-2008 through 2017-2018 made available by DESE. In the analyses, we include data only on students enrolled in grades three through eight, in which students are administered the statewide standardized test in math and ELA, known as the Massachusetts Comprehensive Assessment System (MCAS).

The data consists of four separate datasets: Student-level demographic and socioeconomic variables come from the Student Information Management System (SIMS) data, student test scores in math and ELA come from Massachusetts Comprehensive Assessment System (MCAS) data, teachers' job assignment classifications come from Education Personnel Management System (EPIMS) data, and information on student classroom assignments necessary to link students with their teachers comes from Student Course Schedule (SCS) data. For a detailed description of the data cleaning process, see the Online Appendix.

The EPIMS data places employees within 67 distinct job classifications. Among the seven job classifications categorized as Instructional Staff are separate codes for Teacher and for

Co-Teacher. A teacher is defined as providing instruction, learning experiences, and care to students during a particular period or in a given discipline. A co-teacher is defined as a teacher, equally responsible with another teacher, for providing instruction, learning experiences, and care to students during a particular time period or in a given discipline. The state separately classifies other instructional staff (e.g. Teacher – support content instruction and Instructional Coach) as well as other distinct job classifications for employees who provide instructional support (e.g. paraprofessionals) and special education shared staff (e.g. speech pathologist).[6]

Using these data, we need to construct a student-year-level dataset with student test scores and indicators for the co-teaching status of the classrooms included. To accomplish the task, we first filter the EPIMS data to only include educational personnel who are assigned to math and ELA courses and have a job classified as either "teacher" or "co-teacher". For primary school grades, we also keep non-differentiated courses that cover all content instruction. We then drop observations that correspond to courses taught in alternative education programs to facilitate merging in the MCAS data. The co-taught classrooms are identified here as multiple teachers being assigned to a single year-course-section-term observation. We then link the EPIMS data to the SCS data by year, school, course, section, and term of instruction. We fail to match 0.03% of staff-course-year observations to any student-course-year observation in the SCS data. We drop these observations. From there, we merge in student demographic characteristics from SIMS using the SASID, year, school, and district. Approximately 7% of the sample of student-course-year observations do not have a corresponding student match in the SIMS data. The majority of these matching failures occur in 2011, which is the first year the state tracked student course work. We drop the unmatched observations. We keep one course each in math and ELA for each student-year observation with a priority to a co-taught course. We then separate the student-subject-year data to math and ELA samples and merge them with the MCAS data. We fail to match 8% and 9% of the student-year observations to their test scores in each sample, respectively. We also drop these unmatched observations.

[6]We do not observe consistent information on whether the teacher is certified to teach special education.

We emphasize that even in the best case we are not able to observe the practices of teachers working within co-taught classrooms. Thus, our findings should be viewed as the estimated effect of co-teaching on average as implemented statewide in Massachusetts, which we argue has important implications for both policy and practice. Whether co-teaching is more or less effective when implemented in particular ways is an area for future research.

Table 1 reports relevant descriptive statistics for students with and without disabilities according to whether they are observed in a co-taught classroom sometimes, never, or always within our data. The table only reports data for students within schools in which we observe any co-taught classrooms during the sample period. In addition to observed characteristics, for students who are observed in both co-taught and single-teacher settings the table reports the average percentage of their observations that are in a co-taught classroom.

The table allows for two types of comparisons. First, within each classification type we can compare the characteristics of students we observe in both co-taught and single-teacher settings to those of students who are always or never co-taught. This is a valuable comparison because, as we discuss in the next section, our primary strategy employs student fixed-effects and thus only students in the Sometimes category contribute to the estimate for the effect of co-teaching. For both SWDs and students without disabilities we observe several statistically significant but very small demographic differences between those exposed to both classroom types and students who are always or never co-taught. Students who are never in a co-taught classroom have substantially higher math and ELA scores on average.

Table 1 also compares the observed characteristics of SWDs and students without disabilities who are sometimes, never, or always observed in a co-taught class. Within the "sometimes" category, SWDs and students without disabilities differ significantly on each shared observed characteristic. However, most of the demographic differences are again quite small, with the exceptions that SWDs are substantially more likely to be male and eligible for free lunch. As expected, students without disabilities have significantly higher math and ELA scores than SWDs. On average, SWDs attend slightly smaller classes. The percentage of observations that the student is in a co-taught classroom is about 2 percentage points higher for SWDs than for students without

disabilities.[7]

[TABLE 1 ABOUT HERE]

4 Empirical Method

Our goal is to estimate the causal effect of a student attending a co-taught classroom instead of a classroom headed by a single teacher on the student's standardized test scores. Our ideal strategy would compare the later test scores of students who were randomly assigned to attend a co-taught or single-teacher classroom. However, in practice schools introduce co-teaching, and students within schools are likely assigned to co-taught classrooms non-randomly. We suspect that a naive comparison of the outcomes of students who enrolled in co-taught and single-teacher classrooms would be biased by two sources of selection. First, there could be differences in the characteristics and leadership of schools that adopt co-teaching and those that do not. Second, schools might assign students to co-taught classrooms based on their academic ability or other attributes.

Our primary identification strategy addresses these potential areas of selection by leveraging cross-student variation in the timing of assignment to a co-taught classroom within a regression model that holds constant student and school fixed effects. The model addresses the first source of selection bias by differencing out variation that is fixed across schools and addresses the second source of selection by differencing out fixed attributes for each student.

Our preferred regression model takes the form:

[7]Table B3 in the Online Appendix compares the observed characteristics of students within schools where we observe at least some co-teaching and schools in which we do not observe co-teaching during the sample period. There are again several statistically significant differences, including that schools that enroll more Black students or student eligible for free lunch are slightly more likely to engage in co-teaching. However, none of these differences are of a meaningful magnitude.

$$y_{ist} = \alpha + \sum_{j=1}^{J} \psi_j X_{ist} + \ \delta_i + \ \phi_s + \beta \ coteach_{ist} + \ \epsilon_{ist} \tag{1}$$

where y_{ist} is the standardized test score of student *i* while enrolled in school *s* during year *t*; δ_i and ϕ_s are student and school fixed effects; X_{it} is a vector of time-variant observed characteristics for the student; $coteach_{ist}$ is a dummy variable indicating whether the student attended a co-taught classroom; ϵ_{ist} is a stochastic term clustered by student; and β, ϕ, and ψ_j are parameters to be estimated. We are primarily interested in the coefficient for β, which under the identifying assumptions represents the causal effect of attending a co-taught classroom on student test scores.

Because there is reason to believe that the effect of co-teaching could differ for younger and older students, we also estimate models that incorporate an interaction between co-teaching and whether a student is enrolled in grades three through five ($elem$). Formally:

$$y_{ist} = \alpha + \sum_{j=1}^{J} \psi_j X_{ist} + \ \delta_i + \ \phi_s + \phi elem_{ist} + \ \beta \ coteach_{ist} + \ \lambda \left(coteach_{ist} \times elem_{ist}\right) + \epsilon_{ist} \tag{2}$$

We estimate both equations (1) and (2) separately in samples restricted to include only SWDs or only students without disabilities. We classify a student as a SWD if at any point in our data we observe them with an IEP, and thus each student is included in only one sample.[8] Though a student can be newly classified or declassifed out of special education at any time in their academic career, the large majority of disability classifications are made prior to the third grade. Table B4 in the Online Appendix shows that there are only small differences in the percentage of students in special education within the state between the third and eighth grade. In addition, we restrict the sample to include only students who we observe enrolled at least once in both a co-taught and a single-teacher classroom, because students who are observed in only one setting play no role in identifying either β or λ due to the student fixed-effect.

[8]Table B5 shows that instead classifying students as SWD only during years that they are in special education does not meaningfully impact our results.

Our analysis compares a student's performance in a co-taught class to their performance in a class that is not co-taught, thus exploiting variation in exposure to co-teaching within a student over time. For SWDs, the counterfactual classroom headed by a single teacher might be either a general education class or a class that is specific to SWDs. We argue that the most interesting counterfactual for policy purposes is the class that a student would have attended otherwise under the implementation of co-teaching within the Massachusetts. However, in section 4.2 we discuss an alternative approach that alters the analysis for SWDs to instead compare attending a co-taught classroom separately to attending a special-education-specific (SE-specific) class or a general education classroom.

4.1 Identification

Causal interpretation of β and λ relies on the assumption that there are no unaccounted-for time-variant variables that are correlated both with the timing of a student's assignment to a co-taught classroom and their test score at the end of that year. Formally, the identifying assumption for equation (1) is that:

$$E[y_{ist}|X_{it}, \delta_i, \phi_s, \gamma_t, coteach_{ist}] = E[y_{ist}|X_{it}, \delta_i, \phi_s, \gamma_t] \tag{3}$$

That is, our analysis essentially assumes that, conditional on the other covariates, the timing of assignment to a co-taught classroom in a student's academic career is as good as random. The most substantial threat to our identification strategy is the possibility that students are assigned to co-taught classrooms based on time-variant attributes. For instance, administrators might assign a student to a co-taught classroom after an especially good or bad academic year. Because students move in-and-out of co-teaching environments we cannot look for a lack of a trend in student outcomes prior to entering a co-taught classroom as in a conventional event-study analysis. We instead adopt two approaches to investigate the plausibility that our estimates can be given a causal interpretation.

4.1.1 Regression on Observables

Our first strategy is to evaluate the relationship between observed characteristics and the likelihood that a student is assigned to a co-taught classroom. We are especially interested in the relationship between the student's test score in the prior year and enrollment in a co-taught classroom both because test score is the outcome of interest and also because it is the only time-variant characteristic that we observe in the data but do not directly account for when we estimate equations (1) or (2).

[TABLE 2 ABOUT HERE]

Table 2 reports the results from regressions using observed student characteristics to predict the probability that a student is enrolled in a co-taught class in a given year. For SWDs, without conditioning on other characteristics but comparing within schools, attending a co-taught class last year is associated with a 13.1 percentage point increase in the probability that the student attends a co-taught class this year. This relationship is relatively unchanged when we introduce other covariates, suggesting that it is independent of the observed student characteristics and prior test scores.

Columns (2) and (5) show the results from models that use a within-school comparison to measure the relationship between student characteristics and probability of enrolling in a co-taught classroom. Students in co-taught classrooms are significantly but only slightly more likely to be female, Hispanic, and Black, and they are slightly less likely to be eligible for free or reduced priced lunch. Though interesting, these particular differences are not worrisome for our estimates because each is either time-invariant or directly controlled in the regression. More potentially troublesome for our identification strategy, however, is that these models also find a significant relationship between prior ELA score and enrolling in a co-taught class. However, the magnitude of the effect is quite small, suggesting that a 1σ increase in test scores for an SWD reduces the likelihood of enrolling in a co-taught classroom by about one percentage point.

The regressions in Columns (3) and (6) best matches the primary estimates because it applies both student and school fixed effects. In this approach we no longer observe a significant relationship between the prior year's ELA score and the likelihood that an SWD is enrolled in a

co-taught class. However, this relationship remains significant but very small for students without disabilities.

We interpret the lack of a strong association between the previous year's test score and assignment to a co-taught classroom conditional on other covariates as evidence that there is likely little or no selection bias due to non-random timing in the assignment to a co-taught class in our models.

4.1.2 Bounding the Effect of Co-Teaching

Our second strategy for addressing the potential that dynamic sorting into co-taught classrooms biases our primary estimates is to calculate a bound for the causal effect of co-teaching by comparing the results from equation (2) with estimates from a regression that makes an opposing assumption about the nature of selection bias. This approach follows the suggestion of Angrist and Pischke (2008).

We estimate models that replace the student fixed-effect with a vector of observed time-invariant student characteristics (Z_i) and the student's test score in the respective subject at the end of the prior year. Directly controlling for the student's test score in the previous year leads the model to estimate the effect of attending a co-taught classroom on student test score *gains*. Formally:

$$y_{ist} = \alpha + \sum_{j=1}^{J} \psi_j X_{ist} + \phi_s + \chi y_{is(t-1)} + \chi Z_i + \phi elem_{ist} + \beta\, coteach_{ist} + \lambda\,(coteach_{ist} \times elem_{ist}) + \epsilon_{ist} \tag{4}$$

In contrast to the assumption underlying our primary approach, causal interpretation of β and λ resulting from equation (4) relies on the assumption that assignment to a co-taught classroom is essentially random conditional on the student's test score in the prior year. That is, the effect is identified under the condition that:

$$E[y_{ist}|X_{it}, Z_i, y_{is(t-1)}, \phi_t, \gamma_s, coteach_{ist}] = E[y_{ist}|X_{it}, Z_i, y_{is(t-1)}, \phi_t, \gamma_s] \tag{5}$$

Since they rely on essentially opposing assumptions about the nature of selection bias, the estimates from equation (2) and equation (4) produce a bound for the causal effect of enrolling in a co-taught classroom on student test scores. The intuition here is that if students are sorted into co-taught classrooms based on inherent time-invariant characteristics then β and λ are accurately identified in equation (2); if students are sorted into co-taught classrooms based on the prior year's performance then β and λ are identified in equation (4); and if there is selection on both fixed and time-variant attributes, then the true effect of co-teaching will fall between the estimates from the two equations.

4.2 Investigating Heterogeneous Effects and Alternative Explanations for Estimated Effect of Co-Teaching

We estimate additional versions of equations (1) and (2) in order to demonstrate the robustness of our estimates to some potential alternative explanations for the result. First, we explore heterogeneity by student gender and race/ethnicity. For SWDs, we also evaluate whether the effect differs according to a student's particular disability classification.

Schools are often motivated to adopt co-teaching by the desire to include SWDs in general education classrooms along with students without disabilities. From a policy perspective, it is not obvious whether it is more relevant to isolate the effect of co-teaching from the effect of attending a more inclusive classroom setting. Thus, in addition to our primary estimates we report results from models that include controls for the characteristics of other students in the classroom, including the percentage of students who receive special education services. As we will see, controlling for classroom characteristics has almost no impact on our estimates for the effect of co-teaching. Thus, we treat the more parsimonious model that excludes these variables as our primary analysis.

In order to address the potential concern that attending a co-taught classroom might coincide with a student being tracked into a more or less advanced course (Duflo et al., 2011), we estimate models that include indicators for particular course designation (e.g. Algebra I, Remedial English, etc.). In addition, we also estimate models that distinguish the effect of co-teaching from

classrooms headed by a single teacher but include other instructional staff including a control for whether the classroom includes a paraprofessional or other classroom aide.

Finally, we estimate models that alter the counterfactual used to compare against the performance of SWDs within co-taught classrooms. The estimate for β derived from equation (1) represents the average effect of attending a co-taught class for SWDs relative to the single-teacher classroom that they would have attended otherwise. For SWDs, this counterfactual classroom might be either a general education classroom that they share with students without disabilities or a more isolated SE-specific classroom. However, for SWDs the effect of attending a co-taught class might differ by whether the alternative is attending a general or a SE-specific classroom. To address this issue with interpretation, we estimate regressions for SWDs within a sample restricted to include only students observed at least once within each class type (co-taught, SE-specific, and general education classroom with a single teacher), and thus each student in the sample contributes to identifying the effect of each class type relative to the others. These models also add a variable indicating whether the student is enrolled in a SE-specific classroom, and its interaction with the indicator for elementary grades.

Unfortunately, there is not a direct way to isolate SE-specific classrooms in the data. Thus, we apply several potential definitions to identify SE-specific classrooms, guided by previous literature on SWDs (Gilmour and Wehby, 2020). We report analyses that define a SE-specific classroom as a class with only a single teacher (that is, not co-taught) and A) 80% or more students have an IEP or the teacher's job code is specific to teaching students with disabilities B) 80% or more students have an IEP or 25% of or more students with IEP and teacher's job code is specific to teaching students with disabilities C) 40% or more students have an IEP.

5 Results

Table 3 reports estimates for the average impact of attending a co-taught classroom on student math and ELA scores by grade level for both SWDs and students without disabilities. Columns 3 and 6 add controls for the demographics of students in the classroom, including the percentage

of students who have a disability. For analyses that interact co-teaching by grade level, the row at the bottom of each analysis reports the p-value testing the null hypothesis that the sum of the coefficients on $Co-Teaching$ and $Co-Teaching \times Elementary$ is equal to zero.

[TABLE 3 ABOUT HERE]

We find that on average attending a co-taught classroom improves test scores for SWDs by about 0.016σ in ELA and about 0.026σ in math. The effect of co-teaching on ELA scores does not differ in elementary and middle school grades. However, we do observe significant differences in the effect of co-teaching by grade level on the math test – we estimate that attending a co-taught class improves SWDs' test scores by about 0.036σ in middle school grades and about 0.014σ in elementary grades.

For students without disabilities, we find that on average attending a co-taught classroom leads to a significant increase in math scores of about 0.012σ, and the impact does not significantly differ across grade types. In ELA, we find no significant effect of co-teaching when all grades are considered together. However, we find some evidence that attending a co-taught classroom could reduce ELA scores for students without disabilities by 0.006σ in middle grades and have a positive effect of about 0.004σ in elementary grades, though it's notable that each of these results is only marginally significant.

Columns (3) and (6) report the results from models that add controls for the characteristics of students in a class. Adding these controls has very little impact on the estimated effect of co-teaching on student test scores. This pattern suggests that the effect of co-teaching is independent of the effect of the change in the characteristics of students in the general classroom setting, including the percentage of SWDs in the class.

5.1 Investigating Heterogeneity in the Effect of Co-Teaching

Tables 4 and 5 investigate whether the effect of co-teaching differs by student gender or race/ethnicity for SWDs and students without disabilities, respectively. The estimates for the effect of co-teaching on students without disabilities are all very similar across categorizations. For SWDs, the results

hint at the potential for some heterogeneity in the effect of co-teaching across student characteristics, but in no case is the difference between respective coefficient estimate statistically significant. In particular, the estimated effect in middle school grades appears potentially larger in ELA for females (0.028σ) than for males (0.006σ) and in elementary grades for Black students (0.051σ) than for white students (0.007σ), but neither of these differences is statistically significant.

[TABLE 4 ABOUT HERE]

[TABLE 5 ABOUT HERE]

Figure 3 illustrates the results from models that separately measure the effect of attending a co-taught classroom for students with different disability classifications. We only separate the effect of co-teaching by grade level when evaluating students classified as having a Specific Learning Disability (SLD) because there are too few observations in the other categories to produce meaningful estimates. We estimate a positive effect from co-teaching within each disability category. The coefficient estimates are fairly similar and do not differ significantly across disability classifications.

[FIGURE 3 ABOUT HERE]

5.2 Investigating Alternative Explanations for the Effect of Co-Teaching

Table 6 reports the results from models that investigate some alternative explanations for the effect of co-teaching on student test scores. Columns (1) and (3) report estimates from models that add to the primary analysis a fixed-effect for the student's course (e.g. Algebra I). Columns (2) and (4) report the results from models that add to the main analysis an indicator for whether the class included another instructional staff member (e.g. a paraprofessional). The results from these analyses do not differ meaningfully from those reported in Table 3. The lone potential difference is that adding a control for the student's particular course somewhat reduces the estimated effect of co-teaching for SWDs. However, in no case does the estimate reported on Table 6 differ significantly from the respective estimate reported on Table 3. Further, notice that adding these controls has little to no effect on the r-squared compared to the primary results.

[TABLE 6 ABOUT HERE]

Table 7 reports the results from regressions that clarify the counterfactual classroom for SWDs by restricting the sample to include only SWDs who are observed at least once in each of the three potential classroom settings: co-taught, SE-specific, or general education with a single teacher. The classroom setting types are mutually exclusive. These models add an indicator for if the student attends a SE-specific classroom (and its interaction with the variable indicating elementary grades), and thus the omitted comparison is the student's performance within a general classroom with a single teacher. Recall that because there is no clear way to identify SE-specific classrooms in the data we estimate models in which we identify such classes in one of three ways described previously.

In no case do we find a significant difference between the effect of attending a SE-specific class and attending a general education class with a single teacher. Thus, it is not surprising that despite the change in the counterfactual comparison group in no case does the estimated effect of attending a co-taught class reported on Table 7 differ significantly from the respective estimate from the main specification reported in Table 3.[9] Importantly, despite the much more restricted sample, these models are estimated precisely enough to detect even small differences. We interpret this pattern of results as evidence that the effect of a SWD attending a co-taught classroom is similar regardless of whether the student would otherwise attend a SE-specific class or a general education class headed by a single teacher.

[TABLE 7 ABOUT HERE]

5.3 Binding the Causal Effect of Co-Teaching

Figure C4 in the Online Appendix illustrates the results from our specification test comparing estimates from student fixed effects models (equation 2) with those from models that instead control

[9]Table B7 in the Online Appendix relaxes the restriction by requiring only that a student is observed in a co-taught classroom at least once. The results are very similar.

for observed student demographics and a lagged dependent variable (equation 4).[10] Recall that the purpose of this test is to create a bound for the causal effect of co-teaching by comparing the results from regressions that make opposing assumptions about the nature of selection bias due to non-random sorting of students into co-taught classrooms.

Our results suggest that the feasible range for the effect of co-teaching is quite small. The 95% confidence intervals overlap in six of the eight comparisons shown on Figure C4, and in only three cases is the coefficient estimate for at least one of the models not contained within the 95% confidence interval of the other.

However, though the estimates are often not distinguishable across the two models, in some cases the comparison between the results from the two models does alter the interpretation for the effect of co-teaching. That is largely because the estimated effect within the primary analysis is often small enough that even the marginal decline in the estimate leads the coefficient in the lagged dependent variable model to drop below zero. For example, the finding from the primary analysis that SWDs in middle school benefited from enrolling in a co-taught class on the ELA exam is called into question because the lagged dependent variable analysis yields a negative but statistically insignificant coefficient.

Notably, both the student fixed-effect and lagged dependent variable approaches produce positive and statistically significant estimates for SWDs in elementary grades in ELA and in both grade levels on the math tests. Thus, the result that attending a co-taught classroom benefits SWDs remains intact following this specification test. For students without disabilities, both models find a positive effect in elementary math, a negative effect in middle school ELA, and a positive but insignificant effect in elementary ELA.

6 Summary and Conclusion

We draw upon longitudinal administrative data from Massachusetts to produce the first causal estimate for the effect of co-teaching on student academic performance across a large public school

[10]The relevant coefficients and standard errors are reported on Appendix Table B10

system. Overall, our results suggest that attending a co-taught classroom leads to statistically significant test score benefits for both SWDs and students without disabilities. The results are fairly robust to estimates that employ an opposing assumption about the nature of selection into a co-taught classroom.

Though distinctly positive, the magnitude of the effect of co-teaching on SWDs that we identify is substantially smaller than reported in prior studies that have guided policymakers thus far. For example, among the estimates included in a meta analysis for the effect of co-teaching by Murawski and Swanson (2001) the average effect size was 0.40σ and the largest estimate was 0.95σ. However, prior estimates for the effect of co-teaching are likely misleading because they applied strategies that are not capable of removing selection bias and/or were limited to very small samples within specific contexts (Cook et al., 2011). The magnitude of our credible estimate for the causal effect of co-teaching as applied across Massachusetts falls between the 30th and 40th percentile of effect sizes from randomized control trials of educational interventions on student standardized test scores, and characterized as "small" by the taxonomy recently developed by Kraft (2020). Thus, though our results are overall positive within a broad context, we argue that they do not appear consistent with the enthusiasm that surrounds co-teaching in the special education literature.

From a policy perspective, whether our findings nonetheless support a decision to adopt co-teaching as an instructional model depends on the associated costs, which are thus far unknown. Transitioning to a co-teaching model almost certainly requires additional resources and personnel. However, the ability for schools to reassign employees and change the distribution of students, and the fact that (at least in Massachusetts) schools do not adopt co-teaching wholesale for all students makes it unlikely that adding a co-teacher simply doubles the staffing expenditure associated with students in a classroom. We look forward to future work quantifying the financial implications of co-teaching within schools.

That we analyze co-teaching within the context of a large public school system is arguably both our study's most valuable contribution and its most substantial limitation. Proponents of co-teaching might reasonably argue that the magnitude of our estimates is muted by schools and

teachers that have moved students into classrooms where teachers are purportedly co-teaching but in reality do not apply the best practices necessary for co-teaching to be effective. Indeed, differences in the fidelity of implementation when co-teaching is considered across an entire state rather than within the context of a single school or classroom could at least partially explain why our findings are so much smaller than prior experimentally derived estimates. It is also primarily for this reason that, though co-teaching certainly reduces the adult-to-student ratio within a classroom, we caution against interpreting our results for the effect of co-teaching as the effect of substantially reducing class size. Though our data does not allow for such a black-box analysis, our findings appear consistent with observational studies suggesting that in practice co-teachers often do not work collaboratively to instruct all students in the idealized version recommended by its advocates (Cook et al., 2011; Scruggs et al., 2007; Wexler et al., 2018). From a policy perspective, however, our finding that co-teaching as it is currently implemented across a state has at best modest positive academic effects is highly relevant. Whether the impact of co-teaching at such a scale can be improved by encouraging more consistent application of best practices is an important area for future research.

Further, when interpreting our findings it is important to keep in mind that our analysis is limited to considering the immediate effects of co-teaching on student test scores. However, some of the most important justifications for moving SWDs into inclusive classroom settings, which co-teaching is often meant to facilitate, are not academic but social. If co-teaching were to improve the social skills and friendships of SWDs then it might be worthwhile even if it does not lead to substantial test score improvements. However, it is not clear why co-teaching itself would have positive effects on the socialization of SWDs independent of creating inclusive classroom settings.

We close by noting that co-teaching is not the only example of a policy or practice targeted to improving the outcomes of SWDs that has gained widespread public acceptance within the special education literature and among policymakers despite little rigorous research on its impact at scale. Our paper is one of a few recent studies that use large administrative data to measure causal effects within the context of special education (Schwartz et al., 2019; Ballis and Heath, 2021b; Setren, 2019). We complement this emerging literature. And we encourage the field to

continue building a more robust evidence base surrounding the services schools provide to this especially vulnerable group of students.

References

Andersen, S. C., Beuchert, L., Nielsen, H. S., and Thomsen, M. K. (2020). The effect of teacher's aides in the classroom: Evidence from a randomized trial. *Journal of the European Economic Association*, 18(1):469–505.

Angrist, J. D. and Lavy, V. (1999). Using Maimonides' rule to estimate the effect of class size on scholastic achievement. *The Quarterly Journal of Economics*, 114(2):533–575.

Angrist, J. D. and Pischke, J.-S. (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press.

Austin, V. L. (2001). Teachers' beliefs about co-teaching. *Remedial and Special Education*, 22(4):245–255.

Ballis, B. and Heath, K. (2021a). Direct and spillover effects of limiting minority student access to special education. *EdWorkingPapers*.

Ballis, B. and Heath, K. (2021b). The long-run impacts of special education. *American Economic Journal: Economic Policy*, 13(4):72–111.

Bessette, H. J. (2008). Using students' drawings to elicit general and special educators' perceptions of co-teaching. *Teaching and Teacher Education*, 24(5):1376–1396.

Blatchford, P., Bassett, P., and Brown, P. (2011). Examining the effect of class size on classroom engagement and teacher–pupil interaction: Differences in relation to pupil prior attainment and primary vs. secondary schools. *Learning and Instruction*, 21(6):715–730.

Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W., and Yagan, D. (2011). How does your Kindergarten classroom affect your earnings? Evidence from Project STAR. *The Quarterly Journal of Economics*, 126(4):1593–1660.

Conderman, G. and Hedin, L. R. (2014). Co-teaching with strategy instruction. *Intervention in School and Clinic*, 49(3):156–163.

Cook, B. G., McDuffie-Landrum, K. A., Oshita, L., and Cook, S. C. (2011). Co-teaching for students with disabilities. In Kauffman, J. M., Hallahan, D. P., and Pullen, P. C., editors, *Handbook of Special Education*, pages 187–208. New York, NY: Routledge.

Crosland, K. and Dunlap, G. (2012). Effective strategies for the inclusion of children with autism in general education classrooms. *Behavior Modification*, 36(3):251–269.

Dee, T. S. and West, M. R. (2011). The non-cognitive returns to class size. *Educational Evaluation and Policy Analysis*, 33(1):23–46.

Dieker, L. A. (2001). What are the characteristics of "effective" middle and high school co-taught teams for students with disabilities? *Preventing School Failure: Alternative Education for Children and Youth*, 46(1):14–23.

Duflo, E., Dupas, P., and Kremer, M. (2011). Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya. *American Economic Review*, 101(5):1739–74.

Falch, T., Sandsør, A. M. J., and Strøm, B. (2017). Do smaller classes always improve students' long-run outcomes? *Oxford Bulletin of Economics and Statistics*, 79(5):654–688.

Finn, J. D., Pannozzo, G. M., and Achilles, C. M. (2003). The "why's" of class size: Student behavior in small classes. *Review of Educational Research*, 73(3):321–368.

Fletcher, J. (2010). Spillover effects of inclusion of classmates with emotional problems on test scores in early elementary school. *Journal of Policy Analysis and Management*, 29(1):69–83.

Fontana, K. C. (2005). The effects of co-teaching on the achievement of eighth grade students with learning disabilities. *Journal of At-Risk Issues*, 11(2):17–22.

Fredriksson, P., Öckert, B., and Oosterbeek, H. (2013). Long-term effects of class size. *The Quarterly Journal of Economics*, 128(1):249–285.

Friend, M. (2008). Co-teaching: A simple solution that isn't simple after all. *Journal of Curriculum and Instruction*, 2(2):9–19.

Friend, M. (2015). Welcome to co-teaching 2.0. *Educational Leadership*, 73(4):16–22.

Friend, M., Cook, L., Hurley-Chamberlain, D., and Shamberger, C. (2010). Co-teaching: An illustration of the complexity of collaboration in special education. *Journal of Educational and Psychological Consultation*, 20(1):9–27.

Gately, S. E. and Gately, Jr., F. J. (2001). Understanding coteaching components. *TEACHING Exceptional Children*, 33(4):40–47.

Gilmour, A. F., Fuchs, D., and Wehby, J. H. (2019). Are students with disabilities accessing the curriculum? A meta-analysis of the reading achievement gap between students with and without disabilities. *Exceptional Children*, 85(3):329–346.

Gilmour, A. F. and Henry, G. T. (2018). Who are the classmates of students with disabilities in elementary mathematics classrooms? *Remedial and Special Education*, 41(1):18–27.

Gilmour, A. F. and Wehby, J. H. (2020). The association between teaching students with disabilities and teacher turnover. *Journal of Educational Psychology*, 112(5):1042.

Gottfried, M. A. (2014). Classmates with disabilities and students' noncognitive outcomes. *Educational Evaluation and Policy Analysis*, 36(1):20–43.

Gottfried, M. A., Egalite, A., and Kirksey, J. J. (2016). Does the presence of a classmate with emotional/behavioral disabilities link to other students' absences in kindergarten? *Early Childhood Research Quarterly*, 36:506–520.

Gottfried, M. A. and Harven, A. (2015). The effect of having classmates with emotional and behavioral disorders and the protective nature of peer gender. *The Journal of Educational Research*, 108(1):45–61.

Hang, Q. and Rabren, K. (2009). An examination of co-teaching: Perspectives and efficacy indicators. *Remedial and Special Education*, 30(5):259–268.

Hanushek, E. A., Kain, J. F., and Rivkin, S. G. (2002). Inferring program effects for special populations: Does special education raise achievement for students with disabilities? *Review of Economics and Statistics*, 84(4):584–599.

Harrison, J. R., Vannest, K., Davis, J., and Reynolds, C. (2012). Common problem behaviors of children and adolescents in general education classrooms in the United States. *Journal of Emotional and Behavioral Disorders*, 20(1):55–64.

Hemelt, S. W., Ladd, H. F., and Clifton, C. R. (2021). Do teacher assistants improve student outcomes? evidence from school funding cutbacks in North Carolina. *Educational Evaluation and Policy Analysis*, 43(2):280–304.

Hoxby, C. M. (2000). The effects of class size on student achievement: New evidence from population variation. *The Quarterly Journal of Economics*, 115(4):1239–1285.

Jones, N. D., Bettini, E., and Brownell, M. (2019). Competing strands of educational reform policy: Can collaborative school reform and teacher evaluation reform be reconciled? *The Elementary School Journal*, 119(3):468–486.

Keefe, E. B. and Moore, V. (2004). The challenge of co-teaching in inclusive classrooms at the high school level: What the teachers told us. *American Secondary Education*, 32(3):77–88.

Klinger, J., Vaughn, S., Hughes, M. T., Schumm, J. S., and Elbaum, B. (1998). Outcomes for students with and without learning disabilities. *Learning Disabilities Research & Practice*, 13(3):153–161.

Kohler-Evans, P. A. (2006). Co-teaching: How to make this marriage work in front of the kids. *Education*, 127(2):260–265.

Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher*, 49(4):241–253.

Kraft, M. A., Blazar, D., and Hogan, D. (2018). The effect of teacher coaching on instruction and achievement: A meta-analysis of the causal evidence. *Review of Educational Research*, 88(4):547–588.

Krueger, A. B. (2003). Economic considerations and class size. *The Economic Journal*, 113(485):F34–F63.

Krueger, A. B. and Whitmore, D. M. (2001). The effect of attending a small class in the early grades on college-test taking and middle school test results: Evidence from Project STAR. *The Economic Journal*, 111(468):1–28.

Lane, K. L., Carter, E. W., Pierson, M. R., and Glaeser, B. C. (2006). Academic, social, and behavioral characteristics of high school students with emotional disturbances or learning disabilities. *Journal of Emotional and Behavioral Disorders*, 14(2):108–117.

Leuven, E. and Løkken, S. A. (2020). Long-term impacts of class size in compulsory school. *Journal of Human Resources*, 55(1):309–348.

Leuven, E., Oosterbeek, H., and Rønning, M. (2008). Quasi-experimental estimates of the effect of class size on achievement in Norway. *Scandinavian Journal of Economics*, 110(4):663–693.

Mastropieri, M. A., Scruggs, T. E., Graetz, J., Norland, J., Gardizi, W., and Mcduffie, K. (2005). Case studies in co-teaching in the content areas: Successes, failures, and challenges. *Intervention in School and Clinic*, 40(5):260–270.

McDuffie, K. A., Mastropieri, M. A., and Scruggs, T. E. (2009). Differential effects of peer tutoring in co-taught and non-co-taught classes: Results for content learning and student-teacher interactions. *Exceptional Children*, 75(4):493–510.

Molnar, A., Smith, P., Zahorik, J., Palmer, A., Halbach, A., and Ehrle, K. (1999). Evaluating the sage program: A pilot program in targeted pupil-teacher reduction in wisconsin. *Educational Evaluation and Policy Analysis*, 21(2):165–177.

Murawski, W. W. (2005). Addressing diverse needs through co-teaching. *Kappa Delta Pi Record*, 41(2):77–82.

Murawski, W. W. (2006). Student outcomes in co-taught secondary english classes: How can we improve? *Reading & Writing Quarterly*, 22(3):227–247.

Murawski, W. W. and Bernhardt, P. (2015). An administrator's guide to co-teaching. *Educational Leadership*, 73(4):30–34.

Murawski, W. W. and Dieker, L. A. (2004). Tips and strategies for co-teaching at the secondary level. *TEACHING Exceptional Children*, 36(5):52–58.

Murawski, W. W. and Lochner, W. W. (2011). Observing co-teaching: What to ask for, look for, and listen for. *Intervention in School and Clinic*, 46(3):174–183.

Murawski, W. W. and Swanson, H. L. (2001). A meta-analysis of co-teaching research: Where are the data? *Remedial and Special Education*, 22(5):258–267.

Nierengarten, G. (2013). Supporting co-teaching teams in high schools: Twenty research-based practices. *American Secondary Education*, 42(1):73–83.

Ploessl, D. M., Rock, M. L., Schoenfeld, N., and Blanks, B. (2010). On the same page: Practical techniques to enhance co-teaching interactions. *Intervention in School and Clinic*, 45(3):158–168.

Pratt, S. (2014). Achieving symbiosis: Working through challenges found in co-teaching to achieve effective co-teaching relationships. *Teaching and Teacher Education*, 41:1–12.

Rice, D. and Zigmond, N. (2000). Co-teaching in secondary schools: Teacher reports of developments in Australian and American classrooms. *Learning Disabilities Research & Practice*, 15(4):190–197.

Schulte, A. C. and Stevens, J. J. (2015). Once, sometimes, or always in special education: Mathematics growth and achievement gaps. *Exceptional Children*, 81(3):370–387.

Schwartz, A. E., Hopkins, B. G., and Stiefel, L. (2019). The effects of special education on the academic performance of students with learning disabilities. *EdWorkingPapers*, 19-86. http://www.edworkingpaper.com/ai19-86.

Scruggs, T. E., Mastropieri, M. A., and McDuffie, K. A. (2007). Co-teaching in inclusive classrooms: A metasynthesis of qualitative research. *Exceptional Children*, 73(4):392–416.

Setren, E. (2019). Targeted vs. general education investments: Evidence from special education and English language learners in Boston charter schools. *Journal of Human Resources*, 0219–10040R2.

Sileo, J. M. (2011). Co-teaching: Getting to know your partner. *Teaching Exceptional Children*, 43(5):32–38.

Trent, S. C. (1998). False starts and other dilemmas of a secondary general education collaborative teacher: A case study. *Journal of Learning Disabilities*, 31(5):503–513.

Vaughn, S., Schumm, J. S., and Arguelles, M. E. (1997). The ABCDEs of co-teaching. *Teaching Exceptional Children*, 30(2):4–10.

Walther-Thomas, C., Bryant, M., and Land, S. (1996). Planning for effective co-teaching: The key to successful inclusion. *Remedial and Special Education*, 17(4):255–264.

Walther-Thomas, C. S. (1997). Co-teaching experiences: The benefits and problems that teachers and principals report over time. *Journal of Learning Disabilities*, 30(4):395–407.

Welch, M. (2000). Descriptive analysis of team teaching in two elementary classrooms: A formative experimental approach. *Remedial and Special Education*, 21(6):366–376.

Wexler, J., Kearns, D. M., Lemons, C. J., Mitchell, M., Clancy, E., Davidson, K. A., Sinclair, A. C., and Wei, Y. (2018). Reading comprehension and co-teaching practices in middle school English language arts classrooms. *Exceptional Children*, 84(4):384–402.

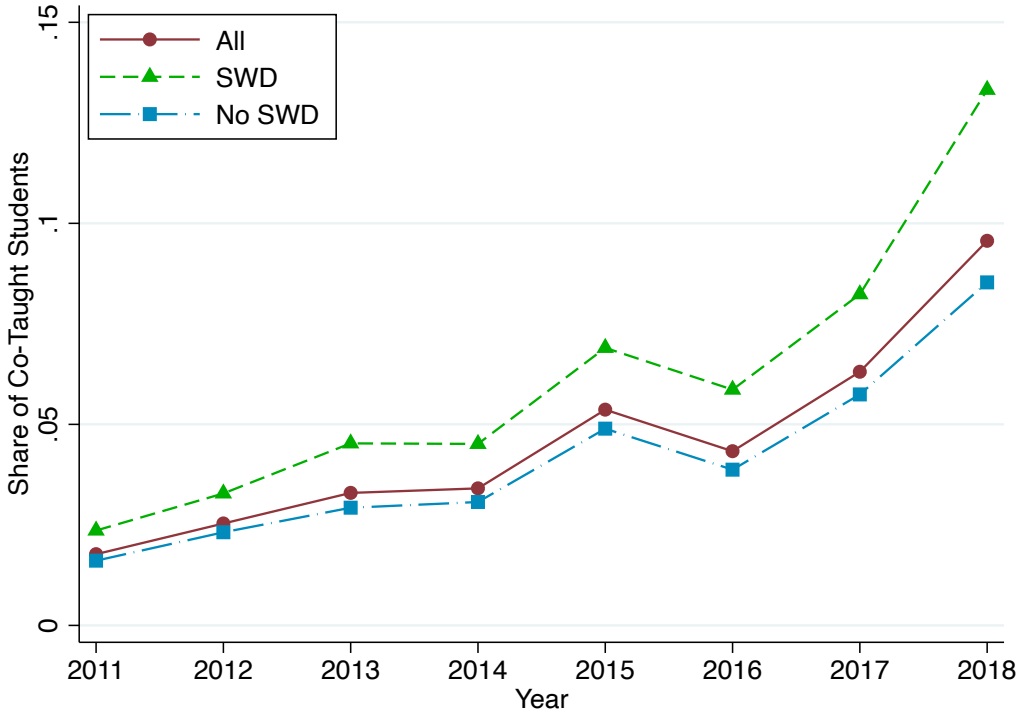
Wilson, G. L. (2005). This doesn't look familiar! A supervisor's guide for observing co-teachers. *Intervention in School and Clinic*, 40(5):271–275.

Wilson, G. L. (2008). Be an active co-teacher. *Intervention in School and Clinic*, 43(4):240–243.

Woessmann, L. and West, M. (2006). Class-size effects in school systems around the world: Evidence from between-grade variation in TIMSS. *European Economic Review*, 50(3):695–736.

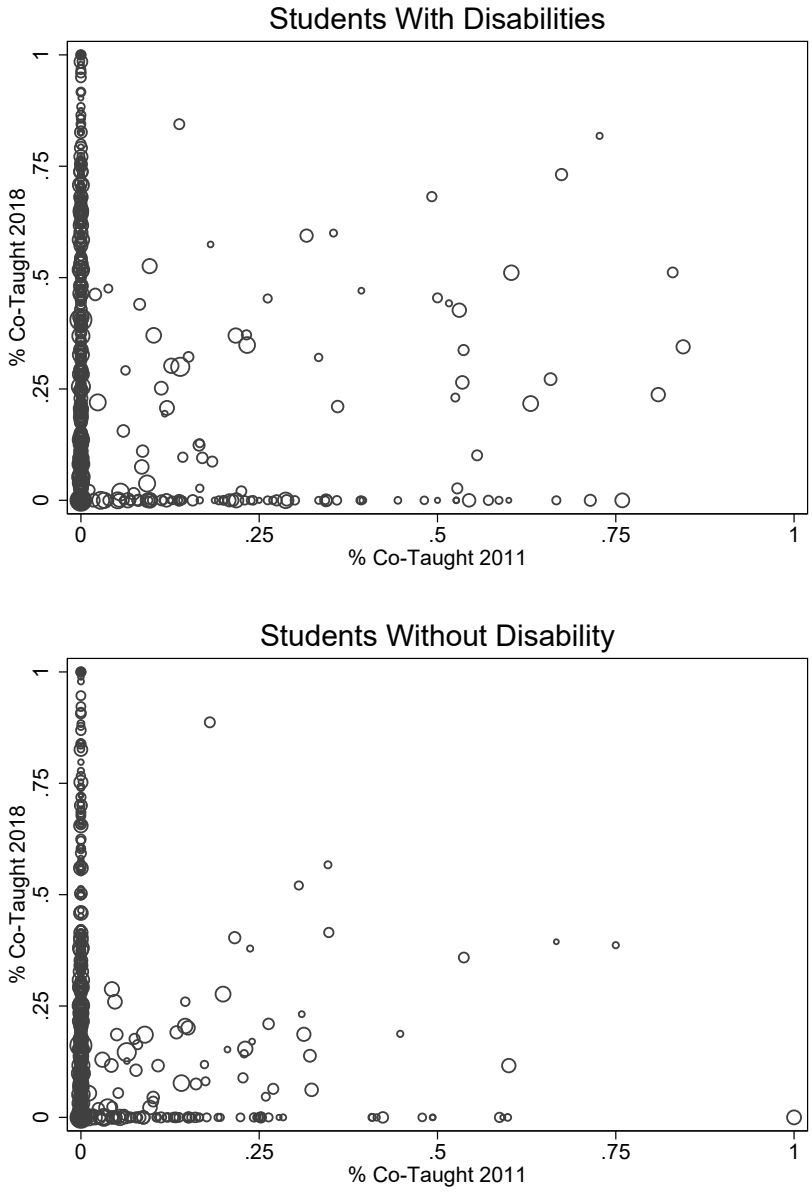
Figures

Figure 1: Change in the Share of Co-Taught Students: ELA Sample



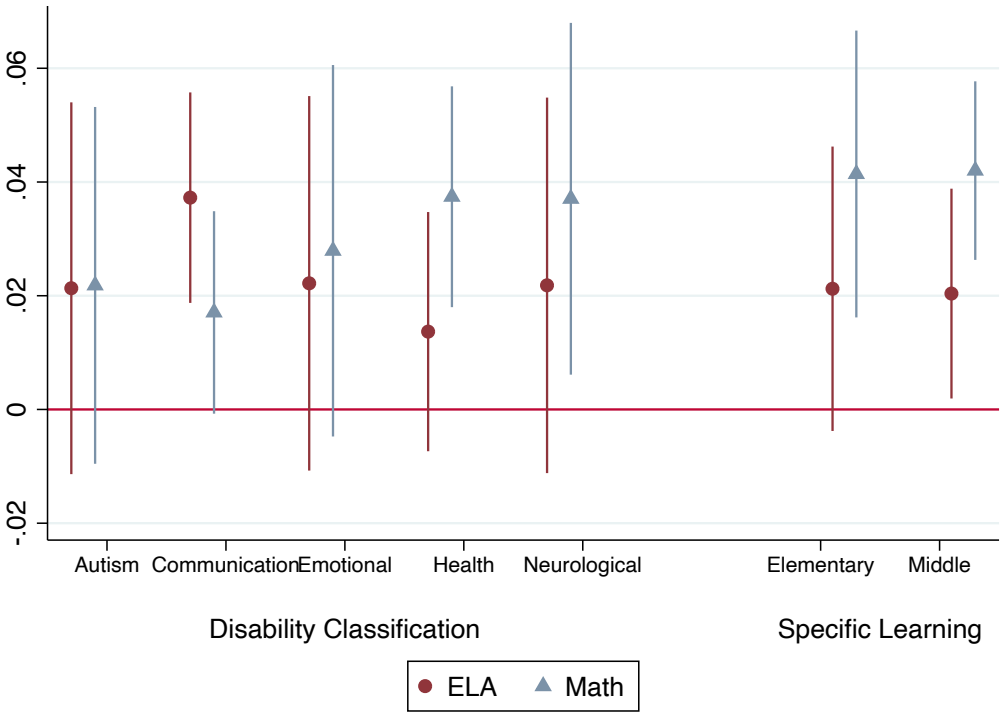
Note: This figure plots the share of co-taught students each year in the ELA sample. Dots are the total share of co-taught students. Triangles are the share of co-taught SWDs. Squares are the share of co-taught students without disabilities.

Figure 2: Percentage of Students in Co-Taught Classrooms by School, 2011 and 2018



Note: This figure plots the percentage of students with disabilities (top panel) and percentage of students without disabilities (bottom panel) observed in a co-taught classroom by school within the ELA sample during the 2011 and 2018 school years. Size of circle illustrates total enrollment within the school in grades 3 through 8.

Figure 3: Effect of Co-Teaching by Special Education Classification



Note: This figure documents the effect of co-teaching for SWDs by special education classifications. Circles are the effect of co-teaching on standardized ELA scores. Triangles are the effect of co-teaching on standardized Math scores. Vertical bars represent 95% confidence intervals. The effect of co-teaching for the students with specific learning disabilities in elementary and middle grades are separately estimated and reported on the right end. For the numerical values underlying this figure, see Table B6.

Tables

Table 1: Descriptive Statistics

	Non-SWD			SWD		
	(1)	(2)	(3)	(4)	(5)	(6)
	Sometimes	Never	Always	Sometimes	Never	Always
Female	0.521	0.529*	0.526*	0.368†	0.366†	0.336†*
White	0.770	0.802	0.705*	0.793†	0.822†*	0.748
Black	0.175	0.125*	0.311	0.197†	0.161†*	0.264†
Hispanic	0.191	0.157*	0.337	0.218†	0.208†*	0.255†
Asian	0.081	0.096*	0.092*	0.037†	0.043†	0.033†
Free Lunch	0.384	0.305*	0.413*	0.480†	0.426†*	0.483†*
Reduced Lunch	0.044	0.039	0.041	0.048†	0.043	0.051†
Autism				0.054	0.061*	0.097*
Communication				0.153	0.145	0.178
Emotional				0.056	0.060*	0.051
Health				0.122	0.113*	0.127*
Neurological				0.051	0.047*	0.051
Specific Learning				0.305	0.275*	0.265
ELA Score	0.091	0.257*	0.006*	-0.874†	-0.730†*	-0.947†*
Math Score	0.073	0.255*	-0.017*	-0.844†	-0.700†*	-0.902†*
Classroom Size	23.891	23.093*	23.478	21.568†	20.333†*	22.287†
% Obs Co-Taught	0.246			0.266†		
Observations	307578	963197	9562	127462	246783	3941

*$p < .05$ Relative to "Never"

†$p < .05$ Relative to Respective Non-SWD

Note: Table reports mean characteristics by special education status and whether student is sometimes, never, or always observed in a co-taught classroom. The samples are restricted to include only observations within schools that are observed to have co-teaching within the sample period. Inferential tests come from regressions that control for school fixed effect but no other controls. The table reports significant differences from two comparisons evaluated separately. *indicates significant difference at 5% level relative to students who were for that cell relative to those who were "Sometimes" observed in a co-taught classroom (i.e. Column 1 vs Column 2; Column 6 vs Column 5, and so on). †indicates significant difference at 5% level relative to students with same co-teaching status (i.e. "Sometimes", "Never", "Always") among students without a disability (i.e. Column 4 vs Column 1; Column 5 vs Column 2, and so on).

Table 2: Regression on Observables

	(1)	(2)	(3)	(4)	(5)	(6)
Lag Co-Teaching	0.131***	0.124***		0.044***	0.044***	
	(0.003)	(0.003)		(0.002)	(0.002)	
Lag ELA		-0.010***	-0.002		-0.008***	-0.003***
		(0.001)	(0.001)		(0.000)	(0.001)
Female		0.003**			0.001	
		(0.001)			(0.001)	
Hispanic		0.006***			0.005***	
		(0.002)			(0.001)	
Black		0.004**			0.003***	
		(0.002)			(0.001)	
Asian		0.004			0.002**	
		(0.003)			(0.001)	
Free Lunch		-0.005***	-0.009**		0.001	-0.005***
		(0.002)	(0.004)		(0.001)	(0.002)
Reduced Lunch		-0.007**	0.002		-0.004**	-0.007**
		(0.003)	(0.006)		(0.001)	(0.003)
Observations	265160	250432	217914	893721	882075	767224
Sample	SWD	SWD	SWD	Non-SWD	Non-SWD	Non-SWD
School FE	Yes	Yes	Yes	Yes	Yes	Yes
Student FE	No	No	Yes	No	No	Yes

Standard errors clustered by student in parentheses

* $p < .1$, ** $p < .05$, *** $p < .01$

Note: This table presents the estimates of regressing co-teaching on observables in the ELA sample, restricted to SWDs. Sample includes students in grades three through eight within schools in which we observe any co-taught classrooms within the sample period. Dependent variable in all models is an indicator that equals 1 if the student is in a co-taught class. "Lagged Co-Teaching" takes a value of 1 if a student was in a co-taught ELA classroom in the previous year. "Lagged ELA Score" is the student's standardized ELA test score in the previous year. Standard errors are clustered at the student level and reported in the parenthesis. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 3: Effect of Co-Teaching on Standardized Test Scores by Grade Levels

	(1)	(2)	(3)	(4)	(5)	(6)
			English Language Arts			
Co-Teaching	0.016***	0.014***	0.014***	-0.001	-0.006*	-0.004
	(0.004)	(0.005)	(0.005)	(0.002)	(0.003)	(0.003)
Co-Teaching ×		0.005	0.005		0.010**	0.009**
Elementary		(0.008)	(0.008)		(0.004)	(0.004)
Observations	129653	129653	129653	284597	284597	284597
r2	0.778	0.778	0.778	0.807	0.807	0.807
F Test p-values						
Co-Teach + Interaction = 0		0.020	0.016		0.302	0.093
			Mathematics			
Co-Teaching	0.026***	0.036***	0.036***	0.012***	0.014***	0.012***
	(0.004)	(0.005)	(0.005)	(0.002)	(0.003)	(0.003)
Co-Teaching ×		-0.022***	-0.022***		-0.003	-0.002
Elementary		(0.008)	(0.008)		(0.004)	(0.004)
Observations	129668	129668	129668	285687	285687	285687
r2	0.793	0.793	0.793	0.845	0.845	0.845
F Test p-values						
Co-Teach + Interaction = 0		0.032	0.024		0.000	0.001
Sample	SWD	SWD	SWD	Non-SWD	Non-SWD	Non-SWD
Student FE	Yes	Yes	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes
Class Demographics	No	No	Yes	No	No	Yes
SWD Share	No	No	Yes	No	No	Yes
SWD Share × SWD	No	No	Yes	No	No	Yes

Note: This table presents the effect of co-teaching on standardized ELA and Math scores. "Co-Teaching" takes a value of 1 if a student was in a co-taught classroom in the subject area. The sample is restricted to students enrolled in grades three through eight that we observe in a co-taught class in the subject area at least once. "Elementary" takes a value of 1 if a student is below grade six. All regressions additionally control for an indicators for if the student is in an elementary grade and for if the student is eligible for free lunch or reduced priced lunch. In columns (1) to (3), the sample is restricted to SWDs. In columns (4) to (6), the sample is restricted to students without disabilities. Standard errors are clustered at the student level and reported in the parenthesis. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 4: Effect of Co-Teaching by Demographics: SWD Sample

	Standardized Test Scores				
	(1)	(2)	(3)	(4)	(5)
	English Language Arts				
Co-Teaching	0.028***	0.006	0.019***	-0.008	0.005
	(0.008)	(0.007)	(0.006)	(0.012)	(0.012)
Co-Teaching ×	-0.011	0.011	-0.012	0.059***	0.021
Elementary	(0.013)	(0.010)	(0.009)	(0.019)	(0.017)
Observations	52583	90623	114873	26552	30359
r2	0.786	0.778	0.782	0.750	0.753
F Test p-values					
Co-Teach + Interaction = 0	0.080	0.033	0.317	0.000	0.049
	Mathematics				
Co-Teaching	0.029***	0.038***	0.038***	0.016	0.023**
	(0.007)	(0.006)	(0.005)	(0.012)	(0.011)
Co-Teaching ×	-0.010	-0.028***	-0.026***	0.009	-0.034*
Elementary	(0.012)	(0.010)	(0.008)	(0.019)	(0.018)
Observations	48490	81020	105196	23064	26136
r2	0.783	0.800	0.796	0.755	0.751
F Test p-values					
Co-Teach + Interaction = 0	0.059	0.210	0.085	0.105	0.429
Sample	Female	Male	White	Black	Hispanic
Student FE	Yes	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes	Yes

Note: This table presents the effect of co-teaching on standardized ELA and Math scores for SWDs by demographics. "Co-Teaching" takes a value of 1 if a student was in a co-taught classroom in the subject area. The sample is restricted to students enrolled in grades three through eight that we observe in a co-taught class in the subject area at least once. "Elementary" takes a value of 1 if a student is below grade six. Standard errors are clustered at the student level and reported in the parenthesis. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 5: Effect of Co-Teaching by Demographics: Students without Disabilities Sample

	Standardized Test Scores				
	(1)	(2)	(3)	(4)	(5)
Sample	Female	Male	White	Black	Hispanic
	English Language Arts				
Co-Teaching	0.001	-0.013***	-0.007**	-0.016**	0.009
	(0.004)	(0.004)	(0.003)	(0.008)	(0.008)
Co-Teaching ×	0.004	0.017***	0.011**	-0.001	-0.010
Elementary	(0.006)	(0.006)	(0.005)	(0.011)	(0.010)
Observations	177151	163359	264267	57743	64617
r2	0.807	0.804	0.801	0.795	0.800
F Test p-values					
Co-Teach + Interaction = 0	0.284	0.378	0.184	0.017	0.864
	Mathematics				
Co-Teaching	0.006*	0.021***	0.009***	0.018***	0.006
	(0.004)	(0.004)	(0.003)	(0.007)	(0.007)
Co-Teaching ×	0.002	-0.007	0.002	-0.016	-0.004
Elementary	(0.006)	(0.006)	(0.005)	(0.011)	(0.010)
Observations	150304	135210	223040	48666	53735
r2	0.847	0.846	0.840	0.826	0.829
F Test p-values					
Co-Teach + Interaction = 0	0.053	0.003	0.001	0.799	0.857
Student FE	Yes	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes	Yes

Note: This table presents the effect of co-teaching on standardized ELA and Math scores for students without disabilities by demographics. "Co-Teaching" takes a value of 1 if a student was in a co-taught classroom in the subject area. The sample is restricted to students enrolled in grades three through eight that we observe in a co-taught class in the subject area at least once. "Elementary" takes a value of 1 if a student is below grade six. Standard errors are clustered at the student level and reported in the parenthesis. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 6: Effect of Co-Teaching: Alternative Specifications

	Standardized Test Scores			
	(1)	(2)	(3)	(4)
	English Language Arts			
Co-Teaching	0.006	0.015***	-0.008***	-0.005*
	(0.005)	(0.005)	(0.003)	(0.003)
Co-Teaching ×	0.002	0.003	0.008*	0.009**
Elementary	(0.008)	(0.008)	(0.004)	(0.004)
Observations	143348	143351	340654	340656
r2	0.780	0.778	0.808	0.807
F Test p-values				
Co-Teach + Interaction = 0	0.181	0.003	0.901	0.250
	Mathematics			
Co-Teaching	0.028***	0.040***	0.013***	0.014***
	(0.005)	(0.005)	(0.003)	(0.003)
Co-Teaching ×	-0.020***	-0.026***	-0.003	-0.004
Elementary	(0.008)	(0.008)	(0.004)	(0.004)
Observations	129663	129668	285682	285687
r2	0.794	0.793	0.846	0.845
F Test p-values				
Co-Teach + Interaction = 0	0.188	0.028	0.001	0.002
Sample	SWD	SWD	No SWD	No SWD
Student FE	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes
Course FE	Yes	No	Yes	No
Helper FE	No	Yes	No	Yes

Note: This table presents the effect of co-teaching on standardized ELA and Math scores using alternative specifications. In columns (1) and (3), we include specific course fixed effects. In columns (2) and (4), we include the indicators for having additional instructional staff members in the classroom. "Co-Teaching" takes a value of 1 if a student was in a co-taught classroom in the subject area. The sample is restricted to students enrolled in grades three through eight that we observe in a co-taught class in the subject area at least once. "Elementary" takes a value of 1 if a student is below grade six. Standard errors are clustered at the student level and reported in the parenthesis. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 7: Effect of Co-Teaching with Clarified Counterfactual

	(1)	(2)	(3)	(4)	(5)	(6)
			English Language Arts			
Co-Teaching	0.010	0.008	0.012*	0.007	0.001	0.017**
	(0.008)	(0.008)	(0.006)	(0.010)	(0.010)	(0.008)
Co-Teaching ×				0.010	0.021	-0.014
Elementary				(0.016)	(0.017)	(0.013)
Special Ed Class	-0.000	-0.007	0.000	0.004	-0.004	0.002
	(0.007)	(0.007)	(0.005)	(0.009)	(0.009)	(0.007)
Special Ed Class ×				-0.015	-0.013	-0.004
Elementary				(0.016)	(0.016)	(0.011)
Observations	51009	49333	82916	51009	49333	82916
r2	0.757	0.754	0.757	0.757	0.754	0.757
F Test p-values						
Co-Teach + Interaction = 0				0.212	0.114	0.700
Sped Class + Interaction = 0				0.422	0.204	0.848
			Mathematics			
Co-Teaching	0.019**	0.019**	0.020***	0.036***	0.035***	0.038***
	(0.007)	(0.008)	(0.006)	(0.009)	(0.009)	(0.007)
Co-Teaching ×				-0.044***	-0.042***	-0.047***
Elementary				(0.016)	(0.016)	(0.012)
Special Ed Class	0.002	-0.001	-0.006	0.011	0.009	0.002
	(0.007)	(0.007)	(0.005)	(0.008)	(0.008)	(0.006)
Sped1 ×				-0.022	-0.024	-0.015
Elementary				(0.015)	(0.015)	(0.010)
Observations	45974	44371	76475	45974	44371	76475
r2	0.757	0.752	0.763	0.757	0.752	0.763
F Test p-values						
Co-Teach + Interaction = 0				0.526	0.612	0.354
Sped Class + Interaction = 0				0.418	0.254	0.117
Sped Class Definition	Type 1	Type 2	Type 3	Type 1	Type 2	Type 3
Student FE	Yes	Yes	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes

Note: This table presents the effect of co-teaching on standardized ELA and Math scores within a sample restricted to include only SWDs observed at least once in the subject area within each of a co-taught, special education, and regular classroom with a single teacher. The omitted comparison group is observations within a regular classroom with a single teacher. Each column represents models using different ways of identifying special education classrooms. In columns (1) and (4), a special education classroom is identified when the teacher has a job assignment of special education or when the share of IEP students is higher than 80%; In columns (2) and (5), a special education classroom is identified when the teacher has a job assignment of special education and the share of IEP students in the class is higher than 25% or when the share of IEP students is higher than 80%; In columns (3) and (6), a special education classroom is identified when the share of IEP students is higher than 40%. "Co-Teaching" takes a value of 1 if a student was in a co-taught classroom in the subject area. The sample is restricted to students enrolled in grades three through eight that we observe in a co-taught class in the subject area at least once. "Elementary" takes a value of 1 if a student is below grade six. All regressions also control for an indicator for elementary grades. In columns (1) to (3), the sample is restricted to SWDs. In columns (4) to (6), the sample is restricted to students without disabilities. Standard errors are clustered at the student level and reported in the parenthesis. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Appendices

A Data Appendix

A.1 SIMS Data

The SIMS data comes from information transmitted from school districts to the state of Massachusetts. The data is reported to the state 3 times per year. We used only the information as reported in spring of the relevant academic year. The unit of observation in these data files is the student. After standardizing variable names to be common across years, we merged the files to generate a panel data set at the student-year level. From there, we resolved inconsistencies in the labeling of the data. For example, the gender variable sometimes coded males as "M" and others as "m"; we ensured such coding was common across all years.

A.2 MCAS Data

The state of Massachusetts provided MCAS data for the spring test administration spanning the years 2008 through 2018. After standardizing variable names across years, we merged the files into a panel data set at the student-year level. For the years 2008-2014 and 2017-2018, we use raw MCAS scores and standardize them within year and grade to have mean 0 and σ 1. For the years after 2015, some students in the state took Partnership for Assessment of Readiness for College and Careers (PARCC) test, some took MCAS, and others took both. For these years, we take the raw MCAS scores wherever available. The state was unable to locate raw PARCC scores for this study. For the 2015 test we used PARCC to MCASS concordance scores and for the 2016 test, we used PARCC theta scores.[11] Within each of these years, we standardized the test scores at the test-type (raw MCAS, PARCC concordance, PARCC theta) and grade level to have mean 0 and σ 1. In addition, we control for test-type (raw MCAS, PARCC concordance, PARCC theta) in the relevant value-added and school quality specifications. For years 2017 and 2018, we only use the

[11] These are transformed versions of the raw scores meant to adjust for question difficulty using techniques from item response theory.

MCAS scores.

A.3 EPIMS Data

The EPIMS data comes from information transmitted from school districts to the state of Massachusetts. The unit of observation in this data is a teacher-school-course-section-term. After standardizing variable names, we merged the files into a single data set at the teacher-year-school-course-term level. From there, we resolved inconsistencies in the data. For example, the gender variable sometimes coded males as "M" and others as "m"; we ensured such coding was common across all years.

A.4 SCS Data

The SCS data come from information transmitted from school districts to DESE. Prior to 2011, the state did not collect data on student coursework. The unit of observation in this data is a student-school-course-section-term. After standardizing variable names, we merged the files into a single data set at the student-year-school-course-section-term level. This data came to us with consistent year-to-year coding and required virtually no cleaning after merging. The research assistant working on this project was very excited about this development.

B Additional Tables

Table B1: Percentage of Students in Co-Taught Classrooms by Grade and Year, ELA Sample

	2011	2012	2013	2014	2015	2016	2017	2018	2018-2011
Students With Disabilities									
Grade 3	0.054	0.082	0.076	0.071	0.140	0.087	0.146	0.239	0.185
Grade 4	0.044	0.073	0.077	0.073	0.143	0.104	0.168	0.259	0.215
Grade 5	0.031	0.059	0.075	0.095	0.145	0.115	0.153	0.268	0.238
Grade 6	0.048	0.055	0.101	0.085	0.124	0.113	0.146	0.232	0.184
Grade 7	0.039	0.056	0.083	0.091	0.112	0.109	0.149	0.241	0.202
Grade 8	0.043	0.043	0.092	0.076	0.097	0.106	0.141	0.212	0.169
Students Without Disabilities									
Grade 3	0.044	0.065	0.065	0.057	0.109	0.069	0.132	0.173	0.128
Grade 4	0.036	0.062	0.063	0.052	0.119	0.084	0.146	0.211	0.175
Grade 5	0.024	0.055	0.057	0.079	0.117	0.089	0.129	0.197	0.172
Grade 6	0.029	0.044	0.064	0.054	0.084	0.069	0.095	0.146	0.117
Grade 7	0.023	0.020	0.043	0.050	0.071	0.068	0.089	0.131	0.108
Grade 8	0.014	0.023	0.044	0.045	0.051	0.048	0.065	0.104	0.090

Note: Table reports the percentage of students within Massachusetts taught in a co-taught classroom by grade and year. Top panel reports results for SWDs and the bottom panel reports results for students without disabilities. Final column reports the change in the percentage of students in co-taught classrooms within the respective grade from 2011 through 2018.

Table B2: Percentage Years Observed Co-Taught After First Co-Taught Year

	Students With Disabilities			Students Without Disabilities		
Obs After 1st Co-Taught	Students	Mean Co-Taught	St.Dev. Co-Taught	Observations	Mean Co-Taught	St.Dev. Co-Taught
1	7,785	0.334	0.472	18,654	0.194	0.396
2	5,552	0.238	0.351	13,212	0.127	0.259
3	5,195	0.221	0.276	13,147	0.145	0.225
4	2,921	0.209	0.258	7,500	0.110	0.181
5	1,669	0.179	0.229	4,694	0.115	0.177
6	904	0.166	0.213	2,500	0.067	0.116
7	241	0.165	0.187	632	0.081	0.127

Note: This table describes the percentage of student observations in years following their first co-taught year in which the student is also co-taught. Information is reported separately according to the number of years that the student is observed following their first co-taught year, and by whether or not the student has a disability. "Obs After 1st Co-Taught" reports the number of years that the student is observed following their first co-taught year, such that 1 indicates that the student is observed exactly 1 subsequent year (not necessarily the immediately following year) after the student is first observed in a co-taught classroom. Table includes all students in the estimation sample. Thus, only students who were first observed co-taught in 2011 could have 7 additional observations, and so on. Students could be no longer observed in the data either because they enter a non-tested grade or if they are no longer enrolled in a Massachusetts public school.

Table B3: Descriptive Statistics by Whether School is Observed to have Co-Teaching During Sample Period

	(1)	(2)	(3)	(4)
	School Co-Teach		School No Co-Teach	
	SWD	No SWD	SWD	No SWD
Female	0.366	0.527	0.368	0.528
White	0.811	0.794	0.874*	0.846*
Black	0.174	0.138	0.115*	0.097*
Hispanic	0.212	0.166	0.204	0.164
Asian	0.041	0.093	0.035*	0.081
Free Lunch	0.445	0.325	0.400*	0.287*
Reduced Lunch	0.045	0.041	0.042*	0.037
Autism	0.059		0.057	
Communication	0.148		0.148	
Emotional	0.059		0.055	
Health	0.116		0.111	
Neurological	0.049		0.048	
Specific Learning	0.285		0.283	
ELA Score	-0.781	0.215	-0.706*	0.273*
Math Score	-0.752	0.209	-0.680*	0.262*
Classroom Size	20.770	23.287	20.768	22.597*
% Obs Co-Taught	0.107	0.071	0.019*	0.009*
Observations	378186	1280337	315926	1080493

*$p < .05$ Relative to Respective Disability Category

Note: Table reports descriptive characteristics by special education status separately for schools in which we do or do not observed co-teaching during the sample period. Inferential statistics derived from regression controlling for school fixed-effects. Columns 1 and 2 include schools in which we observe at least one co-teaching observation, and Columns 3 and 4 include schools in which we do not observe co-teaching.

Table B4: Percentage of Students in Special Education by Grade and Year, ELA Sample

	2011	2012	2013	2014	2015	2016	2017	2018
Grade 3	0.147	0.148	0.148	0.147	0.153	0.150	0.159	0.169
Grade 4	0.160	0.160	0.158	0.160	0.162	0.164	0.169	0.177
Grade 5	0.153	0.166	0.163	0.167	0.169	0.162	0.177	0.182
Grade 6	0.163	0.166	0.172	0.167	0.171	0.167	0.173	0.182
Grade 7	0.163	0.166	0.168	0.166	0.168	0.168	0.172	0.174
Grade 8	0.167	0.162	0.163	0.163	0.167	0.162	0.170	0.174
G8-G3	0.019	0.014	0.014	0.015	0.014	0.011	0.010	0.005

Note: Table reports the percentage of students within Massachusetts who are in special education (i.e. have an Individual Education Plan) by grade and year. Final row reports the difference in the percentage of students in special education between grade 3 and grade 8 in a respective year.

Table B5: Effect of Co-Teaching by Grade Levels: IEP by Year

	(1)	(2)	(3)	(4)	(5)	(6)
			English Language Arts			
Co-Teaching	0.012**	0.013**	0.013**	0.001	-0.004	-0.001
	(0.005)	(0.006)	(0.006)	(0.002)	(0.003)	(0.003)
Co-Teaching ×		-0.003	-0.003		0.009**	0.008*
Elementary		(0.009)	(0.009)		(0.004)	(0.004)
Observations	105355	105355	105355	367936	367936	367936
			Mathematics			
Co-Teaching	0.027***	0.036***	0.036***	0.013***	0.015***	0.014***
	(0.005)	(0.005)	(0.005)	(0.002)	(0.003)	(0.003)
Co-Teaching ×		-0.024***	-0.023**		-0.004	-0.003
Elementary		(0.009)	(0.009)		(0.004)	(0.004)
Observations	95307	95307	95307	310552	310552	310552
Sample	SWD	SWD	SWD	No SWD	No SWD	No SWD
Student FE	Yes	Yes	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes
Class Demographics	No	No	Yes	No	No	Yes
SWD Share	No	No	Yes	No	No	Yes
SWD Share × SWD	No	No	Yes	No	No	Yes

Note: This table presents the effect of co-teaching on standardized ELA and Math scores. "Co-Teaching" takes a value of 1 if a student was in a co-taught classroom in the subject area. Analysis differs from primary regression in that students are defined as having a disability during years that they are in special education rather than being classified as ever having a disability. The sample is restricted to students enrolled in grades three through eight that we observe in a co-taught class in the subject area at least once. "Elementary" takes a value of 1 if a student is below grade six. All regressions also control for an indicator for elementary grades. In columns (1) to (3), the sample is restricted to SWDs. In columns (4) to (6), the sample is restricted to students without disabilities. Standard errors are clustered at the student level and reported in the parenthesis. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table B6: Effect of Co-Teaching by Demographics: SWD Sample

	ELA		Math	
	(1)	(2)	(3)	(4)
Autism	0.021		0.022	
	(0.017)		(0.016)	
Communication	0.037***		0.017*	
	(0.009)		(0.009)	
Emotional	0.022		0.028*	
	(0.017)		(0.017)	
Health	0.014		0.037***	
	(0.011)		(0.010)	
Neurological	0.022		0.037**	
	(0.017)		(0.016)	
Specific_Learning	0.030***	0.020**	0.045***	0.042***
	(0.007)	(0.009)	(0.006)	(0.008)
Specific_Learning		0.001		-0.001
$\times$ Elementary		(0.016)		(0.015)
Observations	143351	42585	129668	39025

Note: This table documents the effect of co-teaching for SWDs by special education classifications. The sample is restricted to SWDs enrolled in grades three through eight that we observe in a co-taught class in the subject area at least once. In columns (2) and (4), we further restrict to students with specific learning disabilities and interact the effect of co-teaching with elementary grades. Standard errors are clustered at the student level and reported in the parenthesis. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table B7: Effect of Co-Teaching: by Special Education Classroom

	English Launguage Arts			Mathematics		
	(1)	(2)	(3)	(4)	(5)	(6)
Co-Teaching	0.008	0.008	0.010*	0.034***	0.036***	0.035***
	(0.007)	(0.007)	(0.006)	(0.006)	(0.006)	(0.005)
Co-Teaching ×	0.010	0.012	0.006	-0.020*	-0.021*	-0.014*
Elementary	(0.011)	(0.012)	(0.009)	(0.011)	(0.011)	(0.009)
Sped Class	0.004	-0.002	0.003	0.008	0.007	0.000
	(0.007)	(0.007)	(0.005)	(0.006)	(0.006)	(0.005)
Sped Class ×	-0.023**	-0.013	-0.006	-0.014	-0.010	-0.009
Elementary	(0.011)	(0.011)	(0.008)	(0.011)	(0.011)	(0.008)
Observations	83171	80037	122546	74511	71957	111954
r2	0.763	0.761	0.767	0.772	0.768	0.777
F Test p-values						
Co-Teach + Interaction = 0	0.048	0.038	0.030	0.123	0.123	0.005
Sped Class + Interaction = 0	0.048	0.128	0.700	0.518	0.748	0.176
Sped Class Definition	Type 1	Type 2	Type 3	Type 1	Type 2	Type 3
Student FE	Yes	Yes	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes

Note: This table presents the effect of co-teaching on standardized ELA and Math scores with special education classroom controls. Each column represents models using different ways of identifying special education classrooms. In columns (1) and (4), a special education classroom is identified when the teacher has a job assignment of special education or when the share of IEP students is higher than 80%; In columns (2) and (5), a special education classroom is identified when the teacher has a job assignment of special education and the share of IEP students in the class is higher than 25% or when the share of IEP students is higher than 80%; In columns (3) and (6), a special education classroom is identified when the share of IEP students is higher than 40%. "Co-Teaching" takes a value of 1 if a student was in a co-taught classroom in the subject area. The sample is restricted to students enrolled in grades three through eight that we observe in a co-taught class in the subject area at least once. "Elementary" takes a value of 1 if a student is below grade six. Standard errors are clustered at the student level and reported in the parenthesis. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table B8: Descriptive Statistics for Heterogeneity Analysis, SWD

	(1) Female	(2) Male	(3) White	(4) Black	(5) Hispanic
Female	1.000	0.000	0.369	0.365	0.383
White	0.795	0.791	1.000	0.158	0.817
Black	0.195	0.198	0.039	1.000	0.183
Hispanic	0.227	0.213	0.225	0.203	1.000
Asian	0.036	0.038	0.008	0.008	0.010
Free Lunch	0.497	0.470	0.423	0.726	0.763
Reduced Lunch	0.046	0.050	0.048	0.045	0.044
Autism	0.022	0.073	0.057	0.038	0.032
Communication	0.155	0.151	0.144	0.157	0.180
Emotional	0.048	0.060	0.052	0.078	0.066
Health	0.090	0.141	0.132	0.096	0.106
Neurological	0.056	0.049	0.054	0.043	0.043
Specific Learning	0.356	0.275	0.306	0.319	0.311
ELA Score	-0.801	-0.917	-0.818	-1.125	-1.159
Math Score	-0.952	-0.781	-0.791	-1.121	-1.099
Classroom Size	21.730	21.475	21.515	21.884	21.677
% Obs Co-Taught	0.251	0.252	0.251	0.260	0.251

Standard errors in parentheses

Note: This table reports sample characteristics for regressions reported in Table 5 Table compares observed characteristics for observations within schools that we observe offering co-teaching at any time ("School With Co-Teach") or that we do not observe with any co-teaching ("School With Co-Teach"). Test for inference compares Column 1 vs Column 3 and Column 2 vs Column 4. Race/ethnicity classifications are not mutually exclusive in the data.

Table B9: Descriptive Statistics for Heterogeneity Analysis, No SWD

	(1) Female	(2) Male	(3) White	(4) Black	(5) Hispanic
Female	1.000	0.000	0.520	0.533	0.534
White	0.768	0.773	1.000	0.154	0.808
Black	0.179	0.170	0.035	1.000	0.188
Hispanic	0.196	0.186	0.200	0.206	1.000
Asian	0.080	0.082	0.014	0.011	0.013
Free Lunch	0.391	0.376	0.318	0.674	0.718
Reduced Lunch	0.044	0.045	0.041	0.054	0.051
ELA Score	0.211	-0.040	0.153	-0.273	-0.310
Math Score	0.059	0.089	0.128	-0.355	-0.329
Classroom Size	23.910	23.869	23.665	24.927	24.432
% Obs Co-Taught	0.237	0.234	0.232	0.261	0.258

Standard errors in parentheses

Note: This table reports sample characteristics for regressions reported in Table 4. Table compares observed characteristics for observations within schools that we observe offering co-teaching at any time ("School With Co-Teach") or that we do not observe with any co-teaching ("School With Co-Teach"). Test for inference compares Column 1 vs Column 3 and Column 2 vs Column 4.Race/ethnicity classifications are not mutually exclusive in the data.

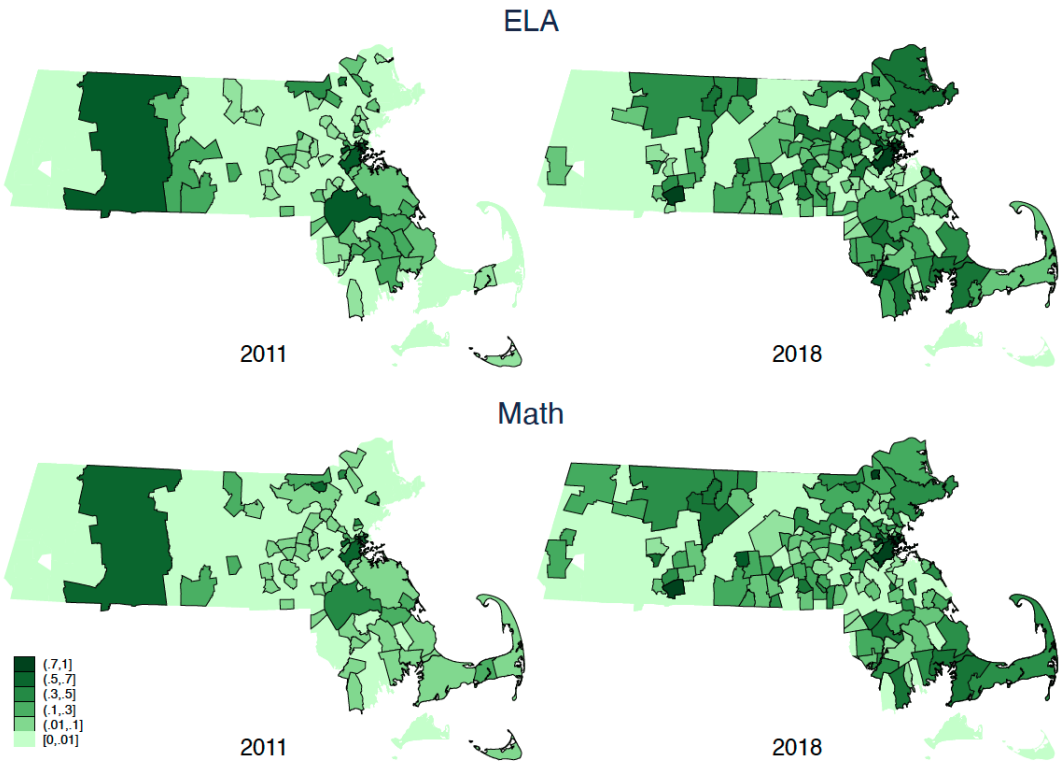
Table B10: Effect of Co-Teaching: Student Fixed Effect and Lagged Dependent Variable

	Standardized Test Scores					
	(1)	(2)	(3)	(4)	(5)	(6)
	English Language Arts					
Co-Teaching	0.013***	-0.005	-0.009	-0.006**	-0.008**	-0.018***
	(0.005)	(0.006)	(0.007)	(0.003)	(0.004)	(0.004)
Co-Teaching ×	0.004	0.026***	0.033**	0.010**	0.011**	0.038***
Elementary	(0.008)	(0.010)	(0.014)	(0.004)	(0.006)	(0.008)
Observations	143351	108204	76582	340656	256711	177380
r2	0.778	0.546	0.482	0.807	0.589	0.542
F Test p-values						
Co-Teach + Interaction = 0	0.004	0.007	0.050	0.236	0.504	0.005
	Mathematics					
Co-Teaching	0.035***	0.014***	0.010	0.013***	-0.006*	-0.013***
	(0.005)	(0.005)	(0.006)	(0.003)	(0.003)	(0.004)
Co-Teaching ×	-0.022***	0.010	0.009	-0.003	0.016***	0.029***
Elementary	(0.008)	(0.010)	(0.014)	(0.004)	(0.006)	(0.008)
Observations	129668	97903	70175	285687	215431	150656
r2	0.793	0.576	0.508	0.845	0.672	0.615
F Test p-values						
Co-Teach + Interaction = 0	0.037	0.004	0.141	0.001	0.030	0.036
Sample	SWD	SWD	SWD	No SWD	No SWD	No SWD
Student FE	Yes	No	No	Yes	No	No
School FE	Yes	Yes	Yes	Yes	Yes	Yes
Prior Year Score	No	Yes	No	No	Yes	No
Two Year Prior	No	No	Yes	No	No	Yes

Note: This table reports the specification test results. In columns (1) and (4), the student fixed effects model is used. In columns (2) and (5), the model with controls for observed student demographics and a lagged dependent variable is used. In columns (3) and (6), the model with controls for observed student demographics and a two-year-lagged dependent variable is used. "Co-Teaching" takes a value of 1 if a student was in a co-taught classroom in the subject area. The sample is restricted to students enrolled in grades three through eight that we observe in a co-taught class in the subject area at least once. "Elementary" takes a value of 1 if a student is below grade six. In columns (1) to (3), the sample is restricted to SWDs. In columns (4) to (6), the sample is restricted to students without disabilities. Standard errors are clustered at the student level and reported in the parenthesis. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

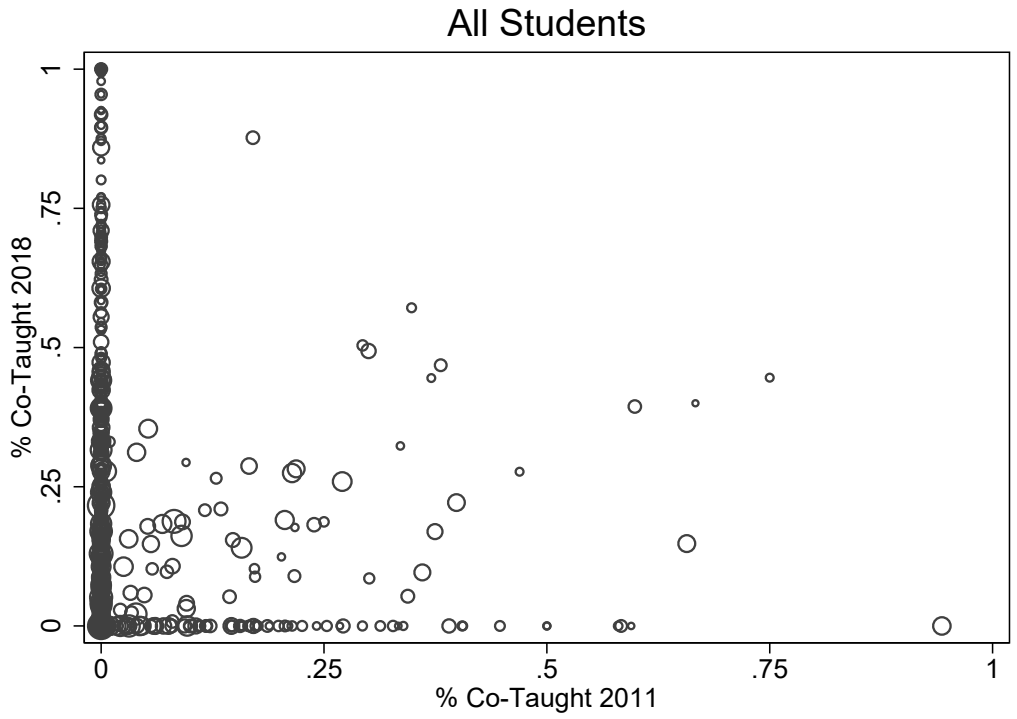
C Additional Figures

Figure C1: Change in the Share of Co-Taught Students by Massachusetts School Districts



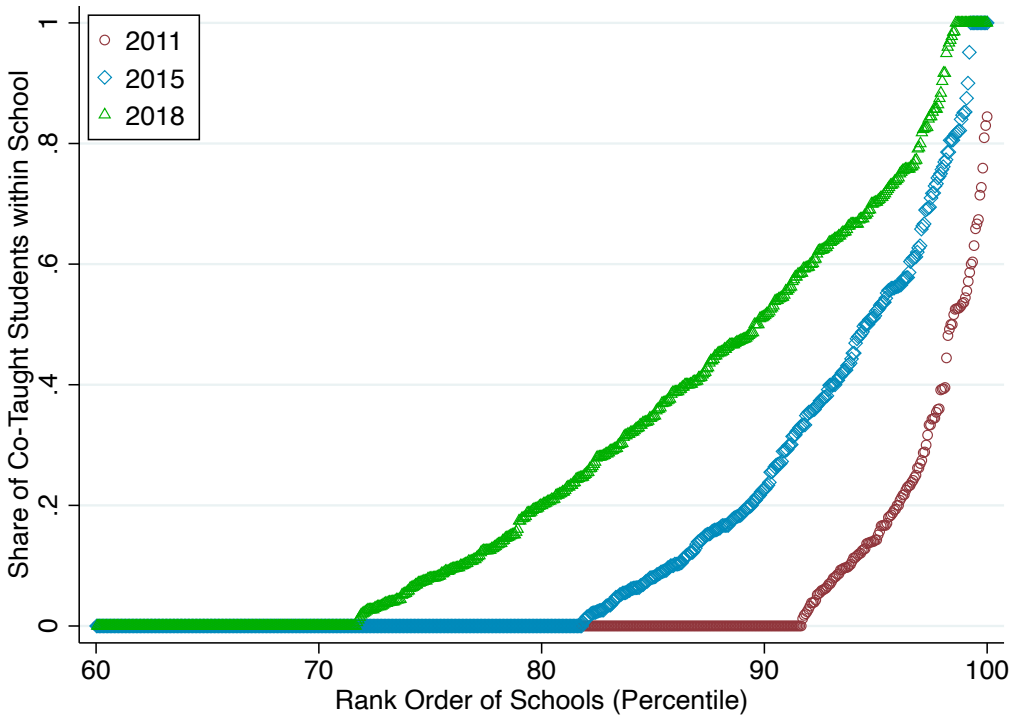
Note: This figure displays the share of co-taught students in 2011 and 2018 by Massachusetts school districts in the ELA and math samples we used to estimate the effect of co-teaching. The school districts with the darkest color have more than 70% of the students co-taught in the subject.

Figure C2: Share of Co-Taught SWDs within Schools: ELA Sample



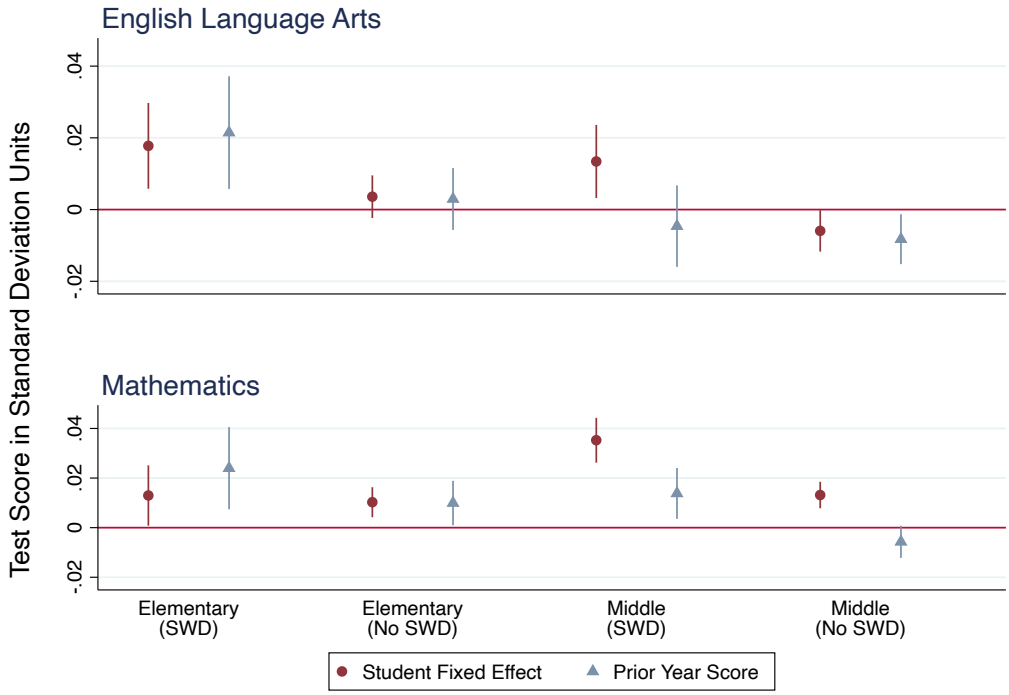
Note: This figure plots the percentage of all students in a co-taught classroom by school during the 2011 and 2018 school years. Each circle on the figure is a school in a given year. Size of circle illustrates total enrollment within the school in grades 3 through 8. The vertical axis is the percentage of students within the school that year who are in a co-taught classroom, and the horizontal axis simply ranks the schools according to their proportion of co-taught students in order to achieve a sloping line.

Figure C3: Share of Co-Taught SWDs within Schools: ELA Sample



Note: This figure plots the share of co-taught students within each school in the ELA sample. The circles, diamonds, and triangles represent the share of co-taught students in each school in the years 2011, 2015, and 2018, respectively. The x-axis plots the percentiles of the within-year distribution of the share of co-taught students.

Figure C4: Effect of Co-Teaching: Student Fixed Effect and Lagged Dependent Variable



Note: This figure illustrates the specification test results. Circles are the estimated effects of co-teaching from the student fixed effects model. Triangles are the estimated effects of co-teaching from the model with controls for observed student demographics and a lagged dependent variable. "Elementary" indicates grades from three through five. "Middle" indicates grades from six through eight. Vertical bars represent 95% confidence intervals. For the numerical values underlying this figure, see Table B10.

Suggested Citation: Jones, N., & Winters, M.A. (2022). *Are two teachers better than one? The effect of co-teaching on students with and without disabilities* (Working Paper 2020-1). Boston, MA: Wheelock Educational Policy Center. Available at wheelockpolicycenter.org.

OUR MISSION

The Wheelock Educational Policy Center (WEPC) conducts and disseminates rigorous, policy-relevant education research in partnership with local, state, and federal policymakers and stakeholders to improve pk-20 educational opportunities and holistic outcomes for underserved students.

www.wheelockpolicycenter.org
wheelockpolicy@bu.edu

Boston University Wheelock College of Education & Human Development
Wheelock Educational Policy Center

WHEELOCK EDUCATIONAL POLICY CENTER