

Service Delivery Models

Impacts for Students with and without Disabilities

Nathan Jones, Lindsey Kaler, Jessica Markham, Josefina Senese,
Marcus A. Winters

Boston University Wheelock College of Education & Human Development



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



Service Delivery Models: Impacts for Students with and without Disabilities

Abstract

Students with and without disabilities may be educated across a range of service delivery models (SDMs)—e.g., general education, co-taught, pull-out, and self-contained—but evidence for the relative effectiveness of these approaches at scale remains limited. We measured the effect of attending different SDMs on students' test scores, attendance rates, and number of disciplinary incidents. Our analysis leveraged within-student variation in SDM assignments and differences across students, applying student fixed effects and lagged outcomes models. These methods helped bound the causal effect within a narrow, policy-relevant range. We found that students with disabilities who participated in multiple instructional settings during their academic journey performed better in less restrictive environments, though the magnitude was often modest and varied across SDMs. For instance, elementary students scored significantly higher in math (0.06σ) in general education settings than in pull-out services, while no notable differences were observed in middle school or on ELA tests at any grade. Co-teaching had a statistically significant positive impact on students without disabilities. For elementary students, the effects ranged from 0.012σ in ELA to 0.017σ in math. The lagged outcome model produced similar estimates across all comparisons. This work contributes to our overall understanding of the effectiveness of inclusive practices as experienced by students statewide.

Keywords: special education, co-teaching, pull-out, self-contained, inclusion, service delivery models, placement

1 Introduction

In 2021, 7.3 million students—15% of the public school population—were identified as having a disability that requires special education (SPED) services (NCES, 2023b). Federal regulations mandate that students with disabilities (SWDs) receive these services in the least restrictive environment (ESSA, 2015; IDEA, 2004). This requirement has promoted a significant shift in the delivery of services from separate settings toward more inclusive alternatives. Since 1990, the percentage of SWDs spending at least 80% of their time in general education (GENED) classrooms has increased from 33% to 67% (NCES, 2023a). However, there exist numerous instructional and staffing models that allow schools to satisfy federal law, and determining the optimal setting for each student remains an ongoing challenge (e.g., Dunn, 1968; Karin et al., 2012; Zigmond, 2003).

Public schools can employ multiple service delivery models (SDMs), which vary in inclusiveness. These models include “push-in” strategies where specially designed instruction occurs within the GENED classroom, “pull-out” methods where students receive supplemental services in separate settings, and “self-contained” models where SWDs spend most of the day separated from their non-disabled peers (Epler & Ross, 2015, 2019; Kaler et al., 2024). SDM decisions are made on a case-by-case basis by a student’s educational team and family, who must weigh the advantages and disadvantages of each model and establish the best setting to fulfill the student’s Individualized Education Program (IEP) (C. M. Cole et al., 2004; Gilmour, 2018; Kaler et al., 2024).

However, there is limited consensus regarding the effectiveness of each SDM at scale (Gilmour, 2018; Ruijs & Peetsma, 2009). Some studies show that SWDs perform better when they spend most of their time in GENED settings, including gains in academic achievement (Gilmour et al., 2019; Rea et al., 2002; Sallin, 2021), attendance (Anderson, 2021; Gottfried et al., 2019; Rea et al., 2002), graduation rates (Malhotra, 2024; Schifter, 2016; Theobald et al., 2019), behavior (Anderson, 2021; Rea et al., 2002), and postsecondary participation (Lombardi et al., 2013; Sallin, 2021). Yet, other research shows no significant impact (C. M. Cole et al., 2004; Malhotra, 2024).

Even studies that show the benefits of GENED report modest effects, which vary by grade, setting, and student subgroup. For example, Anderson (2021) finds that students in less inclusive settings (40%–79% of the day in general education) are 0.59% more likely to receive a disci-

pline referral than those in general education 80% or more of the day. The effect is larger for males (0.86%), White/Asian students (0.71%), and non-economically disadvantaged students (0.69%). Malhotra (2024) finds that including students with disabilities in general classrooms boosts attendance by 0.409 percentage points for all students, with a larger effect for those without disabilities (0.429). However, there are no significant effects for students with disabilities. The author notes that, although the findings are statistically significant, the effect size is too small to be meaningful, with attendance rate changes falling within one percentage point.

Likewise, studies examining the impact of greater inclusion for SWDs on their non-disabled peers have yielded mixed results (Ruijs & Peetsma, 2009). Some scholarship reports negative effects on academic achievement (Fletcher, 2010), attendance (Gottfried et al., 2016; Theobald et al., 2019), and behavior (Gottfried, 2014), while other studies find no impact on academic achievement (Hanushek et al., 2002; Malhotra, 2024) or graduation rates (Malhotra, 2024).

The mixed findings in the literature partly arise from the endogenous nature of student assignments to SDMs. Placements are not random; they are deliberately based on students' individual needs and performance, making it challenging to control for selection bias and derive causal estimates (Anderson, 2021; S. M. Cole et al., 2021). For example, high-achieving SWDs often receive instruction in inclusive classrooms, while their low-performing peers participate in more restrictive settings. Therefore, the better outcomes observed in GENED likely reflect this placement bias (Gilmour, 2018).

To address this issue, we used longitudinal administrative data from Indiana to measure the effect of receiving instruction within different SDM pairings (e.g., GENED vs. pull-out) on student English Language Arts (ELA) and mathematics test scores, attendance rates, and number of disciplinary incidents. We present results from both a student fixed effects model, which leveraged variation in student assignments to one or another SDM, along with a lagged outcome model that drew on differences across students. Both models produced similar estimates, enabling us to draw policy-relevant conclusions.

The possibility that both fixed and dynamic factors influence SDM assignments brings up concerns about whether the identification assumptions behind both the student fixed effect and

lagged outcome models are met, and thus we do not necessarily claim that either produces a clear causal estimate. Rather, we leverage the fact that these strategies rely on opposing assumptions about the nature and direction of selection bias to bind the causal effect within a narrow range (Angrist & Pischke, 2009).

Findings suggest that SWDs who experienced multiple delivery models during their academic journey performed better within the less restrictive setting. However, the magnitude varied and, in many cases, was quite modest. With respect to students without disabilities, transitioning from GENE to co-taught settings increased test scores and reduced the number of disciplinary incidents. Nonetheless, consistent with recent evidence from Massachusetts (Jones & Winters, 2024), the magnitude of the effect when measured at a statewide scale was modest and much smaller than presented by previous experimental analyses (e.g., Friend, 2015; Friend et al., 2010; Murawski & Swanson, 2001).

Our results add to a limited body of research measuring the effect of educational settings on the outcomes for SWDs (e.g., Anderson, 2021; Lombardi et al., 2013), particularly that which measures outcomes at scale. We contribute to the literature both a methodological model for studying this topic and empirical evidence on the impact of multiple SDMs on students with and without disabilities in a longitudinal, statewide sample.

The paper proceeds as follows: in Section 2, we describe our data, in Section 3 we explain our empirical strategy, in Section 4 we report our findings, and in Section 5 we summarize our results, implications, and limitations.

2 Data

In this study, we utilized statewide administrative data on public and charter school students and teachers for 2011–2012 through 2021–2022, made available by the Indiana Department of Education. Student data included gender, race/ethnicity, English Learner and SPED classification, and Free and Reduced Price Meal status.

We conducted separate analyses on samples restricted to students with or without disabilities. We categorized a student as having a disability if, at any point in our dataset, they possessed

an IEP.¹ We observed approximately 180,000 SWDs and 750,000 students without disabilities in our sample (see Figures 1 and B2 for more precise sample sizes).

We focused on four main outcomes: ELA and mathematics test scores, attendance rates, and number of disciplinary incidents. Test outcomes represented standardized scores for students in grades 3–8, so models were limited to students in those grades. In contrast, attendance rates and disciplinary incidents were measured across grades 1–12, allowing models to include students from all grades.

Workforce data included course assignments, class roles, and license records. The data identifies the personnel who actively impart instruction within a given classroom and classifies them into seven categories, including co-teacher, resource room teacher, GENED teacher, and SPED teacher. Unfortunately, the data did not allow us to link non-certified personnel (e.g., paraeducators) to specific classrooms. As a result, our analysis was limited to examining the effects of SDMs that specify teacher assignments. For example, we were able to explore the impacts of co-teaching with two teachers but could not assess the effects of a teacher-paraeducator partnership.

We used a unique classroom identifier to match students to teachers. We then classified each student into one of four SDM types based on their teacher and classroom characteristics: (1) general education, where students are taught by a single GENED teacher; (2) co-taught, where two teachers share instruction in one classroom; (3) pull-out, where students spend part of the day in a GENED classroom and receive additional services from a resource-room teacher in a separate setting; and (4) self-contained, where a special educator provides instruction in a classroom dedicated to SWDs. SWDs may be in any of these SDMs, while students without disabilities participate only in GENED or co-taught settings.

Figure B1 in the Online Appendix illustrates our multi-stage procedure for identifying the SDM to which a student is exposed in a given year, grade, and subject. We defined SDMs based on teacher assignments for each subject—ELA, mathematics, science, and social studies—across all grade levels.² Students may engage in various combinations of these models across different

¹Notably, we observed about 5% of students in our sample as both a student with and without a disability. Removing these students from the estimation sample did not meaningfully impact our results. Thus, we maintained our definition of SWDs as any student who had ever had an IEP, irrespective of the timing of their identification.

²We included only these four subjects in our analysis for two reasons. First, the Indiana Department of Educa-

subjects and over time, and students identified with particular disabilities may be more likely to participate in certain SDM switching patterns than others.³

For a given subject, a student is identified as receiving instruction under a general education model if a single GENED teacher headed their class. A student was co-taught if at least one teacher in that subject was listed as a co-teacher or if we identified one SPED and one GENED educator leading the course. A student received pull-out services if their class had a resource room teacher or if the student attended a class led by a GENED teacher and they were also simultaneously enrolled in another course within the same subject that was comprised exclusively of SWDs and was overseen by a SPED teacher. Finally, a student participated in a self-contained setting if at least 80% of their classmates within that subject had a disability.

For analyses measuring impacts on ELA and math test scores, we used the SDM associated with each respective subject. The independent variable was binary: a student either participated in a specific SDM (e.g., co-taught) for a given subject (ELA or math) or did not. For non-test outcomes, we analyzed all four subject areas (ELA, mathematics, science, and social studies) and calculated the percentage of time each student spent in each SDM throughout the year. The independent variable was then continuous, representing the proportion of time spent in each SDM across all four subjects. If students, especially in the early grades, spent most of their class time with the same peers in the same classroom, the dosage measure was set to 1.

Figure B4 in the Online Appendix shows the switching patterns of students with and without disabilities, respectively, during the sample period. Because our empirical strategy involved comparing students to themselves in different SDMs, our estimations only drew from students who transitioned across models at least once during the sample period. Models that replaced the student fixed effect with a control for the lagged outcome required a minimum of two consecutive observations per student. These restrictions constrained the number of participating students in our sample. Figure 1 reveals that, out of the students who had at least two years of data, approximately

tion specifically identified these as pivotal for promoting future employment and success in post-secondary education (IDOE, 2021a, 2021b). Second, 97% of courses in the remaining subject areas occurred in GENED classrooms, lacking the variability in SDM assignment necessary for our analytical approach.

³See Figure B3 for a breakdown of placement switching types by disability classification. As Figure B4 reflects, students with some disability classifications were likely to engage in certain patterns more than others, which may drive some of our findings.

50% of SWDs engaged in only one SDM for ELA and math classes. Most of the remaining 40% of students participated in two different models. In comparison, less than 10% partook in three models throughout the study period, and almost none of the SWDs transitioned across all four models. Figure B2 in the Online Appendix shows that about 25% of students without disabilities attended both GENED and co-taught classrooms.

3 Empirical Strategy

We aimed to measure the causal effect of attending a given SDM on student outcomes. Because schools likely allocate students to instructional settings based on academic abilities, we suspected a naive comparison of student outcomes across SDMs would suffer from selection bias. To say it another way, students placed in more restrictive settings are more likely to have lower academic and behavioral outcomes than their peers in less restrictive settings, given that academic and behavioral outcomes are used in placement decision-making.

There are also likely unobserved differences between students in more versus less restrictive settings—for example, a student being especially socially withdrawn in a given year, or a parent’s particularly strong advocacy in IEP meetings regarding a student’s placement—that remained undetected in administrative data. Thus, simply comparing the outcomes of students in different SDMs, even when controlling for reported differences, is likely biased.

Unfortunately, we lack random assignment or naturally-occurring exogenous variation in SDM placement necessary for a single model to produce a credibly causal estimate. We addressed this challenge by estimating two models that make opposing assumptions about the source of selection bias. In addition to serving as a robustness check, the nature of selection bias within these models is such that they allow us to bound the causal effect of SDM assignment. In our case, the range of potential values for the causal effect of SDM assignment resulting from our analyses was sufficiently narrow to provide policy-relevant insights.

There are three general aspects of our analyses to understand before we detail the two regression models, their identifying assumptions, and their ability to bound the causal effect of SDM assignment. First, we reduced the potential for selection bias under both strategies by separately

estimating regressions restricted to students observed within each of the six respective combinations for pairs of SDMs.⁴ That is, we estimated models restricted to only students observed at least once in both a co-taught and pull-out settings, models restricted to students observed in both pull-out and self-contained settings, and so on. These pairwise comparisons focus our estimates on students at the intersection of one or another type of SDM assignment and more accurately reflect placement decisions available to school teams. For example, a student being considered for placement into a GENED setting or a co-taught classroom is not likely to also be considered for placement in a highly-restrictive self-contained setting for the same subject.⁵

Second, motivated by concern that the effect of each SDM could differ for younger and older students, for each model we incorporate an interaction between SDM and an indicator for whether the observation occurred within an elementary grade (grade 3–5). This interaction term is necessary to address the context that elementary schools and secondary schools in the U.S. have highly different structures (e.g., most students spend the majority of their day with a single teacher in elementary school but change teachers and classes frequently in middle and high school), coupled with literature documenting the negative impacts of transitioning to middle school on outcomes for students with and without disabilities (O’Hagan et al., 2024).

Third, for ease of interpretation, we focus the below discussion on regressions that measure the effect of SDM assignment on test scores. For these analyses, the treatment variable is an indicator for the student’s SDM in a given course. However, though the general features are the same, the analyses are slightly different when estimating impacts on attendance rate and number of disciplinary incidents. Unlike test scores, which are linked to particular subjects, for students who receive instruction in multiple classrooms during the school year there is no obvious strategy for determining which course is most responsible for non-test outcomes. Thus, when investigating these outcomes we adopted a dosage approach by replacing the indicator for SDM assignment with

⁴There are six possible SDM pair combinations: (1) GENED vs. co-taught, (2) GENED vs. pull-out, (3) GENED vs. self-contained, (4) pull-out vs. self-contained, (5) co-taught vs. pull-out, and (6) co-taught vs. self-contained. As previously noted, we observed very few students across three or more SDM combinations (see Figure 1 and Figure B2 in the Online Appendix).

⁵Figure B3 in the Online Appendix shows the distribution of SWDs across each SDM pairing based on disability type. For instance, 90% of students with language or speech impairments transition between math co-taught and GENED classrooms, while only 30% of students with intellectual disabilities do. Consequently, it is reasonable to anticipate that students with intellectual disabilities may be considered for more restrictive settings.

a variable for the percentage of classes among four core subjects (ELA, mathematics, science, and social studies) that student received instruction within a given SDM.⁶ For these analyses, the coefficient of interest represents the difference in the outcome for a student who spends 0% of their time in a given SDM relative to if they spent 100% of their time in that setting.

3.1 Student Fixed Effect and Lagged Outcome Regressions

We first applied a student fixed effect regression that leverages within-student variation in the timing of assignment to one or another SDM during the student’s academic career. Formally, within samples restricted to students observed attending each of two respective SDMs, our base model takes the form:

$$y_{ist} = \alpha_i + \theta_s + \sum_{j=1}^J \psi_j X_{ist} + \delta^{fe} type_{ist} + \zeta^{elementary}_{ist} + \lambda^{fe} (type_{ist} \times elementary_{ist}) + \varepsilon_{ist} \quad (1)$$

Where y is the outcome (math and ELA test scores, attendance rate, and disciplinary incidents) for student i in school s during the year t ; X is a vector of time-variant observed student characteristics, including an indicator for whether the student had an IEP that year; $Type$ indicates in which of the two SDMs within the pair the student received instruction for that subject;⁷ $elementary$ indicates whether a student was in an elementary or middle school grade (the test score analyses included only elementary and middle grades, while later regressions also incorporated high school grades); α and θ are fixed effects for the student and school; and ε is a stochastic term clustered by student. We are primarily interested in δ^{fe} and λ^{fe} , which respectively represent the relationship between SDM assignment and the outcome in middle grades and the extent to which this relationship differs within elementary grades.

Causal interpretation of $\hat{\delta}^{fe}$ and $\hat{\lambda}^{fe}$ relies on the assumption that there exist no unaccounted-

⁶Because non-test-score outcomes are not standardized, these analyses also include grade fixed effects.

⁷As described above, this variable becomes *Dosage* in models that use number of disciplinary incidents or attendance rate as the outcome.

for time-variant student attributes that correlate with both the timing of a student’s assignment to a given SDM and their outcome that year. That is, identification under this approach hinges on the assumption that students are assigned to SDMs based on fixed attributes. The primary threat to identification is that schools could place students into a less restrictive SDM following a year in which they had a higher outcome.⁸ For example, after a year in which a student placed in a co-taught ELA classroom performs exceptionally well on their end-of-year exam, their team may move them into a GENED classroom with no co-teacher for ELA the following year. We cannot rule out the potential for bias due to such dynamic selection into a more or less restrictive SDM. Indeed, it seems quite plausible that administrators would use a student’s performance at the end of a year to inform their SDM placement the following year.

We next estimate a lagged-outcome model that replaces the student fixed effect with controls for the lagged dependent variable ($y_{is(t-1)}$) and a vector of other observed time-invariant characteristics (F_i). This regression leverages variation in SDM assignments across students with the same lagged outcome and observed characteristics. Formally, we estimate:

$$y_{ist} = \alpha + \sum_{j=1}^J \psi_j X_{ist} + \xi y_{is(t-1)} + v F_i + \vartheta_s + \delta^{lag} type_{ist} + \zeta elementary_{ist} + \lambda^{lag} (type_{ist} \times elementary_{ist}) + \varepsilon_{ist} \quad (2)$$

In contrast to the student fixed effects approach, giving a causal interpretation to $\hat{\delta}^{lag}$ and $\hat{\lambda}^{lag}$ relies on the assumption that students are assigned to SDMs according to their outcome from the preceding year. The key threat to identification for equation 2 is the potential that fixed student attributes, in part, determine SDM assignments. As in the previous case, we cannot rule out, and indeed find it quite plausible, that unobserved fixed student characteristics are associated with both student outcomes and SDM assignment.

⁸Regressions reported in Tables C1 and C2 in the Online Appendix find that, for several of our pairwise comparisons, students’ prior year test scores are predictive of their SDM assignment the following year even after controlling for student fixed effects.

3.2 Bounding the Causal Effect of SDM Assignment

Considered in isolation, there is arguably little *a priori* reason to prefer either the student fixed effect or lagged outcome model, as neither the existence nor the magnitude of selection bias resulting from either strategy is knowable. Further, a regression that includes both a student fixed effect and control for the lagged outcome is consistent only under highly improbable assumptions, ruling out a model that directly accounts for both fixed and dynamic sources of selection (Angrist & Pischke, 2009). However, when considered together, the results from both models have a desirable bracketing property that allows us to bind the causal effect of SDM assignment. This strategy follows the reasoning described by Angrist and Pischke (2009).⁹

As they make opposing assumptions about its nature, the direction for potential selection bias under the student fixed effect and lagged outcome strategies occur in contrary directions. It can be shown that applying a student fixed effect model in the presence of positive dynamic selection leads to estimates for δ^{fe} and λ^{fe} that *understate* the causal effect of assignment to the less-restrictive environment, while applying the lagged outcome model in the presence of positive selection associated with fixed attributes results in estimates that *overstate* the causal effect of assignment to the less-restrictive environment.¹⁰ Therefore, the true underlying causal effect must be contained between the estimates from these two approaches. Consequently, even if neither approach gives a causal estimate in isolation, the results from the two models reveal a range for the value of the true underlying causal effect.

The above relationships, and thus our ability to bind the causal effect of interest, depend on the highly plausible assumption that selectivity based on lagged outcomes and fixed characteristics occur in the same direction. For our purposes, we assume that to the extent such sorting on fixed and/or dynamic attributes exists, students are positively selected into the less restrictive setting. That is, students are more likely to be placed into the less restrictive setting within the pair

⁹See Jones and Winters (2024) and Guryan (2004) for applications of this logic within the context of education research.

¹⁰See Online Appendix A for a more detailed description of the technical details and intuition underlying this decomposition of selection bias within both the student fixed effects and lagged outcome models. Our decomposition and reasoning follows closely the discussion of this issue presented by Angrist and Pischke (2009).

being analyzed if they have a higher lagged outcome and/or a higher fixed effect.¹¹

The extent to which our analyses within this bounding exercise produce informative results depends on the magnitude of the difference between the estimates from the fixed effect and lagged outcome models. If the two models produce very different estimates, then the range of potential values for the causal effect could be too wide to be informative. But, if the models produce fairly similar estimates then the causal effect will fall within a narrow range of potential values.

In practice, our analyses found little discrepancy between the student fixed effect and lagged outcome models. For example, when examining test scores, our estimates differed by more than 0.02σ in only 7 of the 24 comparisons for elementary and secondary grades, and in no case did the estimates vary significantly between the two approaches (see Table 1). The few instances where differences seemed more noticeable involved small samples that led to less precise estimates. While we reported results from both models in Section 4, for clarity, we focused on the estimates from the student fixed effect model and highlighted cases where the lagged outcome model produced notably different estimates. We also clarified the narrow bounds resulting from the two approaches by showing the estimates and 95% confidence intervals for each analysis in Figures B7–B8 in the Online Appendix.

3.3 Other Limitations with Interpretation

We urge readers to consider three additional features of our analyses when interpreting our estimates. First, our results only strictly apply to students who experience each of the two respective SDMs of interest for each comparison within Indiana’s current school system. Our estimates may not generalize to students who are consistently assigned to a particular SDM.¹² For instance, we find it unlikely that the average student who sometimes receives instruction in a pull-out setting and sometimes in a self-contained environment—who makes up the identifying sample in our comparison of these two SDMs—is representative of a student consistently placed in a self-

¹¹The results from the two models would still bind the causal effect if students were negatively sorted according both fixed and dynamic attributes. The bounding exercise would only fail under the seemingly unlikely scenario where selection due to fixed and dynamic sources are in opposing directions.

¹²To provide additional insight, Figures B5–B6 in the Online Appendix depict the percentage of students in each SDM by school district in 2021.

contained setting. Therefore, applying our findings broadly to SWDs whose SDM experiences differ from those in the estimation sample would be inappropriate.

Second, readers should keep in mind that we estimate the average effect of attending a specific SDM without accounting for variations in how schools implement instruction within that setting across the state. That is, we cannot peer inside the SDM “black box” or determine how student performance might change if the model were implemented according to best practices. From a policy perspective, this limitation is not necessarily a weakness, as our goal is to understand the effectiveness of SDMs as they are actually applied statewide, where variations in implementation fidelity are common. However, for those focused on identifying instructional strategies capable of improving student outcomes when implemented under ideal conditions, our estimates arguably have a limited interpretation.¹³

Third, while we defined each SDM strictly for our analysis, they are flexible and adaptable constructs in practice. For example, in an ideal co-taught classroom, both teachers would lead instruction for all students, irrespective of their ability. However, in reality, teachers may adopt a “one teach, one assist” model, where the GENED teacher delivers instruction, and the SPED teacher provides individual assistance to SWDs. As a result, the actual SDM received by SWDs resembles more of a resource room setting than a traditional co-taught one (e.g., Friend, 2015; Friend et al., 2010; Solis et al., 2012).

4 Results

Columns 1 and 2 from Table 1 report the results from student fixed effect and lagged outcome regressions, estimating the impact of attending a given SDM on ELA and math test scores for SWDs. We present estimates from models that restrict the sample to students we observed in both of a given pair of SDMs.¹⁴ Thus, the coefficient for each SDM is measured relative to the omitted SDM in the pair. Simply put, the results show the difference in outcomes for SWDs when

¹³Variations in the quality of execution within a given SDM across the state would bias our estimate for the effect of that strategy when implemented with fidelity toward zero.

¹⁴Table C3 in the Online Appendix presents SDM estimates by disability type. While the results are consistent with our main findings, they are less precise.

enrolled in one SDM compared to the other in the pair. For instance, in the top panel of Table 1, which compares GENED and co-taught, the coefficients reflect the effect of being in GENED vs co-taught, the omitted SDM.

We focus our description of results on the student fixed effects estimates while noting cases where the lagged outcome model appears to produce a meaningfully different estimate. For each pairwise comparison, the table also displays coefficients from the lagged outcome models. In most cases, the two approaches produce similar estimates, suggesting that differential selection to SDMs within the pair does not meaningfully influence our results.

The first three panels compare SWDs' performance when attending a GENED classroom SDM to their performance in a co-taught (Panel A), pull-out (Panel B), and self-contained (Panel C). Our results suggest that students who experience multiple settings benefit from attending the less restrictive one. Comparing GENED to pull-out services (Panel B), we found significant and meaningful (0.06σ) benefits for elementary students in math but no significant difference in middle school or for any grade level on the ELA test. Students in both grade levels and subjects performed substantially better in GENED settings than in self-contained classrooms (Panel C). Middle school students, in particular, benefited from attending a GENED classroom, scoring an estimated 0.058σ and 0.104σ higher than when they participated in a self-contained setting for both subjects. We detect a statistically significant, albeit modest (0.013σ for ELA and 0.019σ for math) benefit from attending an elementary co-taught class relative to a GENED one (Panel A). The respective lagged outcome model produces very similar estimates for each comparison.

Panel D compares SWDs' test scores in the most restrictive environments: receiving pull-out services versus participating in a self-contained classroom. Middle school students performed significantly better on the ELA test during years they were in pull-out settings relative to years they were in a self-contained environment (0.039σ). The coefficient for middle school students in math exhibits a similar trend but is not estimated precisely enough to be detected as significant. There are no significant differences in student performance for elementary students in either subject.

Panels E and F compare student outcomes when attending a co-taught relative to pull-out or self-contained settings. The observed pattern aligns with the comparisons involving these more

restrictive SDMs against GENED, although the estimates are less precise. Middle schoolers score significantly, and at times substantially, higher in co-taught settings than in self-contained or pull-out environments. The lagged outcome model again produces similar results for each comparison pair.

Columns 3 and 4 from Table 1 outline the estimated impact of participating in a given SDM on SWDs' attendance rates and number of disciplinary incidents. The samples encompass all SWDs in grades K through 12. Since these outcomes are not specifically tied to instruction in a particular subject, the variable of interest measures the percentage of the student's day spent within a specific SDM. Therefore, the coefficient measures the difference in the outcome if the student attends 100% of their time in a given setting relative to spending 100% in the comparison environment.

Each analysis reveals a significant difference in student attendance rate across the respective SDMs, and most of them identify significant results for disciplinary incidents. However, we find a precisely estimated but very small effect in each case. Among SWDs statewide, the average attendance rate is 0.948, with a standard deviation of 0.07, and the disciplinary incidents are averaged at 0.389, with a standard deviation of 1.337.

High school SWDs show a significant but modest improvement in attendance rates when enrolled in the least restrictive SDMs (GENED and co-taught) relative to when they participated in pull-out and self-contained settings. To illustrate, Table 1 Panel D highlights that receiving 100% of their instruction in GENED, as opposed to receiving pull-out services, leads to a 0.5% increase in high schoolers' attendance rate (equivalent to 0.07σ). We find no significant effects in middle grades.

In terms of disciplinary incidents, high schoolers benefited from attending a GENED classroom compared to receiving pull-out services (Panel B, Column 4). Receiving 100% of their instruction in GENED resulted in a reduction of 0.2 disciplinary incidents, equivalent to 0.17σ .

[TABLE 1 ABOUT HERE]

Table 2 presents the effect of attending a GENED classroom relative to a co-taught setting for students without disabilities. Notably, both student fixed effects and lagged dependent variable

models yield remarkably similar estimates, suggesting a limited influence of selection bias on our results. The fixed effects estimates reveal statistically significant yet modest benefits from attending a co-taught classroom across various grade levels and subjects. Specifically, for elementary students, these coefficients range from 0.012σ in ELA to 0.017σ in mathematics.

Finally, spending 100% of their day in a co-taught setting has statistically significant but small effects on attendance rates and disciplinary incidents. Among students without disabilities statewide, the average attendance rate is 0.955, with a standard deviation of 0.06. Additionally, disciplinary incidents have an average of 0.217 and a standard deviation of 0.996. Table 2 Columns 3–4 highlights that receiving 100% of their instruction in co-taught settings leads to a 0.1% increase in elementary students' attendance rate (equivalent to 0.02σ) and a 0.012 decrease in disciplinary incidents (equivalent to 0.01σ).

[TABLE 2 ABOUT HERE]

5 Discussion

The service delivery model (SDM) in which students receive their core instruction inarguably shapes their educational experience. Research suggests that different SDMs may influence outcomes not only for SWDs but also for students without disabilities. Yet, there is limited consensus on the effectiveness of each SDM at scale (Gilmour, 2018; Ruijs & Peetsma, 2009). Some studies show that SWDs perform better in GENED settings (e.g., Anderson, 2021; Gilmour et al., 2019; Sallin, 2021), while others find no significant impact (e.g., Malhotra, 2024). Research on non-disabled peers is also mixed, with some studies reporting negative effects (e.g., Gottfried, 2014; Gottfried et al., 2016; Theobald et al., 2019) and others finding no impact (Hanushek et al., 2002; Malhotra, 2024). Additionally, few studies explore SDMs beyond the time spent in general education classrooms (i.e., over 80%, 40–79%, or under 40%).

Our analysis contributes to this current literature by identifying SDMs via staffing and enrollment data, which allows us to discern some details on staffing models across classrooms. As such, we examined the impacts of four distinct SDMs on a continuum of restrictiveness: (1)

general education classrooms instructed by a single general education teacher, (2) co-taught general education classrooms served by both a general education and special education teacher, (3) resource room classrooms, and (4) self-contained special education classrooms.

Overall, our results suggest that SWDs who experience two SDMs tend to benefit from receiving instruction in the least restrictive one, though the magnitude is often modest and varies across SDMs. These findings align with recent literature, which shows modest benefits of instruction in inclusive settings, varying by grade and student subgroup (e.g., Anderson, 2021; Malhotra, 2024).

We also show that students without disabilities experience a statistically significant but small positive impact from co-teaching, particularly in elementary grades. These differences by school level may reflect variations in the structure of elementary, middle, and secondary schools or could be related to disruptions associated with the transition from elementary to middle school, impacting student outcomes (O'Hagan et al., 2024).

This paper provides new insights relevant to ongoing policy efforts toward increasing the service provision to SWDs in more inclusive environments. Some who favor instructing SWDs within inclusive settings may be disappointed by our generally null or small positive results. However, our findings also indicate that attending a more inclusive SDM does not meaningfully harm students' outcomes, even when examined at scale, which is encouraging systemically.

Moreover, our findings produce new evidence to supplement recent literature examining co-teaching's impact on outcomes for students with and without disabilities at scale (Jones & Winters, 2024). Consistent with that prior study, we uncover only minor positive impacts from co-teaching, suggesting that the exceptionally large results reported in some earlier investigations might not accurately reflect the typical experience of public school students.

Though we caution that our estimates only strictly apply to students who may experience both SDMs from a given pair, our results could inform practitioners as they consider further advocating for the appropriate educational environment for students they believe may benefit from one or another setting.

It is also important to recognize that this study is based on a specific state context. While

it is encouraging that our findings align with those of Jones and Winters (2024) in Massachusetts, and we account for contextual factors like urbanicity through school fixed effects, there are still unique aspects of this context that should be considered when interpreting our results. Therefore, researchers should continue exploring the differential impacts of various SDMs across different contexts to better equip practitioners to make evidence-based decisions about student placement.

Future research should also explore questions such as: How might these effects vary in non-public school settings? Are there differential impacts of each SDM based on factors like teacher quality, experience, or working conditions? What external factors (e.g., school funding mechanisms and teacher labor markets) influence the likelihood of students being placed in a particular SDM, and how might this affect their academic trajectories?

Many questions remain unanswered, highlighting the need to better understand how educational service delivery functions at scale, to evaluate the effectiveness of each model, and to identify strategies to improve outcomes for students with and without disabilities in school and beyond.

References

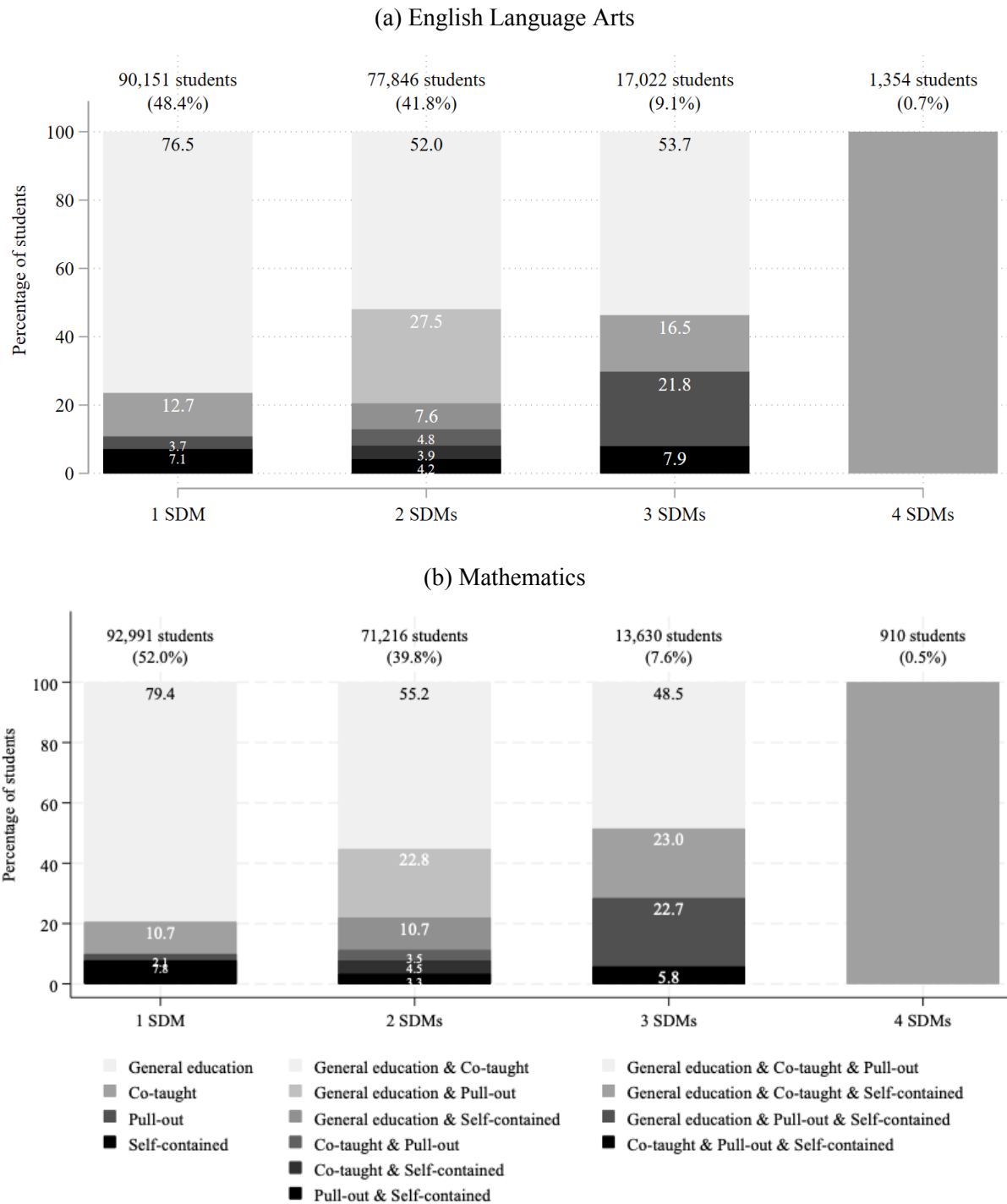
- Anderson, K. P. (2021). The relationship between inclusion, absenteeism, and disciplinary outcomes for students with disabilities. *Educational Evaluation and Policy Analysis*, 43(1), 32–59. <https://doi.org/10.3102/0162373720968558>
- Angrist, J. D., & Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.
- Cole, C. M., Waldron, N., & Majd, M. (2004). Academic progress of students across inclusive and traditional settings. *Mental retardation*, 42(2), 136–144. [https://doi.org/10.1352/0047-6765\(2004\)42<136:aposai>2.0.co;2](https://doi.org/10.1352/0047-6765(2004)42<136:aposai>2.0.co;2)
- Cole, S. M., Murphy, H. R., Frisby, M. B., Grossi, T. A., & Bolte, H. R. (2021). The relationship of special education placement and student academic outcomes. *The Journal of Special Education*, 54(4), 217–227. <https://doi.org/10.1177/0022466920925033>
- Dunn, L. M. (1968). Special education for the mildly retarded—Is much of it justifiable? *Exceptional Children*, 35(1), 5–22. <https://doi.org/10.1177/001440296803500101>
- Epler, P. L., & Ross, R. (2015). Providing services to students with special needs across various settings. In *Models for effective service delivery in special education programs* (pp. 1–28). IGI Global. <https://doi.org/10.4018/978-1-4666-7397-7.ch001>
- Epler, P. L., & Ross, R. (2019). Special education service delivery models around the globe. In *Pre-service and in-service teacher education: Concepts, methodologies, tools, and applications* (pp. 599–621). IGI Global. <https://doi.org/10.4018/978-1-5225-7305-0.ch029>
- ESSA. (2015). Every Student Succeeds Act [20 U.S.C. § 6301].
- 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. <https://doi.org/10.1002/pam.20479>
- Friend, M. (2015). Welcome to co-teaching 2.0. *Educational Leadership*, 73(4), 16–22.
- Friend, M., Cook, L., Hurley-Chamberlain, D., & 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. <https://doi.org/10.1080/10474410903535380>

- Gilmour, A. F. (2018). Has inclusion gone too far? Weighing its effects on students with disabilities, their peers, and teachers. *Education Next*, 18(4), 8–16.
- Gilmour, A. F., Fuchs, D., & 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. <https://doi.org/10.1177/0014402918795830>
- Gottfried, M. A. (2014). Classmates with disabilities and students' noncognitive outcomes. *Educational Evaluation and Policy Analysis*, 36(1), 20–43. <https://doi.org/10.3102/0162373713493130>
- Gottfried, M. A., Egalite, A., & 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. <https://doi.org/10.1016/j.ecresq.2016.02.002>
- Gottfried, M. A., Stiefel, L., Schwartz, A. E., & Hopkins, B. (2019). Showing up: Disparities in chronic absenteeism between students with and without disabilities in traditional public schools. *Teachers College Record*, 121(8), 1–34. <https://doi.org/10.1177/016146811912100808>
- Guryan, J. (2004). Desegregation and black dropout rates. *American Economic Review*, 94(4), 919–943. <https://doi.org/10.1257/0002828042002679>
- Hanushek, E. A., Kain, J. F., & 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. <https://doi.org/10.1162/003465302760556431>
- IDEA. (2004). *Individuals With Disabilities Education Act*. <https://sites.ed.gov/idea/>
- IDOE. (2021a). *Elementary and middle level subjects and descriptions 2021–2022*. <https://www.in.gov/doe/students/indiana-academic-standards>
- IDOE. (2021b). *High school college and career readiness course titles and descriptions 2021–2022*. <https://www.in.gov/doe/students/indiana-academic-standards>
- Jones, N., & Winters, M. A. (2024). Are two teachers better than one? the effect of coteaching on students with and without disabilities. *Journal of Human Resources*, 59(4), 1180–1206. <https://doi.org/10.3368/jhr.0420-10834R3>

- Kaler, L., Markham-Anderson, J., & Jones, N. (2024). Service delivery models and outcomes for students with disabilities. *Remedial and Special Education*. <https://doi.org/10.1177/07419325241268747>
- Karin, H., Evelien, C., Mieke, H., & Katja, P. (2012). Don't pull me out!? Preliminary findings of a systematic review of qualitative evidence on experiences of pupils with special educational needs in inclusive education. *Procedia-Social and Behavioral Sciences*, *69*, 1709–1713. <https://doi.org/10.1016/j.sbspro.2012.12.118>
- Lombardi, A., Doren, B., Gau, J. M., & Lindstrom, L. E. (2013). The influence of instructional settings in reading and math on postsecondary participation. *Journal of Disability Policy Studies*, *24*(3), 170–180. <https://doi.org/10.1177/1044207312468766>
- Malhotra, K. P. (2024). Whose idea is this? An examination of the effectiveness of inclusive education. *Educational Evaluation and Policy Analysis*. <https://doi.org/10.3102/01623737241257951>
- Murawski, W. W., & Swanson, H. L. (2001). A meta-analysis of co-teaching research: Where are the data? *Remedial and Special Education*, *22*(5), 258–267. <https://doi.org/10.1177/074193250102200501>
- Naqvi, A. (2024, October 16). *Stata package “sankey”* (Version 1.81). <https://github.com/asjadnaqvi/stata-sankey>
- NCES. (2023a). *Percentage distribution of school-age students served under Individuals with Disabilities Education Act (IDEA), Part B, by educational environment and type of disability: Selected years, Fall 1989 through Fall 2021*. National Center for Education Statistics, U.S. Department of Education, Institute of Education Sciences. https://nces.ed.gov/programs/digest/d22/tables/dt22_204.60.asp
- NCES. (2023b). *Students with disabilities*. National Center for Education Statistics, U.S. Department of Education, Institute of Education Sciences. <https://nces.ed.gov/programs/coe/indicator/cgg>
- O'Hagan, K. G., Stiefel, L., & Schwartz, A. E. (2024). The academic effects of moving to middle school on students with disabilities relative to their general education peers. *Educational Evaluation and Policy Analysis*. <https://doi.org/10.3102/016237372412299>

- Rea, P. J., McLaughlin, V. L., & Walther-Thomas, C. (2002). Outcomes for students with learning disabilities in inclusive and pullout programs. *Exceptional Children*, 68(2), 203–222. <https://doi.org/10.1177/001440290206800204>
- Ruijs, N. M., & Peetsma, T. T. (2009). Effects of inclusion on students with and without special educational needs reviewed. *Educational Research Review*, 4(2), 67–79.
- Sallin, A. (2021). Estimating returns to special education: Combining machine learning and text analysis to address confounding. *arXiv*. <https://doi.org/10.48550/arXiv.2110.08807>
- Schifter, L. A. (2016). Using survival analysis to understand graduation of students with disabilities. *Exceptional Children*, 82(4), 479–496. <https://doi.org/10.1177/0014402915619418>
- Solis, M., Vaughn, S., Swanson, E., & Mcculley, L. (2012). Collaborative models of instruction: The empirical foundations of inclusion and co-teaching. *Psychology in the Schools*, 49(5), 498–510. <https://doi.org/10.1002/pits.21606>
- Theobald, R. J., Goldhaber, D. D., Gratz, T. M., & Holden, K. L. (2019). Career and technical education, inclusion, and postsecondary outcomes for students with learning disabilities. *Journal of Learning Disabilities*, 52(2), 109–119. <https://doi.org/10.1177/0022219418775121>
- Zigmond, N. (2003). Where should students with disabilities receive special education services? Is one place better than another? *The Journal of Special Education*, 37(3), 193–199. <https://doi.org/10.1177/00224669030370030901>

Figure 1: Share of Students with Disabilities by Number of Service Delivery Models



Notes: This figure shows the percentage of students with disabilities who have attended 1, 2, 3, or 4 service delivery models (SDMs) across the sample period (labeled “1 SDM,” “2 SDMs,” “3 SDMs,” and “4 SDMs,” respectively). The sample is restricted to students enrolled in grades three through eight for ELA (Panel A) and Mathematics (Panel B), respectively. Sample sizes and proportions are indicated above each bar, reflecting the number and share of students with disabilities who experienced 1, 2, 3, or 4 SDMs in each subject across the sample period.

Table 1: Service Delivery Model Effects. Students with Disabilities

	(1)		(2)		(3)		(4)	
	English Language Arts		Mathematics		Attendance Rate		Disciplinary Incidents	
<i>(a) General education vs. Co-taught</i>								
GENED	0.005 (0.006)	0.012 (0.009)	-0.014* (0.008)	0.001 (0.011)	0.003*** (0.001)	0.004*** (0.001)	-0.033*** (0.012)	-0.102*** (0.012)
GENED × Elementary	-0.018***††† (0.009)	-0.023* (0.014)	-0.005††† (0.010)	-0.003 (0.015)	-0.004***††† (0.001)	-0.003*** (0.001)	0.048***††† (0.014)	0.115***† (0.014)
GENED × Middle					-0.004*** (0.001)	-0.004*** (0.001)	-0.072***††† (0.019)	-0.053***††† (0.020)
Observations	51,150	32,199	46,859	28,817	498,278	407,372	500,211	409,276
R^2	0.849	0.615	0.846	0.607	0.549	0.357	0.540	0.279
<i>(b) General education vs. Pull-out</i>								
GENED	0.009 (0.007)	0.009 (0.010)	-0.003 (0.009)	-0.012 (0.013)	0.005*** (0.001)	0.006*** (0.001)	-0.053*** (0.017)	-0.121*** (0.017)
GENED × Elementary	0.003 (0.012)	0.038***††† (0.018)	0.063***††† (0.014)	0.085***††† (0.022)	-0.005*** (0.001)	-0.005***††† (0.001)	0.021 (0.028)	0.049*††† (0.029)
GENED × Middle					-0.003** (0.001)	-0.003*††† (0.001)	-0.169***††† (0.033)	-0.172***††† (0.035)
Observations	37,449	23,686	26,335	16,623	216,793	177,747	217,480	178,488
R^2	0.818	0.549	0.817	0.567	0.537	0.353	0.540	0.290
Lagged Outcome	No	Yes	No	Yes	No	Yes	No	Yes
Student FE	Yes	No	Yes	No	Yes	No	Yes	No
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grade FE	No	No	No	No	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; $H_0 : \beta = 0$

† $p < 0.10$, †† $p < 0.05$, ††† $p < 0.01$; $H_0 : \beta + \lambda = 0$

Notes: This table presents the effect of attending a general education (GENED) classroom versus a co-taught (top panel) and pull-out (bottom panel) setting on SWDs' standardized ELA and Math scores, attendance rate, and the number of disciplinary incidents. The sample includes SWDs in grades 3–8 for test scores and grades 1–12 for non-test outcomes, with students observed in each service delivery model pair (e.g., GENED vs. co-taught) at least once. For test scores, “GENED” equals 1 if a student was in a GENED classroom for that subject. For non-test outcomes, “GENED” equals 1 if the student received instruction for in GENED classrooms for all subjects (ELA, Mathematics, Social Studies, and Science). “Elementary” equals 1 for students below sixth grade, and “Middle” equals 1 for grades 6–8. Each outcome includes two columns, the first of which (left column) reflects results from the school, and student fixed effects model, and the second of which (right column) reflects results from the lagged dependent model with school fixed effects. The lagged outcome used for the second approach is the student's ELA, Math, attendance rate, and number of disciplinary incidents during the previous year. Both models include student demographics as controls: gender, race, SPED classification, disability type, English Learner status, and free or reduced meal eligibility. For non-test outcomes, we also add grade fixed effects for all the models. Standard errors are clustered at the student level and reported in parentheses.

	(1)		(2)		(3)		(4)	
	English Language Arts		Mathematics		Attendance Rate		Disciplinary Incidents	
<i>(c) General education vs. Self-contained</i>								
General education	0.058*** (0.015)	0.069*** (0.024)	0.104*** (0.018)	0.138*** (0.026)	-0.003*** (0.001)	0.001 (0.001)	-0.022 (0.016)	-0.095*** (0.014)
GENED × Elementary	-0.008†† (0.027)	-0.023 (0.042)	-0.041††† (0.027)	-0.038††† (0.039)	0.006***††† (0.001)	0.002††† (0.001)	0.062***†† (0.022)	0.115*** (0.021)
GENED × Middle					0.003** (0.001)	0.001††† (0.001)	0.001 (0.024)	0.013††† (0.024)
Observations	8,458	5,229	10,076	6,303	261,659	206,372	262,640	207,334
R^2	0.779	0.515	0.742	0.466	0.539	0.358	0.548	0.302
<i>(d) Pull-out vs. Self-contained</i>								
Pull-out	0.039** (0.018)	0.028 (0.029)	0.034 (0.025)	0.069* (0.039)	-0.006*** (0.002)	-0.008*** (0.002)	0.031 (0.041)	0.014 (0.036)
Pull-out × Elementary	-0.042 (0.026)	-0.035 (0.042)	-0.008 (0.035)	-0.035 (0.054)	0.007*** (0.002)	0.007** (0.003)	0.121**††† (0.058)	0.178***††† (0.065)
Pull-out × Middle					0.004 (0.003)	0.003†† (0.003)	0.489***††† (0.083)	0.595***††† (0.086)
Observations	8,096	4,718	5,838	3,466	62,114	50,362	62,280	50,514
R^2	0.782	0.519	0.758	0.474	0.631	0.373	0.647	0.344
Lagged Outcome	No	Yes	No	Yes	No	Yes	No	Yes
Student FE	Yes	No	Yes	No	Yes	No	Yes	No
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grade FE	No	No	No	No	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; $H_0 : \beta = 0$

† $p < 0.10$, †† $p < 0.05$, ††† $p < 0.01$; $H_0 : \beta + \lambda = 0$

Notes: This table presents the effect of attending a general education (GENED) classroom (top panel) and a pull-out setting (bottom panel) versus a self-contained one on SWDs' standardized ELA and Math scores, attendance rate, and the number of disciplinary incidents. The sample includes SWDs in grades 3–8 for test scores and grades 1–12 for non-test outcomes, with students observed in each service delivery model pair (e.g., GENED vs. self-contained) at least once. For test scores, "GENED" equals 1 if a student was in a GENED classroom for that subject. For non-test outcomes, "GENED" equals 1 if the student received instruction for in GENED classrooms for all subjects (ELA, Mathematics, Social Studies, and Science). "Elementary" equals 1 for students below sixth grade, and "Middle" equals 1 for grades 6–8. Each outcome includes two columns, the first of which (left column) reflects results from the school, and student fixed effects model, and the second of which (right column) reflects results from the lagged dependent model with school fixed effects. The lagged outcome used for the second approach is the student's ELA, Math, attendance rate, and number of disciplinary incidents during the previous year. Both models include student demographics as controls: gender, race, SPED classification, disability type, English Learner status, and free or reduced meal eligibility. For non-test outcomes, we also add grade fixed effects for all the models. Standard errors are clustered at the student level and reported in parentheses.

	(1)		(2)		(3)		(4)	
	English Language Arts		Mathematics		Attendance Rate		Disciplinary Incidents	
<i>(e) Co-taught vs. Pull-out</i>								
Co-taught	0.035*** (0.012)	0.022 (0.019)	0.051*** (0.017)	0.020 (0.026)	0.005*** (0.002)	0.007*** (0.002)	0.003 (0.037)	0.026 (0.034)
Co-taught × Elementary	-0.013 (0.022)	-0.023 (0.038)	0.031††† (0.030)	0.016 (0.051)	-0.004* (0.002)	-0.005* (0.002)	-0.086*†† (0.050)	-0.086*† (0.047)
Co-taught × Middle					-0.003 (0.003)	-0.002††† (0.003)	-0.086 (0.071)	-0.062 (0.070)
Observations	10,636	6,754	6,850	4,321	58344	49429	58601	49,669
R^2	0.806	0.515	0.806	0.528	0.614	0.363	0.612	0.302
<i>(f) Co-taught vs. Self-contained</i>								
Co-taught	0.071*** (0.021)	0.076** (0.034)	0.214*** (0.028)	0.226*** (0.040)	0.003** (0.001)	0.004*** (0.001)	-0.009 (0.028)	-0.039 (0.030)
Co-taught × Elementary	-0.036 (0.037)	0.041†† (0.058)	-0.172*** (0.044)	-0.171** (0.073)	-0.002 (0.002)	-0.006*** (0.002)	0.017 (0.038)	0.063 (0.039)
Co-taught × Middle					-0.002 (0.002)	-0.004** (0.002)	0.085*†† (0.047)	0.080 (0.049)
Observations	4,609	2,801	4,273	2,592	97,135	78,426	97,557	78,825
R^2	0.759	0.445	0.730	0.440	0.580	0.361	0.617	0.334
Lagged Outcome	No	Yes	No	Yes	No	Yes	No	Yes
Student FE	Yes	No	Yes	No	Yes	No	Yes	No
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grade FE	No	No	No	No	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; $H_0: \beta = 0$

† $p < 0.10$, †† $p < 0.05$, ††† $p < 0.01$; $H_0: \beta + \lambda = 0$

Notes: This table presents the effect of attending a co-taught classroom versus a pull-out (top panel) and a self-contained (bottom panel) setting on SWDs' standardized ELA and Math scores, attendance rate, and the number of disciplinary incidents. The sample includes SWDs in grades 3–8 for test scores and grades 1–12 for non-test outcomes, with students observed in each service delivery model pair (e.g., co-taught vs. pull-out) at least once. For test scores, “co-taught” equals 1 if a student was in a co-taught classroom for that subject. For non-test outcomes, “co-taught” equals 1 if the student received instruction for in co-taught classrooms for all subjects (ELA, Mathematics, Social Studies, and Science). “Elementary” equals 1 for students below sixth grade, and “Middle” equals 1 for grades 6–8. Each outcome includes two columns, the first of which (left column) reflects results from the school, and student fixed effects model, and the second of which (right column) reflects results from the lagged dependent model with school fixed effects. The lagged outcome used for the second approach is the student's ELA, Math, attendance rate, and number of disciplinary incidents during the previous year. Both models include student demographics as controls: gender, race, SPED classification, disability type, English Learner status, and free or reduced meal eligibility. For non-test outcomes, we also add grade fixed effects for all the models. Standard errors are clustered at the student level and reported in parentheses.

Table 2: Service Delivery Model Effects. Students Without Disabilities

	(1)	(2)	(3)	(4)
	English Language Arts	Mathematics	Attendance Rate	Disciplinary Incidents
<i>General education vs. Co-taught</i>				
GENED	-0.006** (0.002)	-0.021*** (0.003)	0.003*** (0.000)	-0.032*** (0.006)
GENED × Elementary	-0.006*††† (0.003)	0.004††† (0.004)	-0.004***††† (0.000)	0.040***††† (0.007)
GENED × Middle			-0.002*** (0.000)	0.015*††† (0.008)
Observations	332,336	275,386	2,134,591	2,142,583
R ²	0.844	0.858	0.521	0.479
Lagged Outcome	No	Yes	No	No
Student FE	Yes	No	Yes	Yes
School FE	Yes	Yes	Yes	Yes
Grade FE	No	No	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; $H_0 : \beta = 0$

† $p < 0.10$, †† $p < 0.05$, ††† $p < 0.01$; $H_0 : \beta + \lambda = 0$

Notes: This table presents the effect of attending a general education (GENED) classroom versus a co-taught on standardized ELA and Math scores, attendance rate, and the number of disciplinary incidents. The sample includes students without disabilities in grades 3–8 for test scores and grades 1–12 for non-test outcomes, with students observed GENED and co-taught classroom at least once. For test scores, “GENED” equals 1 if a student was in a GENED classroom for that subject. For non-test outcomes, “GENED” equals 1 if the student received instruction for in GENED classrooms for all subjects (ELA, Mathematics, Social Studies, and Science). “Elementary” equals 1 for students below sixth grade, and “Middle” equals 1 for grades 6–8. Each outcome includes two columns, the first of which (left column) reflects results from the school, and student fixed effects model, and the second of which (right column) reflects results from the lagged dependent model with school fixed effects. The lagged outcome used for the second approach is the student’s ELA, Math, attendance rate, and number of disciplinary incidents during the previous year. Both models include student demographics as controls: gender, race, SPED classification, disability type, English Learner status, and free or reduced meal eligibility. For non-test outcomes, we also add grade fixed effects for all the models. Standard errors are clustered at the student level and reported in parentheses.

A Bounding Exercise

This Appendix offers additional technical details and explains the intuition behind our claim. Specifically, it demonstrates how the results from the student fixed-effects and lagged outcome models produce bounds for the causal effect of assignment to one SDM versus another. Throughout this section, we follow closely the presentation of this issue documented in Section 5.4 of Angrist and Pischke (2009). The only difference is that, in our context, selection bias occurs in the opposite direction of the example they discuss. We restate their case here to help readers unfamiliar with this argument. We do not claim ours as a novel application or description of the issue.

We begin by setting some basic parameters. For each case, we consider a simplified model comparing across two SDM pairs, ignoring covariates, and assuming only two periods. For this discussion, we also remove the potential for differential impacts by grade level, and thus do not include the grade-by-SDM interactions found in our primary analyses. Further, we assume that all students are in the more-restrictive setting of the pair during the first period, and we define the “treatment” ($D_{it} = 1$) as assignment to the less-restrictive SDM of the pair during period t . All resulting insights also apply to more general settings.

We assume that students are positively sorted into the less-restrictive setting. That is, for the respective case we assume that the source of bias is that, all else equal, students are more likely to be placed into the less-restrictive setting if they scored higher at the end of the previous year or that they have more desirable fixed attributes. We further assume that the true causal effect of attending a less-restrictive setting is positive. The reasoning behind the bounding exercise still holds. However, the bias direction for each model might change if the assumptions about selection bias or the treatment effect are different.

We use $\hat{\delta}$ to distinguish the model’s estimate for the true underlying causal effect, δ . A well-identified regression that accurately models the source of student sorting into the less-restrictive environment would produce in expectation $\hat{\delta} = \delta$. The below exercises document the systematic difference between $\hat{\delta}$ and δ that occur if the regression does not accurately model the source of sorting into the treatment. We cannot know the magnitude of selection bias. Rather,

our goal here is to determine the sign for the bias under alternative miss-specifications for the true source of selectivity.

Let us foreshadow the results. In the presence of positive selection based on fixed attributes the bias resulting from estimating a lagged-outcome model is a function of the portion of assignment to a less-restrictive classroom that y_{it-1} does not explain, and the relationship between y_{it-1} and likelihood of being placed into a less-restrictive setting, both of which are positive, leading to an estimate that is biased upward. In contrast, in the presence of positive selection based on lagged outcomes the bias resulting from estimating a student fixed-effect model is the product of the difference between the true and first-differenced relationship between y_{it} and y_{it-1} , which must be negative, and the (positive) relationship between the student's lagged outcome and likelihood they are assigned to the less-restrictive classroom, leading to an estimate that is biased downward.

That is, in light of positive selectivity the direction for potential selection bias under the two approaches are in opposing directions. It follows that, under the above plausible conditions, the true causal effect must be found between the results produced by the lagged-outcome and student fixed-effect models.

A.1 Lagged Outcome Model Produces Upper-Bound for True Causal Effect

Let us first consider the case where assignment to a more inclusive setting in year t is positively correlated with an unobserved fixed effect, but we incorrectly estimate a lagged outcome model. The following equation accurately describes the outcome:

$$y_{it} = \alpha_i + \delta D_{it} + \varepsilon_{it} \tag{A1}$$

where ε_{it} is serially uncorrelated and uncorrelated with both α_i and D_{it} . Since α_i is fixed and by construction students are in the more restrictive classroom the prior year, we can also write:

$$y_{it-1} = \alpha_i + \varepsilon_{it-1} \tag{A2}$$

where α_i and ε_{it-1} are uncorrelated.

Under this scenario, a student fixed-effect regression would accurately model the source of selectivity and thus produce an unbiased estimate for δ . Instead, we mistakenly estimate a lagged outcome regression that models dynamic sorting into the less-restrictive setting. The lagged-outcome regression takes the form:

$$y_{it} = \lambda y_{it-1} + \delta D_{it} + \varepsilon_{it} \quad (\text{A3})$$

The probability limit for δ is $\frac{\text{Cov}(y_{it}, \tilde{D}_{it})}{V(\tilde{D}_{it})}$, where \tilde{D}_{it} is the residual from a regression describing the relationship between lagged outcome and probability the student is placed into the less-restrictive environment:

$$D_{it} = \gamma y_{it-1} + \tilde{D}_{it} \quad (\text{A4})$$

Note that γ represents the relationship between the student's lagged outcome and the likelihood that they are placed in the less-restrictive setting during year t . If lagged outcomes did not influence SDM placement then we would find $\gamma = 0$. However, by construction of this exercise where students are assigned to a less-restrictive setting based on having a higher lagged outcome, we know that $\gamma > 0$.

We can now substitute $\alpha_i = y_{it-1} - \varepsilon_{it-1}$ from A2 into A1 to get:

$$y_{it} = y_{it-1} + \delta D_{it} + \varepsilon_{it} - \varepsilon_{it-1} \quad (\text{A5})$$

And from here, with some algebra we reveal:

$$\hat{\delta} = \frac{\text{Cov}(y_{it}, \tilde{D}_{it})}{V(\tilde{D}_{it})} \quad (\text{A6})$$

$$\hat{\delta} = \delta - \frac{\text{Cov}(\varepsilon_{it-1}, \tilde{D}_{it})}{V(\tilde{D}_{it})}$$

The intuition underlying the equation derived thus far is that the regression's estimate for the causal effect of treatment contains both the true causal effect and selection bias associated with

the portion of assignment to a less-restrictive classroom that y_{it-1} does not explain. We can further derive this equation to:

$$\hat{\delta} = \delta + \frac{\gamma\sigma_{\varepsilon}^2}{V(\tilde{D}_{it})} \quad (\text{A7})$$

where σ_{ε}^2 is the variance of ε_{it-1} . This equation shows that the selection bias embedded within $\hat{\delta}$ is a specific function of the variances for ε_{it-1} and the portion of assignment to a less-restrictive classroom that y_{it-1} does not explain, and the relationship between y_{it-1} and likelihood of being placed into a less-restrictive setting.

We can clearly sign the direction of bias from this equation. As variances, both σ_{ε}^2 and $V(\tilde{D}_{it})$ must be positive. By construction of this exercise, we know that $\gamma > 0$. Thus, the selection bias term must be positive. Consequently, if students are positively sorted based on fixed characteristics, the results from a lagged outcome model will *overstate* and thus serve as an upper-bound for the the true causal effect of assignment to the less-restrictive SDM.

A.2 Student Fixed-Effect Model Produces Lower-Bound for True Causal Effect

Now, let's consider the case where assignment to a more inclusive setting in year t is positively correlated with the student's lagged outcome, but we incorrectly estimate a student fixed-effect model. The following equation accurately describes the outcome:

$$y_{it} = \alpha + \theta y_{it-1} + \delta D_{it} + \varepsilon_{it} \quad (\text{A8})$$

where ε_{it} is serially uncorrelated and uncorrelated with D_{it} .

Under this scenario, a lagged-outcome regression would accurately model the source of selectivity and thus produce an unbiased estimate for δ . Instead, we mistakenly estimate a student fixed-effect regression that models sorting into the less-restrictive setting based on fixed student attributes. The student fixed-effect regression takes the form:

$$y_{it} = \alpha_i + \delta D_{it} + \varepsilon_{it} \quad (\text{A9})$$

Though it can be shown that the result generalizes, for this illustration we consider the most simple form where $D_{it-1} = 0$ for everyone, which is equivalent to the first-differenced estimator with the probability limit:

$$\frac{Cov(y_{it} - y_{it-1}, D_{it} - D_{it-1})}{V(D_{it} - D_{it-1})} = \frac{Cov(y_{it} - y_{it-1}, D_{it})}{V(D_{it})} \quad (\text{A10})$$

Subtracting y_{it-1} from both sides of Equation A8 yields:

$$y_{it} - y_{it-1} = \alpha_i + (\theta - 1)y_{it-1} + \delta D_{it} + \varepsilon_{it} \quad (\text{A11})$$

Substituting Equation A11 into Equation A10 leaves:

$$\hat{\delta} = \frac{Cov(y_{it} - y_{it-1}, D_{it})}{V(D_{it})} \quad (\text{A12})$$

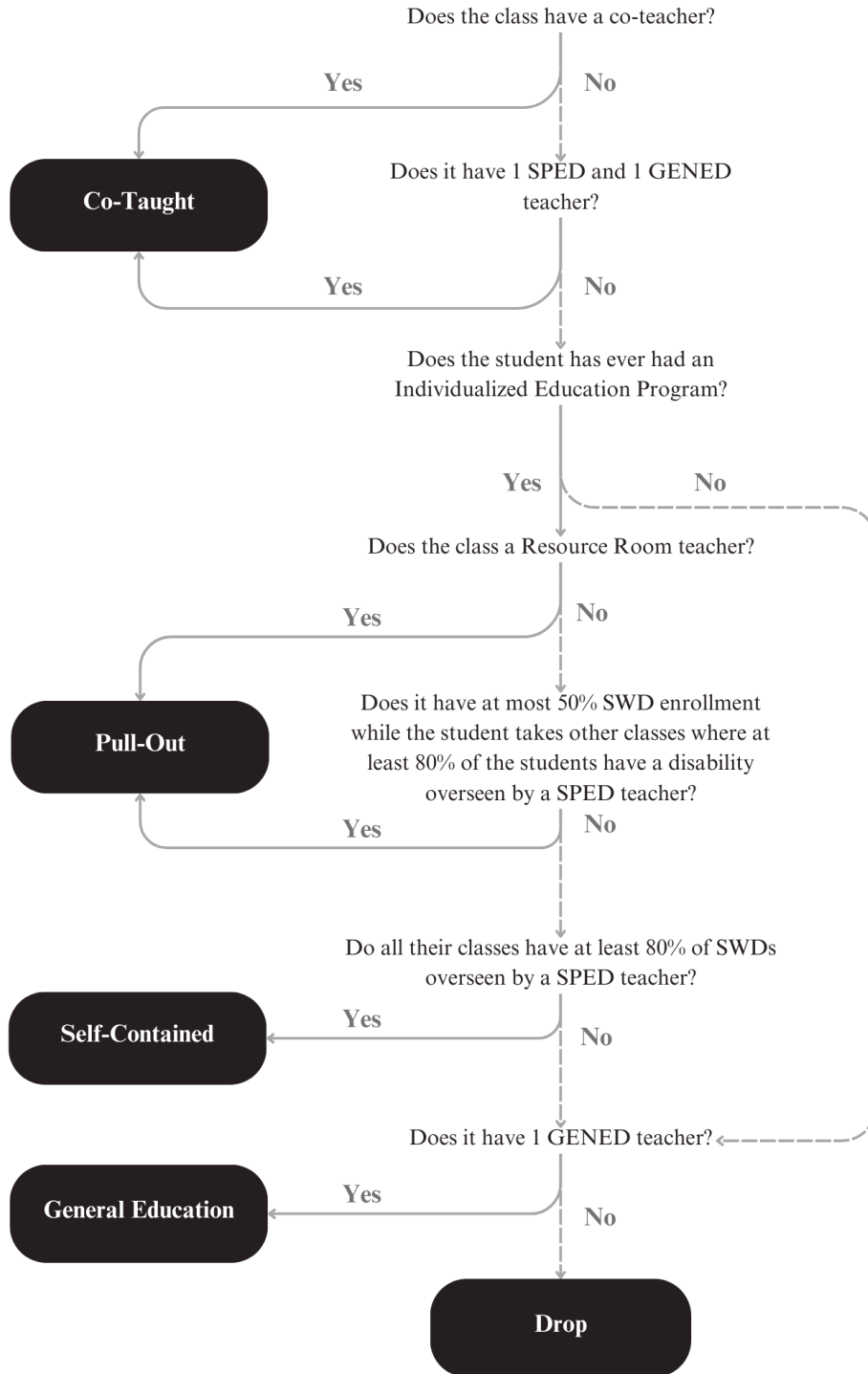
$$\hat{\delta} = \delta + (\theta - 1) \frac{Cov(y_{it-1}, D_{it})}{V(D_{it})}$$

Equation A12 says that the estimate for the causal effect contains both the true causal effect and selection bias associated with the unaccounted-for relationship between the student's lagged outcome and the likelihood they are assigned to a less-restrictive classroom. Intuitively, in this case selection bias is the product of the difference between the true and first-differenced relationship between y_{it} and y_{it-1} and the relationship between the student's lagged outcome and likelihood they are assigned to the less-restrictive classroom in year t .

We can easily sign the direction of bias described in the above equation. In general, we know that the relationship between y_{it} and y_{it-1} is a positive number less than one, so $(\theta - 1)$ is negative. By construction of this exercise where students are positively sorted into less-restrictive classrooms based on their lagged outcome, we know that the second term is positive. Thus, the direction of bias is negative. That is, in the presence of positive sorting based on lagged outcome, the student fixed-effect model will *understate* and thus serve as a lower-bound for the true causal effect of assignment to the less-restrictive SDM.

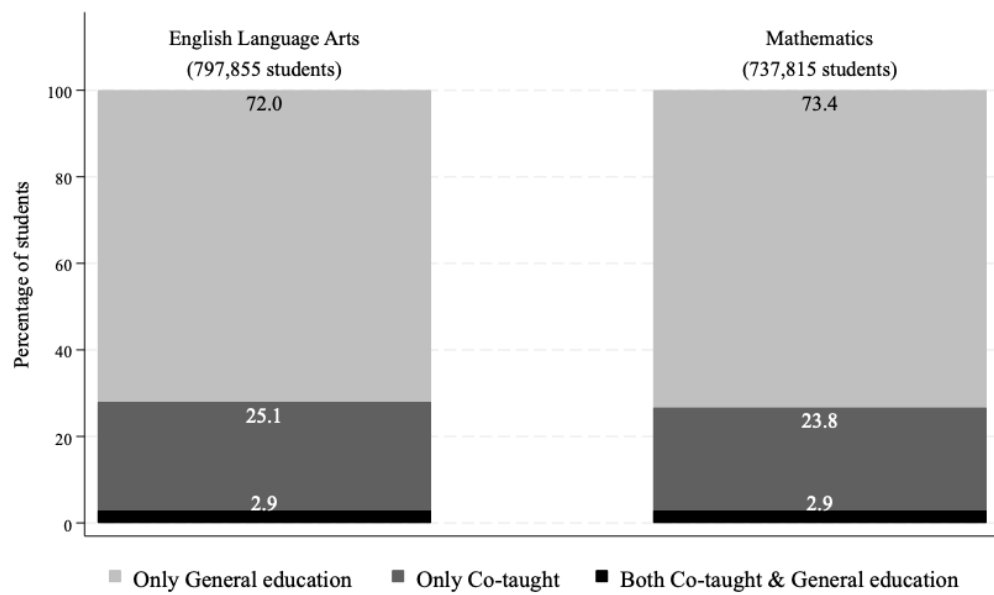
B Figures

Figure B1: Service Delivery Model Assignment Mechanism



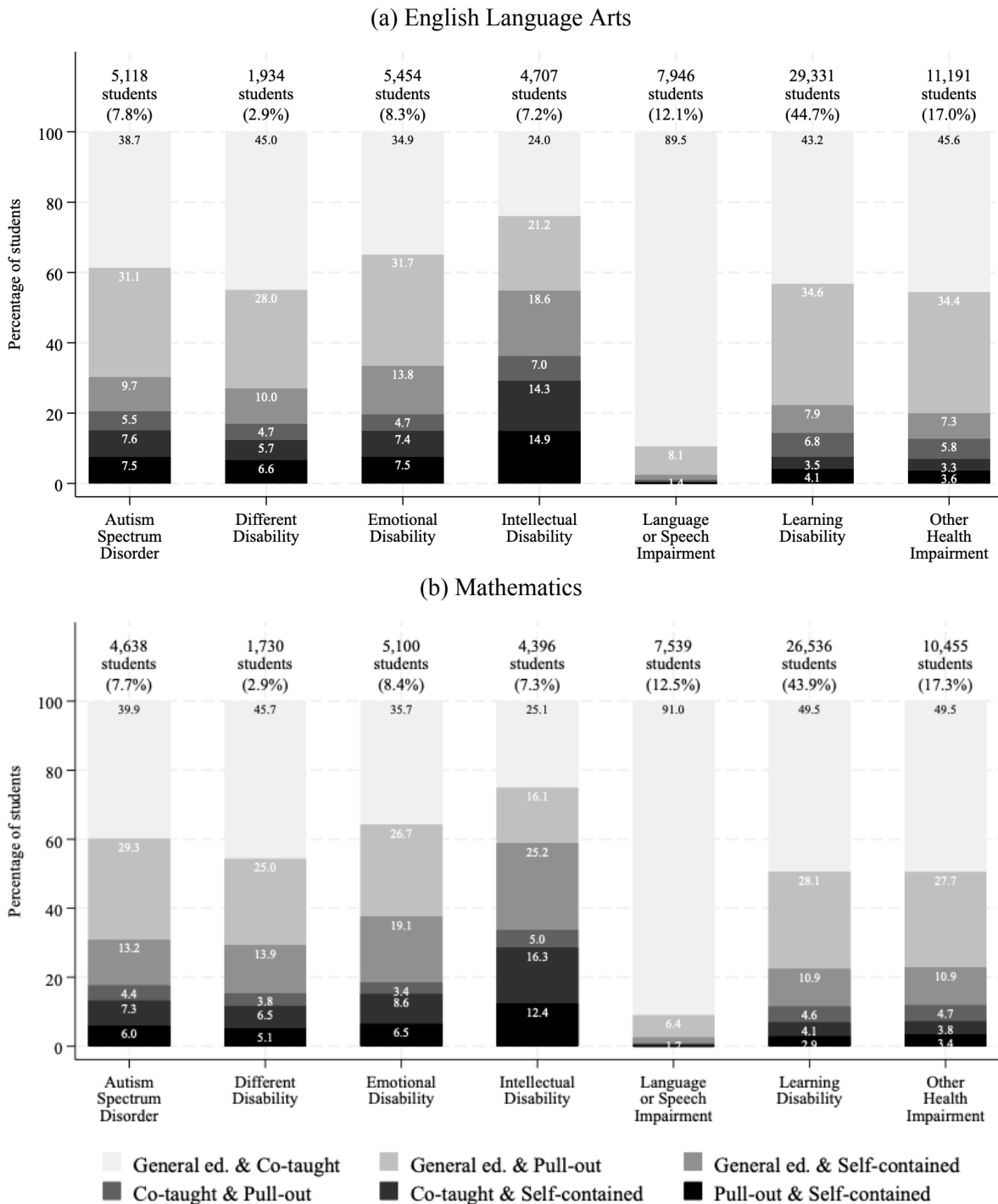
Notes: This figure shows the multi-stage procedure that assigned each student-subject combination to one of the four service delivery models (SDMs): regular general education, co-taught, pull-out, and self-contained. Given its conditional nature, if a student could belong to more than one SDM, they were assigned to the first one specified in the sequential assessment. Students without disabilities can only participate in regular general education or co-taught models.

Figure B2: Share of Students Without Disabilities by Number of Service Delivery Models



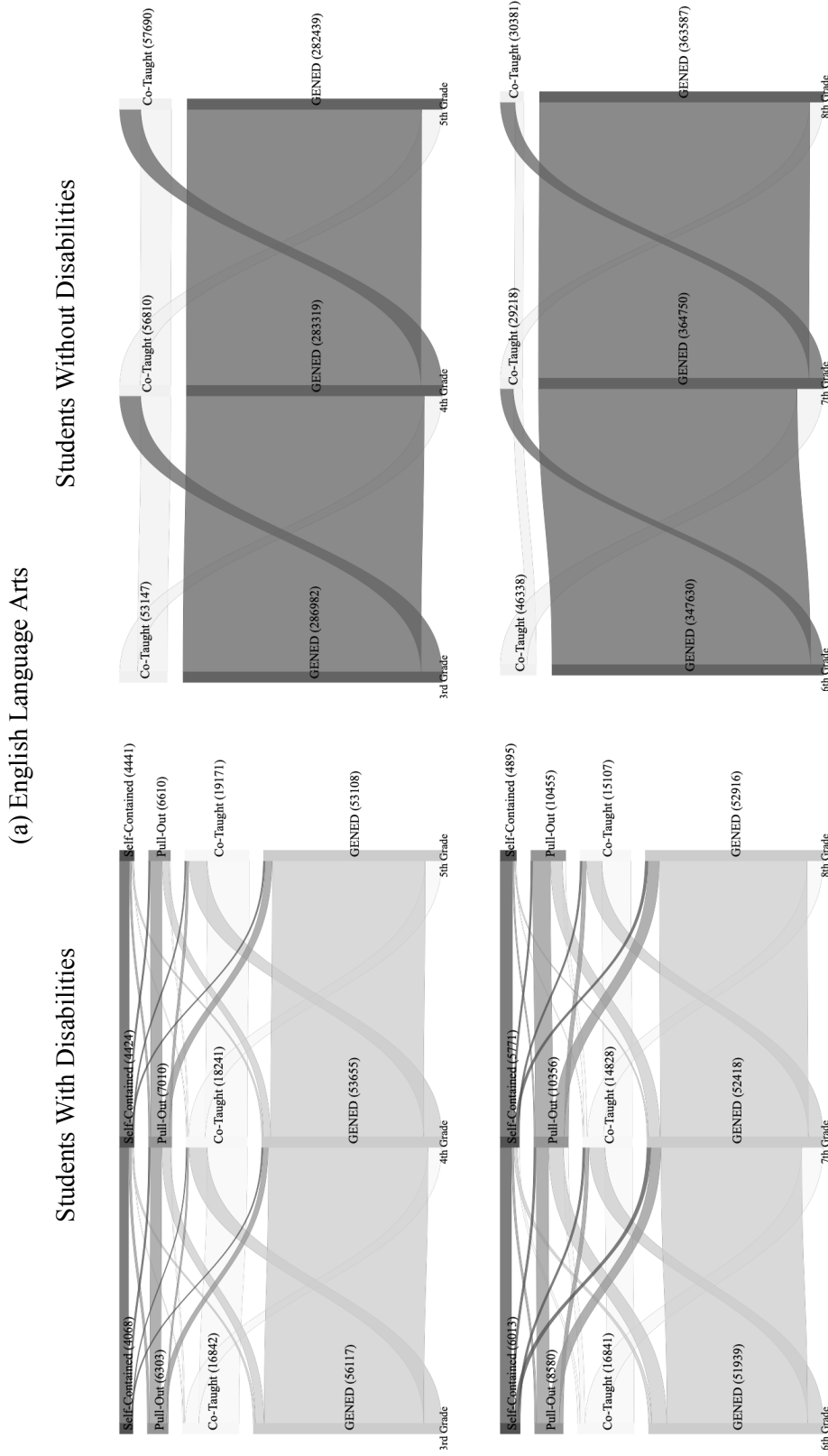
Notes: This figure shows the proportion of students without disabilities in the sample enrolled in ELA and Mathematics, distributed by the number of SDMs attended over the sample period. Sample sizes are indicated above each bar.

Figure B3: Service Delivery Model by Disability Type



Notes: This figure shows the percentage of students with disabilities, within each Special Education classification, that experience each pair of service delivery models (SDMs) in ELA (Panel A) and Mathematics (Panel B). The sample is restricted to students enrolled in grades three through eight that we observe in both SDMs for each subject area at least once. Sample sizes and proportions are indicated above each bar. The number represents the individual count of students included in each category, while the percentage reflects the proportion of students included in each category, out of the full sample (by subject area). The category "Different Disability" encompasses students with Blind, Brain Injury, Deaf, Deaf-Blind, Developmental Delay, Multiple Disabilities, and Orthopedic Impairment.

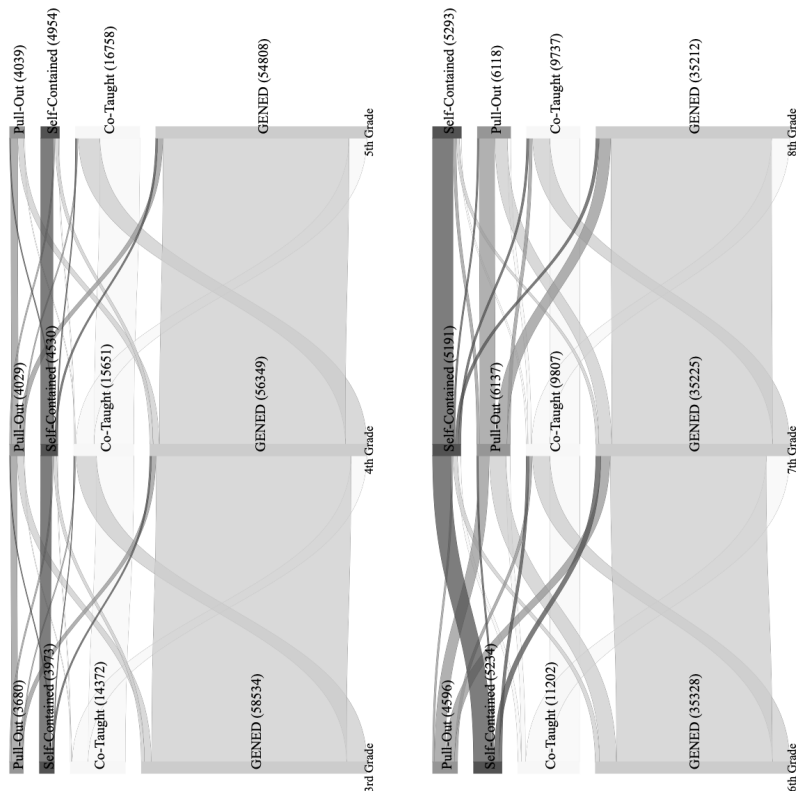
Figure B4: Students' Fluctuations across Service Delivery Models



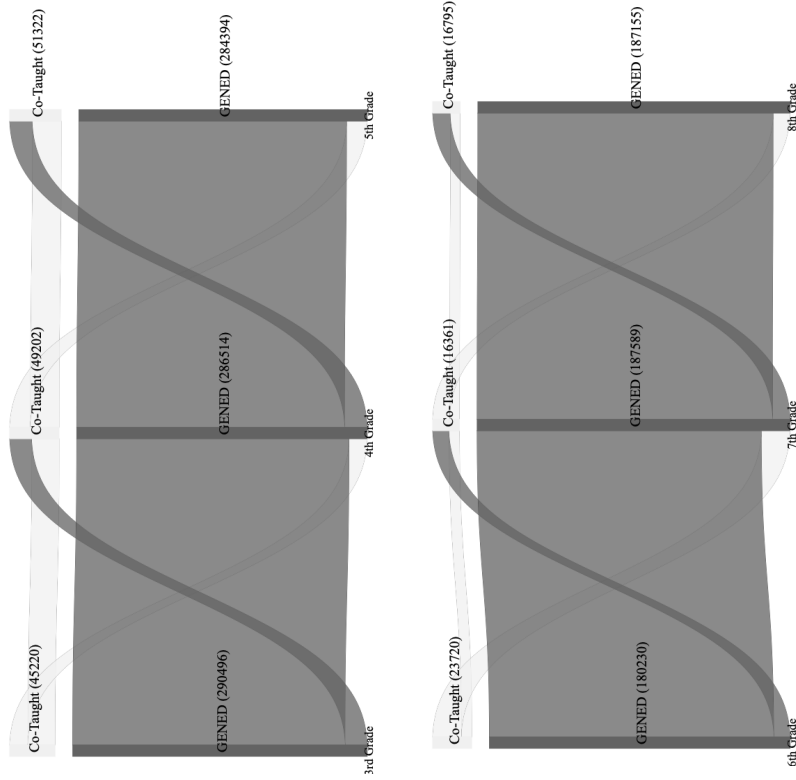
Notes: This figure shows the mobility of students with and without disabilities across service delivery models in English Language Arts (Panel A) and Mathematics (Panel B). In both panels, only students observed in grades 3 through 8 are included in the calculations. As such, in both panels, the top left quadrant shows the movement of students with disabilities across SDMs in grades 3–5, while the bottom right quadrant shows the same for students without disabilities. The bottom left quadrant shows the movement of students with disabilities across SDMs in grades 6–8, and the top right quadrant shows the same for students without disabilities. Numbers in parentheses represent sample sizes. The figure was created using the Sankey package for Stata, developed by Naqvi (2024).

(b) Mathematics

Students With Disabilities



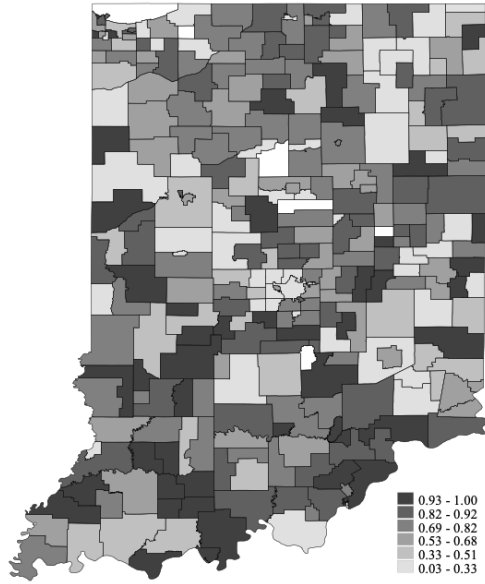
Students Without Disabilities



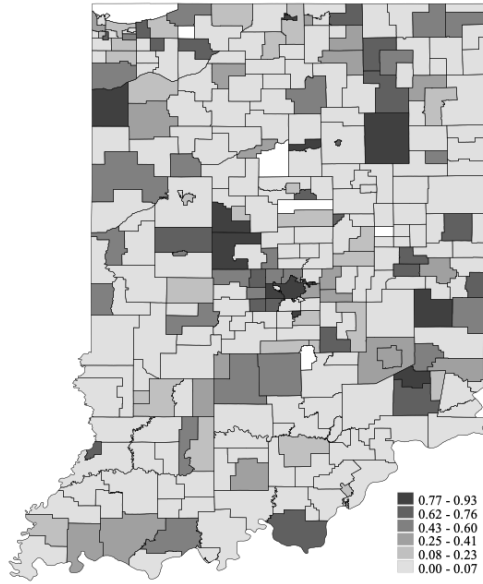
Notes: This figure shows the mobility of students with and without disabilities across service delivery models in English Language Arts (Panel A) and Mathematics (Panel B). In both panels, only students observed in grades 3 through 8 are included in the calculations. As such, in both panels, the top left quadrant shows the movement of students with disabilities across SDMs in grades 3–5, while the top right quadrant shows the same for students without disabilities. The bottom left quadrant shows the movement of students with disabilities across SDMs in grades 6–8, and the bottom right quadrant shows the same for students without disabilities. Numbers in parentheses represent sample sizes. The figure was created using the Sankey package for Stata, developed by Naqvi (2024).

Figure B5: Share of SWDs by Service Delivery Model

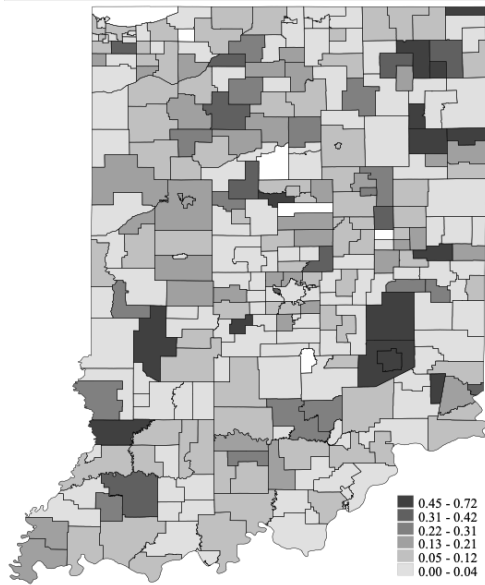
(a) English Language Arts



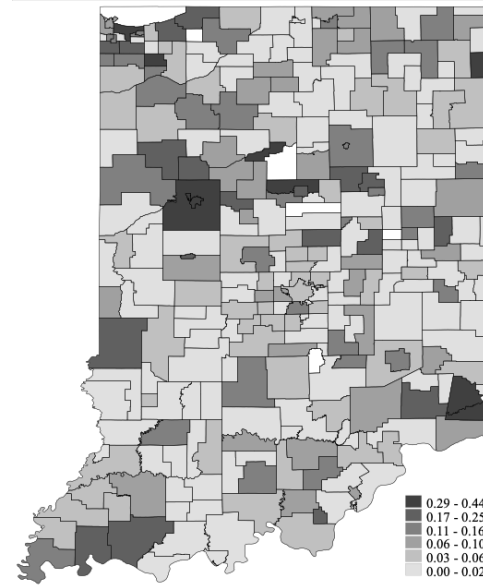
(i) General education



(ii) Co-taught



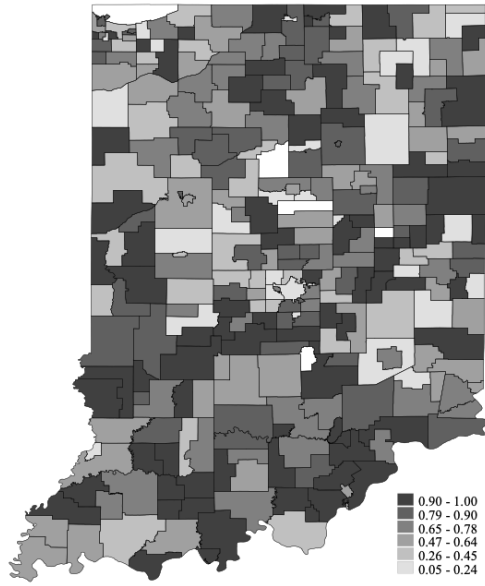
(iii) Pull-out



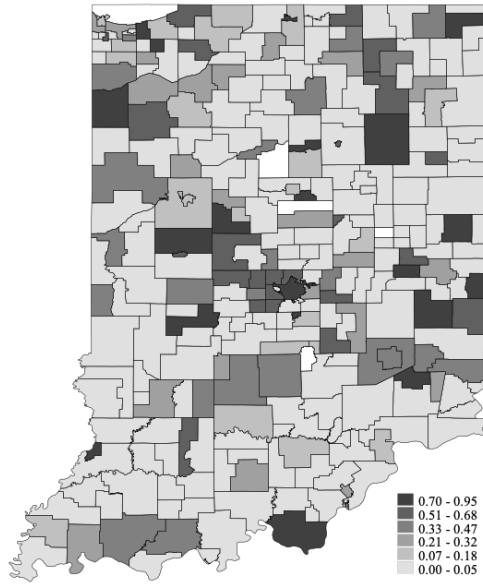
(iv) Self-contained

Notes: This figure plots the percentage of students with disabilities observed in each service delivery model for ELA in grades 3 through 8 in 2021 by school district. We categorize a student as having a disability if, at any given time in our dataset, they have an IEP.

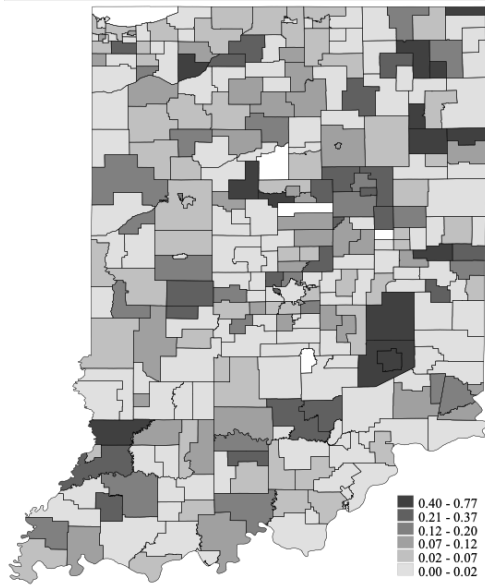
(b) Mathematics



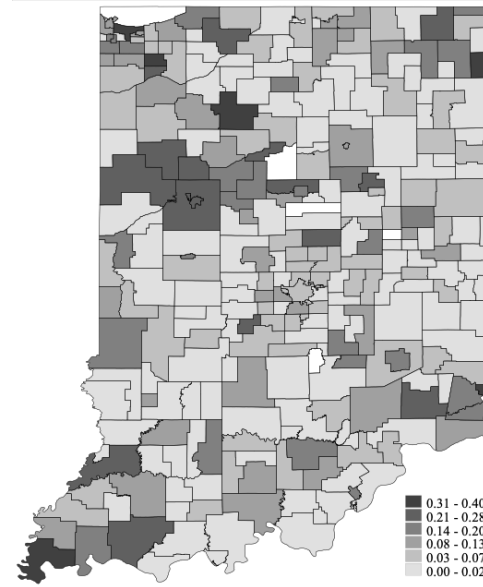
(i) General education



(ii) Co-taught



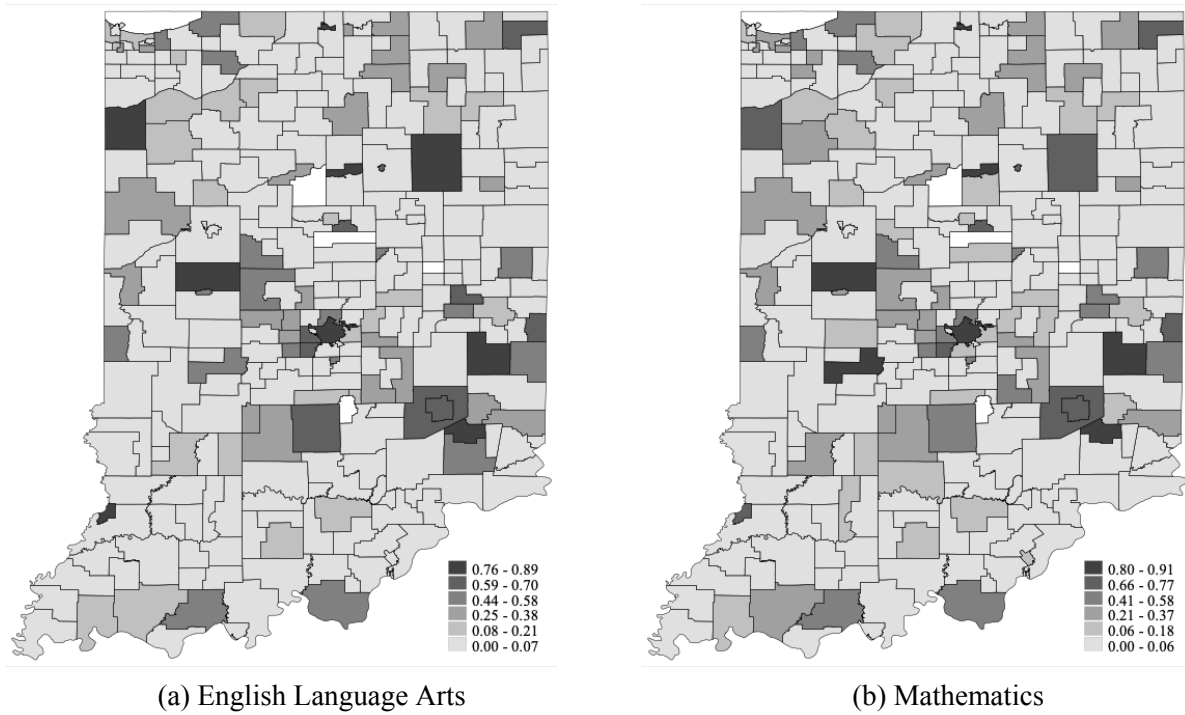
(iii) Pull-out



(iv) Self-contained

Notes: This figure plots the percentage of students with disabilities observed in each service delivery model for Mathematics in grades 3 through 8 in 2021 by school district.

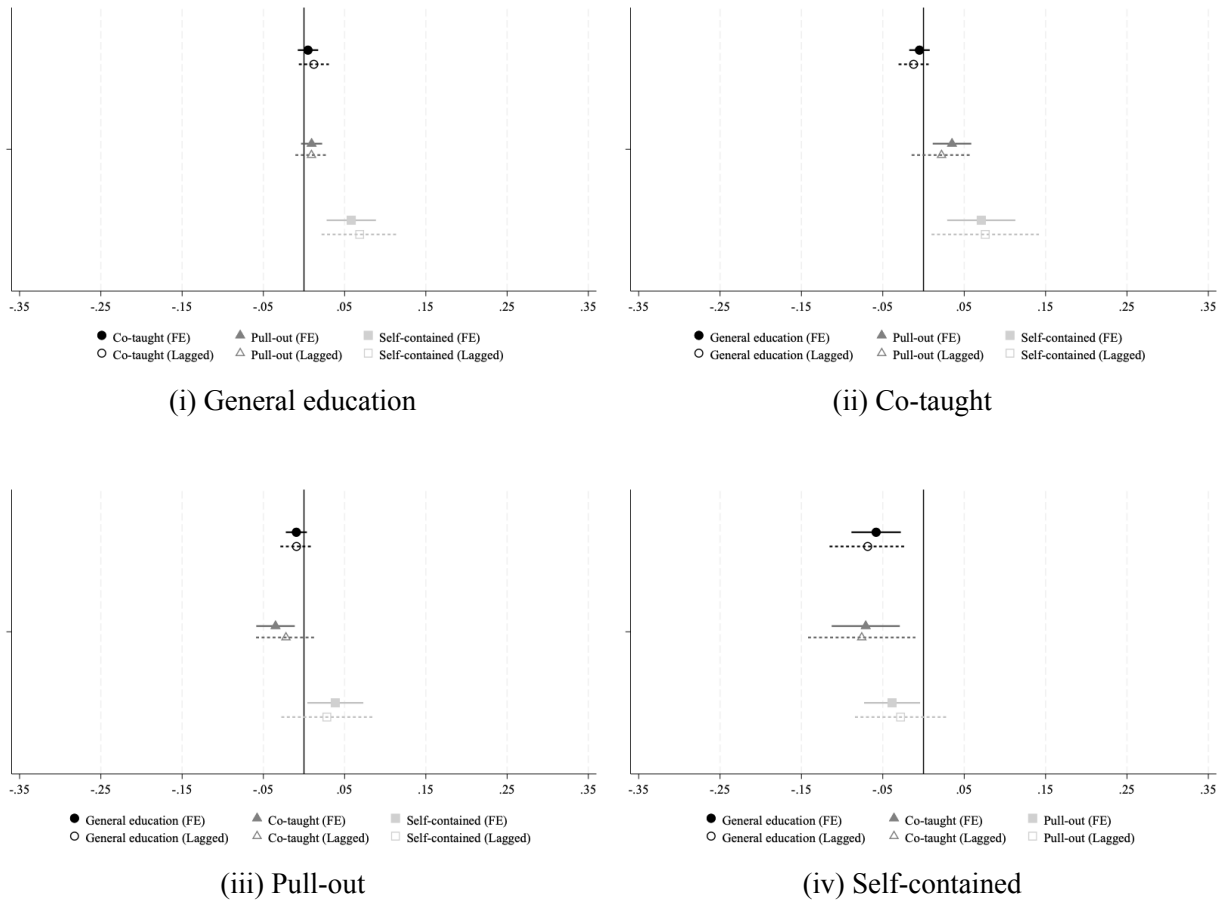
Figure B6: Share of Students Without Disabilities in Co-taught Classrooms



Notes: This figure plots the percentage of students without disabilities observed in co-taught classrooms for English Language Arts and Mathematics in grades 3 through 8 in 2021 school by district.

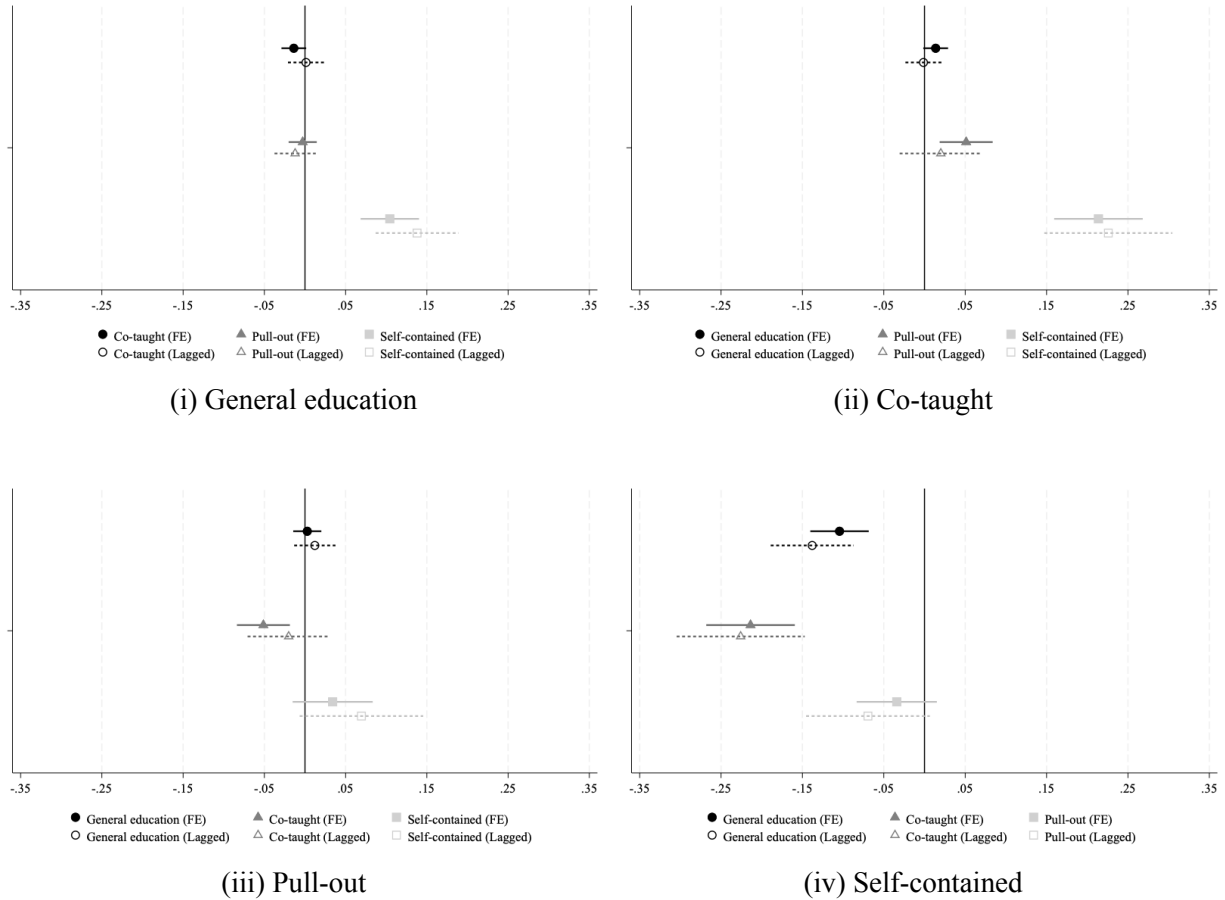
Figure B7: Service Delivery Models Effects on Students with Disabilities

(a) English Language Arts



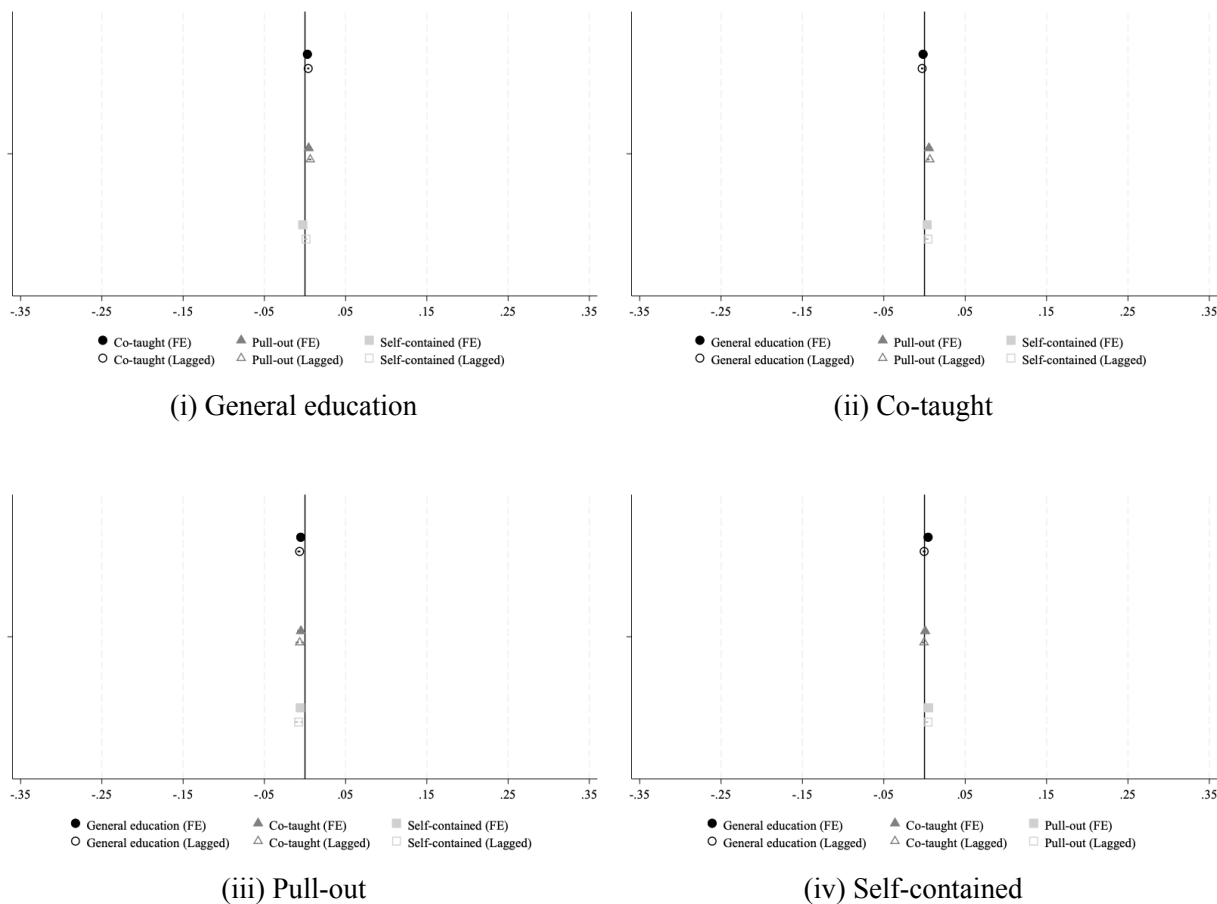
Notes: This figure illustrates the effect of attending a specific service delivery model (SDM) on ELA test scores, compared to alternative settings. For example, the top left quadrant presents the effect of attending a general education (GENED) classroom relative to co-taught, pull-out, and self-contained settings. The sample is limited to students with disabilities in grades three through eight. Each comparison is represented by two symbols: filled symbols denote results from the school and student fixed effects model, while hollow symbols indicate results from the lagged dependent variable model with school fixed effects. For instance, in quadrant (i), the black circle shows the effect of attending a GENED classroom compared to a co-taught classroom using the school and student fixed effects model, while the hollow circle shows the effect based on the previous year's ELA test scores and school fixed effects. Horizontal bars represent 95% confidence intervals.

(b) Mathematics



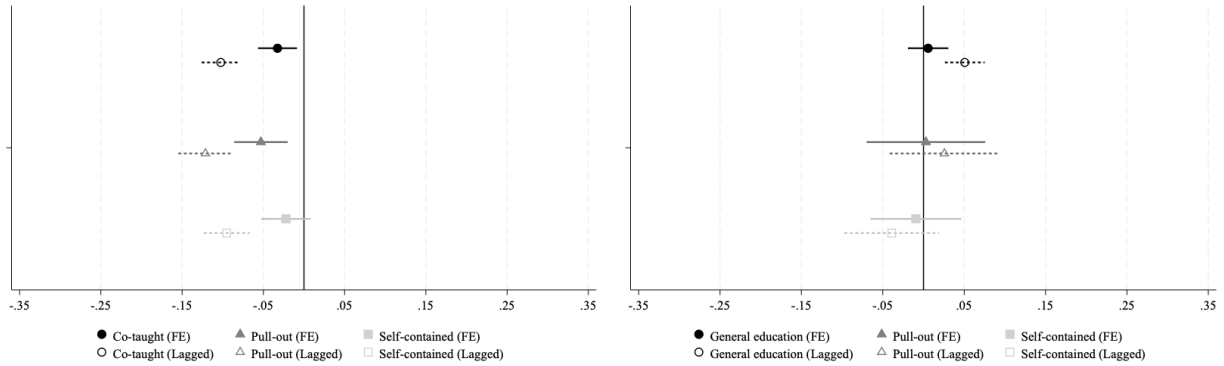
Notes: This figure illustrates the effect of attending a specific service delivery model (SDM) on Math test scores, compared to alternative settings. For example, the top left quadrant presents the effect of attending a general education (GENED) classroom relative to co-taught, pull-out, and self-contained settings. The sample is limited to students with disabilities in grades three through eight. Each comparison is represented by two symbols: filled symbols denote results from the school and student fixed effects model, while hollow symbols indicate results from the lagged dependent variable model with school fixed effects. For instance, in quadrant (i), the black circle shows the effect of attending a GENED classroom compared to a co-taught classroom using the school and student fixed effects model, while the hollow circle shows the effect based on the previous year's Math test scores and school fixed effects. Horizontal bars represent 95% confidence intervals.

(c) Attendance Rate



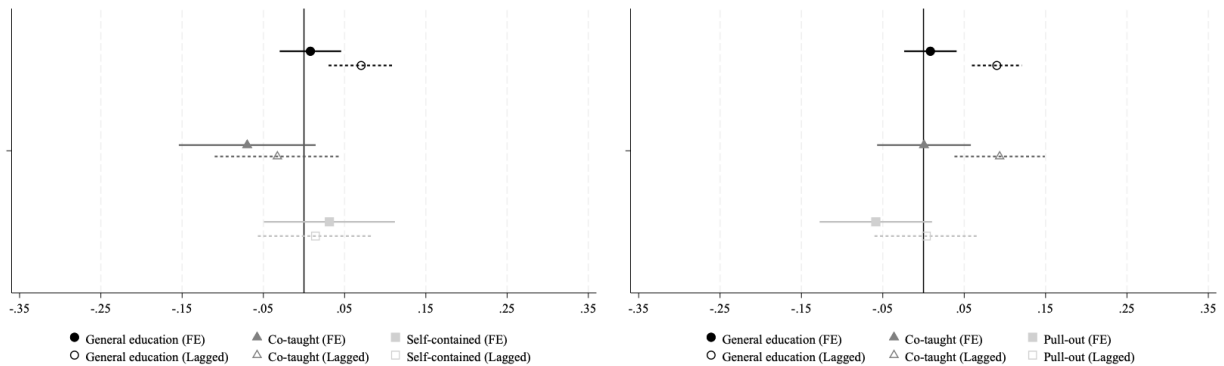
Notes: This figure illustrates the effect of attending a specific service delivery model (SDM) on students with disabilities' attendance rate compared to alternative settings. For example, the top left quadrant presents the effect of receiving instruction for all subjects (ELA, Mathematics, Social Studies, and Science) in a general education (GENED) classroom relative to co-taught, pull-out, and self-contained settings. The sample is limited to students with disabilities in grades one through twelve. Each comparison is represented by two symbols: filled symbols denote results from the student, school, and grade fixed effects model, while hollow symbols indicate results from the lagged dependent variable model with school and grade fixed effects. For instance, in quadrant (i), the black circle shows the effect of attending all subjects in a GENED classroom compared to a co-taught setting using the student, school, and grade fixed effects, while the hollow circle shows the effect based on the previous year's attendance rate, as well as school and grade fixed effects. Horizontal bars represent 95% confidence intervals.

(d) Disciplinary Incidents



(i) General education

(ii) Co-taught

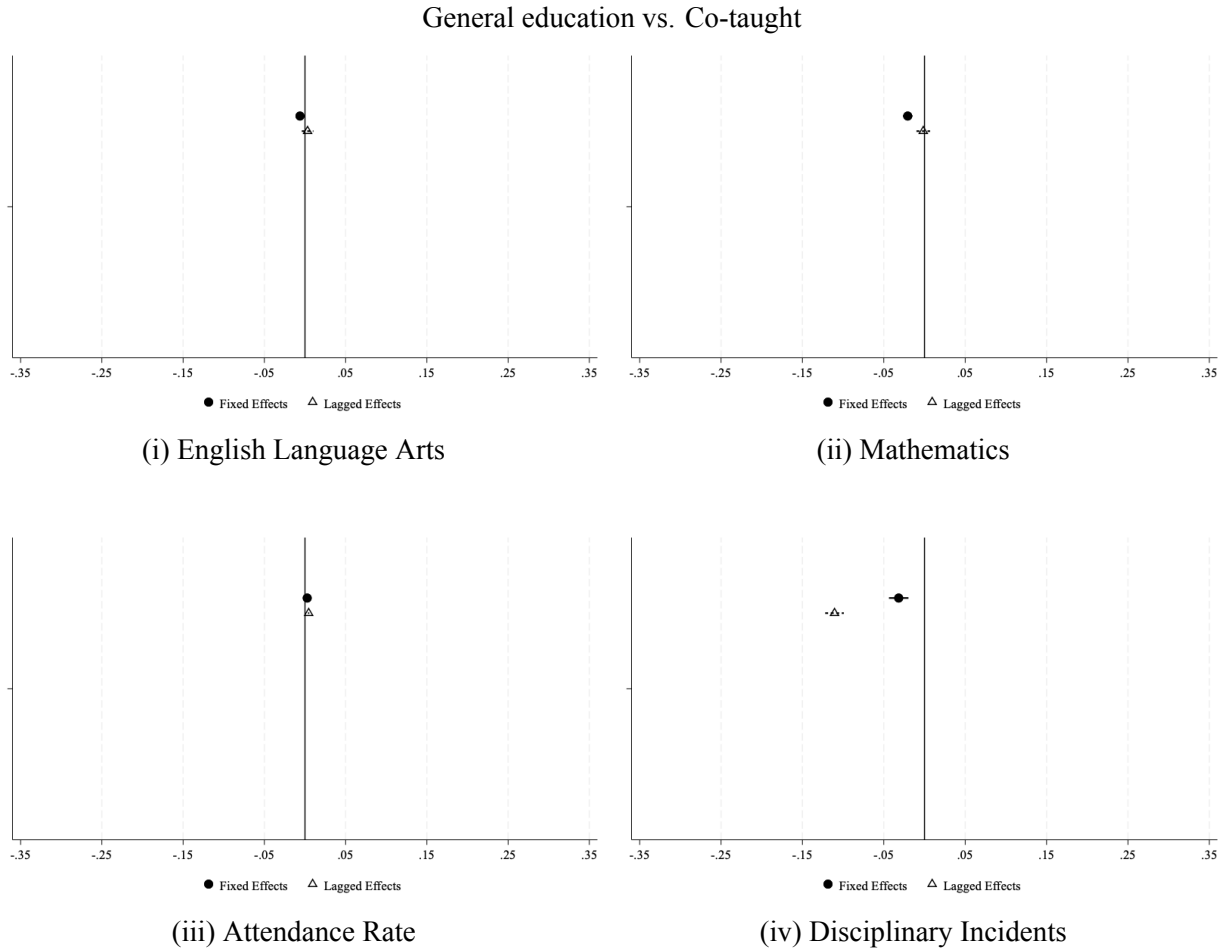


(iii) Pull-out

(iv) Self-contained

Notes: This figure illustrates the effect of attending a specific service delivery model (SDM) on students with disabilities' number of disciplinary incidents compared to alternative settings. For example, the top left quadrant presents the effect of receiving instruction for all subjects (ELA, Mathematics, Social Studies, and Science) in a general education (GENED) classroom relative to co-taught, pull-out, and self-contained settings. The sample is limited to students with disabilities in grades one through twelve. Each comparison is represented by two symbols: filled symbols denote results from the student, school, and grade fixed effects model, while hollow symbols indicate results from the lagged dependent variable model with school and grade fixed effects. For instance, in quadrant (i), the black circle shows the effect of attending all subjects in a GENED classroom compared to a co-taught setting using the student, school, and grade fixed effects, while the hollow circle shows the effect based on the previous year's disciplinary incidents, as well as school and grade fixed effects. Horizontal bars represent 95% confidence intervals.

Figure B8: Service Delivery Models Effects on Students without Disabilities



Notes: This figure illustrates the effect of attending a general education (GENED) classroom versus a co-taught setting on standardized ELA and Math scores, attendance rates, and disciplinary incidents. For non-test outcomes, GENED refers to receiving instruction in all subjects (ELA, Mathematics, Social Studies, and Science) within a general education classroom, as opposed to a co-taught setting. The sample includes students without disabilities in grades 3 through 8 for test outcomes (top quadrants) and grades 1 through 12 for non-test outcomes (bottom quadrants). Each comparison is depicted using two symbols: black circles represent results from the student and school fixed effects model, while hollow triangles denote results from the lagged dependent variable model with school fixed effects. In the case of non-test outcomes, all models also incorporate grade fixed effects. For example, in quadrant (iii), the black circle indicates the effect of receiving instruction in all subjects within a GENED classroom compared to a co-taught setting on attendance rates, accounting for student, school, and grade fixed effects. Meanwhile, the hollow circle represents the effect based on the previous year's attendance rate, as well as school and grade fixed effects. Horizontal bars indicate 95% confidence intervals.

C Tables

Table C1: Regression on the Observables. Students With Disabilities

<i>General education vs. Co-taught</i>												
	English Language Arts			Mathematics			Attendance Rate			Disciplinary Incidents		
<i>Lagged Effects</i>												
Outcome of interest	0.039*** (0.002)	0.025*** (0.003)	0.021*** (0.007)	0.025*** (0.002)	0.013*** (0.003)	-0.013 (0.008)	0.150*** (0.008)	0.108*** (0.008)	0.058*** (0.011)	-0.006*** (0.000)	-0.005*** (0.000)	-0.003*** (0.000)
Co-taught	0.335*** (0.004)	0.327*** (0.004)	0.345*** (0.005)	0.351*** (0.004)	0.345*** (0.004)	0.367*** (0.005)	-0.117*** (0.002)	-0.104*** (0.002)	0.061*** (0.002)	-0.117*** (0.002)	-0.104*** (0.002)	0.061*** (0.002)
Pull-out	0.021 (0.016)	0.031* (0.016)	0.056* (0.031)	0.031 (0.021)	0.043** (0.021)	0.100*** (0.038)	-0.267*** (0.005)	-0.224*** (0.005)	-0.021*** (0.006)	-0.266*** (0.005)	-0.224*** (0.005)	-0.021*** (0.006)
Self-contained	0.034 (0.029)	0.038 (0.029)	0.086 (0.053)	0.050* (0.028)	0.061** (0.028)	0.106** (0.051)	-0.386*** (0.003)	-0.339*** (0.003)	-0.095*** (0.004)	-0.385*** (0.003)	-0.339*** (0.003)	-0.095*** (0.004)
<i>Characteristics</i>												
Elementary		0.051*** (0.009)	0.054*** (0.011)		0.091*** (0.009)	0.104*** (0.011)						
English learner		0.014 (0.011)	0.108* (0.059)		-0.004 (0.012)	-0.025 (0.061)		-0.016*** (0.002)	-0.015*** (0.006)		-0.015*** (0.002)	-0.015** (0.006)
Free & reduced-price meal		-0.007 (0.004)	0.006 (0.016)		-0.004 (0.005)	-0.012 (0.017)		-0.011*** (0.001)	-0.001 (0.001)		-0.012*** (0.001)	-0.001 (0.001)
Female		-0.000 (0.004)			-0.003 (0.004)			0.002*** (0.001)			0.001 (0.001)	
<i>Race</i>												
Asian		0.018 (0.021)			0.024 (0.025)			0.017*** (0.005)			0.017*** (0.005)	
Black		0.003 (0.008)			-0.011 (0.008)			-0.012*** (0.001)			-0.010*** (0.001)	
Hispanic/latinx		0.001 (0.008)			0.009 (0.009)			0.004** (0.002)			0.004** (0.002)	
Other race		0.006 (0.009)			-0.004 (0.009)			-0.002 (0.002)			-0.001 (0.002)	
<i>SPED Classification</i>												
Autism spectrum disorder		-0.020* (0.012)			-0.034*** (0.012)			-0.048*** (0.002)			-0.048*** (0.002)	
Different disability		-0.006 (0.015)			-0.018 (0.016)			-0.030*** (0.003)			-0.031*** (0.003)	
Emotional disability		-0.009 (0.012)			-0.016 (0.012)			-0.034*** (0.002)			-0.032*** (0.002)	
Intellectual disability		-0.041*** (0.014)			-0.058*** (0.014)			-0.084*** (0.002)			-0.084*** (0.002)	
Language/speech impairment		0.015 (0.009)			0.006 (0.009)			0.030*** (0.001)			0.030*** (0.001)	
Learning disability		-0.047*** (0.009)			-0.051*** (0.010)			-0.053*** (0.001)			-0.053*** (0.001)	
Other health impairment		-0.036*** (0.010)			-0.048*** (0.010)			-0.043*** (0.001)			-0.043*** (0.001)	
<i>Sample</i>												
Observations	51,300	47,125	39,213	45,382	42,204	33,696	541,400	540,213	529,675	541,825	540,638	530,113
R ²	0.208	0.207	0.405	0.219	0.220	0.409	0.341	0.354	0.520	0.341	0.354	0.520
Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Student FE	No	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grade FE	No	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: This table presents the estimates of regressing general education (GENED) on observables. The sample is restricted to students with disabilities enrolled in 3–8 eight for ELA and Math, and 1–12 for attendance rate and disciplinary incidents dataset. Students included in the regression attended at least one GENED and one co-taught classroom during the sample period. The dependent variable is always GENED. We used three approximations: (1) lagged measures and school fixed effects, (2) lagged measures and school fixed effects with controls, and (3) lagged measures, as well as student and school fixed effects with controls. For non-test outcomes, we also add grade fixed effects for all the models. Standard errors are clustered at the student level and reported in parenthesis. “Lagged Effects” includes the values for each variable in the previous year. Within Lagged Effects, each SDM takes a value of 1 if a student received that instruction type for that subject area in the previous year. GENED is the reference category. “Elementary” takes a value of 1 if a student is below grade six. We designate a student as having a disability if they have an IEP at any point in our dataset. The reference category for students’ race is White. “Other Race” includes students identified as American Indian, Multiracial, Native Hawaiian or Other Pacific Islander. A student is classified as having a specific disability type if, at any given moment in our dataset, they are identified as having that particular disability. The category “Different Disability” encompasses students with Blind, Brain Injury, Deaf, Deaf-Blind, Developmental Delay, Multiple Disabilities, and Orthopedic Impairment.

General education vs. Pull-out

	English Language Arts			Mathematics			Attendance Rate			Disciplinary Incidents		
<i>Lagged Effects</i>												
Outcome of interest	0.042*** (0.003)	0.039*** (0.003)	0.027*** (0.010)	0.028*** (0.004)	0.023*** (0.004)	-0.013 (0.013)	0.159*** (0.012)	0.153*** (0.012)	0.100*** (0.018)	-0.008*** (0.001)	-0.008*** (0.001)	-0.006*** (0.001)
Co-taught	0.225*** (0.020)	0.227*** (0.020)	0.133*** (0.041)	0.242*** (0.024)	0.248*** (0.025)	0.178*** (0.049)	-0.113*** (0.004)	-0.110*** (0.004)	0.023*** (0.004)	-0.112*** (0.004)	-0.109*** (0.004)	0.023*** (0.004)
Pull-out	0.341*** (0.005)	0.342*** (0.005)	0.311*** (0.007)	0.382*** (0.006)	0.385*** (0.006)	0.342*** (0.009)	-0.070*** (0.003)	-0.062*** (0.003)	0.091*** (0.004)	-0.069*** (0.003)	-0.062*** (0.003)	0.091*** (0.004)
Self-contained	-0.021 (0.023)	-0.021 (0.023)	-0.208*** (0.038)	0.014 (0.029)	0.019 (0.029)	-0.136*** (0.049)	-0.321*** (0.004)	-0.303*** (0.004)	-0.095*** (0.005)	-0.320*** (0.004)	-0.302*** (0.004)	-0.095*** (0.005)
<i>Characteristics</i>												
Elementary		0.016 (0.013)	0.036** (0.017)		0.096*** (0.016)	0.125*** (0.020)						
English learner		-0.006 (0.016)	0.051 (0.122)		-0.000 (0.019)	0.139 (0.139)		-0.015*** (0.003)	-0.036*** (0.012)		-0.014*** (0.003)	-0.035*** (0.012)
Free & reduced-price meal		0.001 (0.006)	-0.004 (0.024)		-0.007 (0.007)	-0.020 (0.029)		-0.010*** (0.001)	-0.004* (0.002)		-0.011*** (0.001)	-0.004* (0.002)
Female			-0.002 (0.005)		-0.014** (0.006)			0.004*** (0.001)			0.002* (0.001)	
<i>Race</i>												
Asian		0.035 (0.033)			0.068** (0.030)			0.001 (0.007)			0.002 (0.007)	
Black		-0.006 (0.011)			-0.004 (0.013)			-0.011*** (0.002)			-0.007*** (0.002)	
Hispanic/latinx		-0.006 (0.012)			-0.003 (0.014)			0.004* (0.002)			0.005** (0.002)	
Other race		0.014 (0.012)			-0.006 (0.015)			0.001 (0.003)			0.002 (0.003)	
<i>SPED Classification</i>												
Autism spectrum disorder		-0.055*** (0.016)			-0.082*** (0.018)			-0.022*** (0.003)			-0.023*** (0.003)	
Different disability		-0.034* (0.021)			-0.079*** (0.024)			-0.019*** (0.004)			-0.021*** (0.004)	
Emotional disability		-0.050*** (0.015)			-0.071*** (0.018)			-0.021*** (0.002)			-0.017*** (0.002)	
Intellectual disability		-0.054*** (0.016)			-0.107*** (0.020)			-0.068*** (0.003)			-0.068*** (0.003)	
Language/speech impairment		0.044*** (0.013)			0.041** (0.016)			0.015*** (0.002)			0.015*** (0.002)	
Learning disability		-0.061*** (0.014)			-0.086*** (0.016)			-0.006*** (0.002)			-0.006*** (0.002)	
Other health impairment		-0.063*** (0.014)			-0.080*** (0.016)			-0.010*** (0.002)			-0.009*** (0.002)	
<i>Sample</i>												
Observations	29,084	29,084	19,964	20,151	20,151	13,280	192,928	192,928	188,142	193,096	193,096	188,314
R ²	0.226	0.228	0.422	0.252	0.257	0.424	0.267	0.273	0.454	0.268	0.273	0.454
Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Student FE	No	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grade FE	No	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: This table presents the estimates of regressing general education (GENED) on observables. The sample is restricted to students with disabilities enrolled in 3–8 eight for ELA and Math, and 1–12 for attendance rate and disciplinary incidents dataset. Students included in the regression attended at least one GENED and one pull-out classroom during the sample period. The dependent variable is always GENED. We used three approximations: (1) lagged measures and school fixed effects, (2) lagged measures and school fixed effects with controls, and (3) lagged measures, as well as student and school fixed effects with controls. For non-test outcomes, we also add grade fixed effects for all the models. Standard errors are clustered at the student level and reported in parenthesis. “Lagged Effects” includes the values for each variable in the previous year. Within Lagged Effects, each SDM takes a value of 1 if a student received that instruction type for that subject area in the previous year. GENED is the reference category. “Elementary” takes a value of 1 if a student is below grade six. We designate a student as having a disability if they have an IEP at any point in our dataset. The reference category for students’ race is White. “Other Race” includes students identified as American Indian, Multiracial, Native Hawaiian or Other Pacific Islander. A student is classified as having a specific disability type if, at any given moment in our dataset, they are identified as having that particular disability. The category “Different Disability” encompasses students with Blind, Brain Injury, Deaf, Deaf-Blind, Developmental Delay, Multiple Disabilities, and Orthopedic Impairment.

<i>General education vs. Self-contained</i>												
	English Language Arts			Mathematics			Attendance Rate			Disciplinary Incidents		
<i>Lagged Effects</i>												
Outcome of interest	0.079*** (0.008)	0.082*** (0.008)	0.094*** (0.024)	0.055*** (0.007)	0.053*** (0.007)	0.012 (0.019)	0.232*** (0.013)	0.234*** (0.013)	0.157*** (0.018)	-0.010*** (0.001)	-0.010*** (0.001)	-0.008*** (0.001)
Co-taught	0.155*** (0.044)	0.156*** (0.044)	-0.036 (0.104)	0.165*** (0.037)	0.177*** (0.037)	0.124 (0.108)	-0.124*** (0.004)	-0.117*** (0.003)	0.032*** (0.004)	-0.124*** (0.004)	-0.116*** (0.003)	0.033*** (0.004)
Pull-out	0.037 (0.024)	0.036 (0.024)	-0.126** (0.051)	-0.072*** (0.022)	-0.065*** (0.022)	-0.095* (0.051)	-0.262*** (0.005)	-0.244*** (0.005)	-0.066*** (0.006)	-0.260*** (0.005)	-0.243*** (0.005)	-0.065*** (0.006)
Self-contained	0.234*** (0.013)	0.234*** (0.013)	0.187*** (0.016)	0.209*** (0.011)	0.217*** (0.011)	0.164*** (0.014)	-0.392*** (0.003)	-0.358*** (0.003)	-0.099*** (0.003)	-0.392*** (0.003)	-0.358*** (0.003)	-0.099*** (0.003)
<i>Characteristics</i>												
Elementary		0.056* (0.031)	0.091** (0.039)		0.113*** (0.026)	0.142*** (0.033)						
English learner		0.035 (0.030)	0.048 (0.123)		-0.005 (0.030)	0.188 (0.185)		-0.014*** (0.004)	-0.033** (0.013)		-0.012*** (0.004)	-0.032** (0.013)
Free & reduced-price meal		-0.000 (0.013)	0.043 (0.055)		-0.027** (0.012)	-0.014 (0.049)		-0.009*** (0.001)	-0.006** (0.002)		-0.011*** (0.001)	-0.006** (0.002)
Female		-0.006 (0.012)			0.002 (0.010)			0.006*** (0.001)			0.003* (0.001)	
<i>Race</i>												
Asian		0.033 (0.075)			-0.062 (0.078)			-0.006 (0.010)			-0.005 (0.010)	
Black		-0.045** (0.022)			-0.019 (0.019)			-0.012*** (0.002)			-0.006** (0.002)	
Hispanic/latinx		-0.025 (0.024)			-0.001 (0.024)			0.010*** (0.003)			0.010*** (0.003)	
Other race		-0.032 (0.025)			0.025 (0.024)			0.005 (0.003)			0.006* (0.003)	
<i>SPED Classification</i>												
Autism spectrum disorder		-0.064* (0.035)			-0.047 (0.031)			-0.056*** (0.003)			-0.057*** (0.003)	
Different disability		-0.053 (0.042)			-0.054 (0.039)			-0.057*** (0.005)			-0.061*** (0.005)	
Emotional disability		-0.079** (0.032)			-0.043 (0.029)			-0.029*** (0.003)			-0.024*** (0.003)	
Intellectual disability		-0.014 (0.032)			-0.063** (0.029)			-0.107*** (0.003)			-0.107*** (0.003)	
Language/speech impairment		-0.043 (0.030)			-0.015 (0.025)			0.020*** (0.002)			0.020*** (0.002)	
Learning disability		-0.047 (0.029)			-0.049* (0.026)			0.013*** (0.002)			0.012*** (0.002)	
Other health impairment		-0.060** (0.030)			-0.066** (0.027)			-0.006** (0.003)			-0.006** (0.003)	
<i>Sample</i>												
Observations	6,118	6,118	3,669	7,670	7,670	5,029	219,165	219,165	212,915	219,395	219,395	213,150
R ²	0.297	0.299	0.432	0.244	0.249	0.413	0.358	0.373	0.571	0.358	0.373	0.571
Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Student FE	No	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grade FE	No	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: This table presents the estimates of regressing general education (GENED) on observables. The sample is restricted to students with disabilities enrolled in 3–8 eighth for ELA and Math, and 1–12 for attendance rate and disciplinary incidents dataset. Students included in the regression attended at least one GENED and one self-contained classroom during the sample period. The dependent variable is always GENED. We used three approximations: (1) lagged measures and school fixed effects, (2) lagged measures and school fixed effects with controls, and (3) lagged measures, as well as student and school fixed effects with controls. For non-test outcomes, we also add grade fixed effects for all the models. Standard errors are clustered at the student level and reported in parenthesis. “Lagged Effects” includes the values for each variable in the previous year. Within Lagged Effects, each SDM takes a value of 1 if a student received that instruction type for that subject area in the previous year. GENED is the reference category. “Elementary” takes a value of 1 if a student is below grade six. We designate a student as having a disability if they have an IEP at any point in our dataset. The reference category for students’ race is White. “Other Race” includes students identified as American Indian, Multiracial, Native Hawaiian or Other Pacific Islander. A student is classified as having a specific disability type if, at any given moment in our dataset, they are identified as having that particular disability. The category “Different Disability” encompasses students with Blind, Brain Injury, Deaf, Deaf-Blind, Developmental Delay, Multiple Disabilities, and Orthopedic Impairment.

<i>Pull-out vs. Self-contained</i>												
	English Language Arts			Mathematics			Attendance Rate			Disciplinary Incidents		
<i>Lagged Effects</i>												
Outcome of interest	0.046*** (0.008)	0.044*** (0.008)	0.059** (0.025)	0.016* (0.009)	0.017* (0.009)	-0.015 (0.025)	-0.030 (0.021)	-0.014 (0.021)	0.054 (0.040)	0.004*** (0.001)	0.002*** (0.001)	-0.001 (0.001)
Co-taught	-0.074 (0.064)	-0.061 (0.064)	-0.102 (0.128)	-0.086 (0.070)	-0.059 (0.070)	-0.012 (0.146)	0.005 (0.007)	0.005 (0.007)	0.001 (0.010)	0.004 (0.007)	0.004 (0.007)	0.000 (0.010)
Pull-out	-0.376*** (0.028)	-0.368*** (0.028)	-0.355*** (0.048)	-0.541*** (0.026)	-0.528*** (0.026)	-0.624*** (0.035)	-0.157*** (0.006)	-0.158*** (0.006)	-0.310*** (0.008)	-0.158*** (0.006)	-0.159*** (0.006)	-0.310*** (0.008)
Self-contained	0.021 (0.028)	0.033 (0.028)	-0.011 (0.051)	-0.095*** (0.029)	-0.075*** (0.029)	-0.196*** (0.047)	-0.062*** (0.004)	-0.061*** (0.004)	-0.035*** (0.006)	-0.062*** (0.004)	-0.062*** (0.004)	-0.035*** (0.006)
<i>Characteristics</i>												
Elementary		0.139*** (0.031)	0.192*** (0.044)		0.155*** (0.034)	0.155*** (0.046)						
English learner		0.034 (0.032)	0.318*** (0.042)		0.033 (0.037)	-0.368*** (0.037)		-0.003 (0.005)	0.009 (0.028)		-0.003 (0.005)	0.010 (0.028)
Free & reduced-price meal		-0.011 (0.012)	0.019 (0.057)		-0.012 (0.015)	0.031 (0.066)		0.002 (0.002)	-0.008 (0.006)		0.002 (0.002)	-0.008 (0.006)
Female		0.010 (0.012)			0.002 (0.013)			-0.002 (0.002)			-0.002 (0.002)	
<i>Race</i>												
Asian		-0.094 (0.079)			0.152* (0.091)			0.006 (0.013)			0.006 (0.013)	
Black		0.008 (0.023)			-0.027 (0.025)			0.007* (0.004)			0.005 (0.004)	
Hispanic/latinx		-0.003 (0.025)			0.019 (0.030)			-0.003 (0.004)			-0.003 (0.004)	
Other race		-0.045* (0.026)			0.026 (0.032)			-0.002 (0.005)			-0.002 (0.005)	
<i>SPED Classification</i>												
Autism spectrum disorder		0.040 (0.033)			0.001 (0.043)			-0.011** (0.005)			-0.011** (0.005)	
Different disability		-0.036 (0.043)			-0.085* (0.046)			-0.014** (0.006)			-0.013** (0.006)	
Emotional disability		-0.009 (0.031)			-0.021 (0.038)			0.025*** (0.004)			0.023*** (0.004)	
Intellectual disability		-0.034 (0.031)			-0.041 (0.037)			-0.020*** (0.004)			-0.020*** (0.004)	
Language/speech impairment		0.060* (0.032)			0.092*** (0.035)			0.008** (0.004)			0.008** (0.004)	
Learning disability		-0.029 (0.029)			-0.017 (0.035)			-0.010*** (0.004)			-0.010*** (0.004)	
Other health impairment		-0.020 (0.029)			-0.023 (0.036)			-0.002 (0.004)			-0.003 (0.004)	
<i>Sample</i>												
Observations	5,574	5,574	3,328	4,113	4,113	2,464	50,468	50,468	45,891	50,514	50,514	45,939
R ²	0.375	0.381	0.478	0.372	0.380	0.466	0.221	0.223	0.460	0.221	0.223	0.460
Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Student FE	No	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grade FE	No	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: This table presents the estimates of regressing pull-out on observables. The sample is restricted to students with disabilities enrolled in 3–8 eight for ELA and Math, and 1–12 for attendance rate and disciplinary incidents dataset. Students included in the regression received instruction in at least one pull-out and one self-contained setting during the sample period. The dependent variable is always pull-out. We used three approximations: (1) lagged measures and school fixed effects, (2) lagged measures and school fixed effects with controls, and (3) lagged measures, as well as student and school fixed effects with controls. For non-test outcomes, we also add grade fixed effects for all the models. Standard errors are clustered at the student level and reported in parenthesis. “Lagged Effects” includes the values for each variable in the previous year. Within Lagged Effects, each SDM takes a value of 1 if a student received that instruction type for that subject area in the previous year. GENED is the reference category. “Elementary” takes a value of 1 if a student is below grade six. We designate a student as having a disability if they have an IEP at any point in our dataset. The reference category for students’ race is White. “Other Race” includes students identified as American Indian, Multiracial, Native Hawaiian or Other Pacific Islander. A student is classified as having a specific disability type if, at any given moment in our dataset, they are identified as having that particular disability. The category “Different Disability” encompasses students with Blind, Brain Injury, Deaf, Deaf-Blind, Developmental Delay, Multiple Disabilities, and Orthopedic Impairment.

Co-taught vs. Pull-out												
	English Language Arts			Mathematics			Attendance Rate			Disciplinary Incidents		
<i>Lagged Effects</i>												
Outcome of interest	0.011* (0.006)	0.011* (0.006)	0.054*** (0.020)	-0.003 (0.007)	-0.004 (0.008)	0.064** (0.029)	0.004 (0.025)	0.001 (0.025)	-0.006 (0.043)	-0.003*** (0.001)	-0.003*** (0.001)	-0.004*** (0.001)
Co-taught	-0.144*** (0.020)	-0.153*** (0.020)	-0.126*** (0.033)	-0.216*** (0.025)	-0.227*** (0.025)	-0.258*** (0.042)	0.035*** (0.005)	0.035*** (0.005)	-0.085*** (0.006)	0.036*** (0.005)	0.036*** (0.005)	-0.085*** (0.006)
Pull-out	0.311*** (0.019)	0.303*** (0.019)	0.361*** (0.030)	0.290*** (0.024)	0.283*** (0.024)	0.264*** (0.041)	0.145*** (0.007)	0.146*** (0.007)	0.196*** (0.009)	0.146*** (0.007)	0.147*** (0.007)	0.196*** (0.009)
Self-contained	-0.064 (0.042)	-0.054 (0.041)	-0.162** (0.069)	-0.114** (0.057)	-0.108* (0.055)	-0.149 (0.092)	-0.001 (0.007)	0.001 (0.007)	0.001 (0.010)	0.000 (0.007)	0.002 (0.007)	0.002 (0.010)
<i>Characteristics</i>												
Elementary		-0.227*** (0.029)	-0.283*** (0.041)		-0.199*** (0.040)	-0.260*** (0.057)						
English learner		0.021 (0.024)	0.049 (0.216)		0.027 (0.028)	0.039 (0.396)		-0.006 (0.006)	-0.021 (0.020)		-0.006 (0.006)	-0.020 (0.020)
Free & reduced-price meal		-0.004 (0.010)	0.010 (0.046)		-0.014 (0.013)	-0.029 (0.068)		0.004* (0.003)	0.008 (0.005)		0.005* (0.003)	0.008 (0.005)
Female		0.002 (0.009)			0.005 (0.011)			0.002 (0.002)			0.002 (0.002)	
<i>Race</i>												
Asian		-0.059 (0.061)			-0.083 (0.071)			0.007 (0.018)			0.006 (0.018)	
Black		-0.011 (0.016)			-0.008 (0.018)			0.004 (0.004)			0.005 (0.004)	
Hispanic/latinx		0.001 (0.021)			-0.014 (0.023)			0.007 (0.005)			0.007 (0.005)	
Other race		-0.017 (0.019)			-0.003 (0.023)			-0.004 (0.005)			-0.004 (0.005)	
<i>SPED Classification</i>												
Autism spectrum disorder		-0.010 (0.030)			-0.011 (0.030)			0.017*** (0.005)			0.016*** (0.005)	
Different disability		-0.038 (0.040)			-0.040 (0.041)			0.001 (0.007)			0.001 (0.007)	
Emotional disability		-0.041 (0.030)			-0.034 (0.029)			-0.011** (0.005)			-0.009* (0.005)	
Intellectual disability		-0.014 (0.030)			-0.007 (0.033)			0.001 (0.005)			0.000 (0.005)	
Language/speech impairment		-0.006 (0.025)			-0.067** (0.030)			-0.008** (0.004)			-0.008** (0.004)	
Learning disability		-0.008 (0.027)			-0.004 (0.026)			0.011*** (0.004)			0.011*** (0.004)	
Other health impairment		-0.007 (0.028)			0.001 (0.026)			0.005 (0.004)			0.005 (0.004)	
<i>Sample</i>												
Observations	7,835	7,835	5,031	4,860	4,860	2,866	51,319	51,319	47,767	51,375	51,375	47,832
R ²	0.315	0.324	0.446	0.370	0.377	0.431	0.440	0.440	0.619	0.440	0.440	0.619
Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Student FE	No	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grade FE	No	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: This table presents the estimates of regressing co-taught on observables. The sample is restricted to students with disabilities enrolled in 3–8 eight for ELA and Math, and 1–12 for attendance rate and disciplinary incidents dataset. Students included in the regression attended at least one co-taught and one pull-out classroom during the sample period. The dependent variable is always GENED. We used three approximations: (1) lagged measures and school fixed effects, (2) lagged measures and school fixed effects with controls, and (3) lagged measures, as well as student and school fixed effects with controls. For non-test outcomes, we also add grade fixed effects for all the models. Standard errors are clustered at the student level and reported in parenthesis. “Lagged Effects” includes the values for each variable in the previous year. Within Lagged Effects, each SDM takes a value of 1 if a student received that instruction type for that subject area in the previous year. GENED is the reference category. “Elementary” takes a value of 1 if a student is below grade six. We designate a student as having a disability if they have an IEP at any point in our dataset. The reference category for students’ race is White. “Other Race” includes students identified as American Indian, Multiracial, Native Hawaiian or Other Pacific Islander. A student is classified as having a specific disability type if, at any given moment in our dataset, they are identified as having that particular disability. The category “Different Disability” encompasses students with Blind, Brain Injury, Deaf, Deaf-Blind, Developmental Delay, Multiple Disabilities, and Orthopedic Impairment.

Co-taught vs. Self-contained

	English Language Arts			Mathematics			Attendance Rate			Disciplinary Incidents		
<i>Lagged Effects</i>												
Outcome of interest	0.038*** (0.011)	0.041*** (0.011)	0.025 (0.032)	0.054*** (0.011)	0.056*** (0.011)	0.074** (0.032)	0.076*** (0.020)	0.078*** (0.020)	0.139*** (0.032)	-0.002*** (0.001)	-0.002*** (0.001)	-0.002** (0.001)
Co-taught	-0.190*** (0.042)	-0.217*** (0.042)	-0.082 (0.075)	-0.269*** (0.036)	-0.299*** (0.035)	-0.307*** (0.063)	0.045*** (0.004)	0.049*** (0.004)	-0.087*** (0.005)	0.045*** (0.004)	0.049*** (0.004)	-0.086*** (0.005)
Pull-out	-0.148*** (0.055)	-0.170*** (0.053)	-0.261*** (0.096)	-0.132** (0.059)	-0.153*** (0.056)	-0.259** (0.104)	-0.080*** (0.008)	-0.074*** (0.008)	-0.077*** (0.010)	-0.080*** (0.008)	-0.074*** (0.008)	-0.077*** (0.010)
Self-contained	0.125*** (0.040)	0.113*** (0.040)	0.237*** (0.072)	0.077** (0.034)	0.055* (0.033)	0.028 (0.062)	-0.119*** (0.004)	-0.106*** (0.004)	-0.043*** (0.005)	-0.119*** (0.004)	-0.106*** (0.004)	-0.043*** (0.005)
<i>Characteristics</i>												
Elementary		-0.292*** (0.037)	-0.372*** (0.054)		-0.322*** (0.037)	-0.475*** (0.053)						
English learner		-0.057 (0.042)	-0.997*** (0.004)		-0.045 (0.034)	-0.939*** (0.026)		-0.012** (0.005)	0.004 (0.022)		-0.012** (0.005)	0.004 (0.022)
Free & reduced-price meal		0.021 (0.016)	-0.085 (0.070)		0.027 (0.017)	0.004 (0.076)		0.001 (0.002)	0.003 (0.004)		0.000 (0.002)	0.003 (0.004)
Female		0.002 (0.015)			-0.009 (0.016)			0.002 (0.002)			0.001 (0.002)	
<i>Race</i>												
Asian		-0.054 (0.088)			-0.006 (0.096)			0.001 (0.018)			0.002 (0.018)	
Black		0.032 (0.025)			0.015 (0.025)			-0.003 (0.003)			-0.001 (0.003)	
Hispanic/latinx		0.029 (0.030)			-0.014 (0.029)			0.004 (0.004)			0.004 (0.004)	
Other race		0.032 (0.029)			0.026 (0.032)			-0.002 (0.004)			-0.002 (0.004)	
<i>SPED Classification</i>												
Autism spectrum disorder		-0.143*** (0.046)			-0.042 (0.050)			-0.019*** (0.005)			-0.019*** (0.005)	
Different disability		-0.129** (0.060)			-0.090 (0.057)			-0.036*** (0.006)			-0.038*** (0.006)	
Emotional disability		-0.112** (0.045)			-0.051 (0.046)			-0.020*** (0.004)			-0.019*** (0.004)	
Intellectual disability		-0.088** (0.040)			-0.011 (0.043)			-0.045*** (0.004)			-0.046*** (0.004)	
Language/speech impairment		-0.053 (0.045)			-0.053 (0.058)			0.010*** (0.003)			0.010*** (0.003)	
Learning disability		-0.071* (0.038)			0.005 (0.042)			0.001 (0.004)			0.000 (0.004)	
Other health impairment		-0.073* (0.040)			-0.036 (0.043)			-0.007* (0.004)			-0.007* (0.004)	
<i>Sample</i>												
Observations	3,332	3,332	2,044	3,104	3,104	1,870	80,307	80,307	76,133	80,379	80,379	76,216
R ²	0.288	0.311	0.430	0.295	0.322	0.397	0.402	0.404	0.598	0.402	0.404	0.598
Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Student FE	No	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grade FE	No	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: This table presents the estimates of regressing co-taught on observables. The sample is restricted to students with disabilities enrolled in 3–8 eight for ELA and Math, and 1–12 for attendance rate and disciplinary incidents dataset. Students included in the regression received instruction in at least one co-taught and one self-contained setting during the sample period. The dependent variable is always pull-out. We used three approximations: (1) lagged measures and school fixed effects, (2) lagged measures and school fixed effects with controls, and (3) lagged measures, as well as student and school fixed effects with controls. For non-test outcomes, we also add grade fixed effects for all the models. Standard errors are clustered at the student level and reported in parenthesis. “Lagged Effects” includes the values for each variable in the previous year. Within Lagged Effects, each SDM takes a value of 1 if a student received that instruction type for that subject area in the previous year. GENED is the reference category. “Elementary” takes a value of 1 if a student is below grade six. We designate a student as having a disability if they have an IEP at any point in our dataset. The reference category for students’ race is White. “Other Race” includes students identified as American Indian, Multiracial, Native Hawaiian or Other Pacific Islander. A student is classified as having a specific disability type if, at any given moment in our dataset, they are identified as having that particular disability. The category “Different Disability” encompasses students with Blind, Brain Injury, Deaf, Deaf-Blind, Developmental Delay, Multiple Disabilities, and Orthopedic Impairment.

Table C2: Regression on the Observables. Students Without Disabilities

	<i>General education vs. Co-taught</i>											
	English Language Arts			Mathematics			Attendance Rate			Disciplinary Incidents		
<i>Lagged Effects</i>												
Outcome of interest	0.019*** (0.001)	0.019*** (0.001)	0.020*** (0.004)	0.017*** (0.001)	0.018*** (0.001)	0.012*** (0.004)	0.132*** (0.004)	0.123*** (0.004)	0.057*** (0.006)	-0.006*** (0.000)	-0.005*** (0.000)	-0.002*** (0.000)
Co-taught	0.365*** (0.002)	0.366*** (0.002)	0.361*** (0.002)	0.371*** (0.002)	0.373*** (0.002)	0.370*** (0.002)	-0.037*** (0.001)	-0.036*** (0.001)	0.084*** (0.001)	-0.037*** (0.001)	-0.037*** (0.001)	0.084*** (0.001)
<i>Characteristics</i>												
Elementary		0.055*** (0.004)	0.068*** (0.005)		0.066*** (0.004)	0.082*** (0.005)						
English learner		0.010* (0.005)	0.052*** (0.018)		0.027*** (0.005)	0.100*** (0.018)		-0.011*** (0.001)	0.004** (0.002)		-0.011*** (0.001)	0.004** (0.002)
Free & reduced-price meal		-0.007*** (0.002)	0.008 (0.007)		-0.006*** (0.002)	0.008 (0.008)		-0.009*** (0.000)	-0.000 (0.001)		-0.009*** (0.000)	-0.000 (0.001)
Female		-0.002 (0.002)			-0.002 (0.002)			0.003*** (0.000)			0.002*** (0.000)	
<i>Race</i>												
Asian		0.021*** (0.007)			0.005 (0.008)			0.009*** (0.001)			0.010*** (0.001)	
Black		-0.011*** (0.003)			0.000 (0.004)			-0.009*** (0.001)			-0.007*** (0.001)	
Hispanic/latinx		-0.007** (0.003)			-0.002 (0.003)			0.001 (0.001)			0.001 (0.001)	
Other race		-0.002 (0.004)			-0.000 (0.004)			-0.002** (0.001)			-0.002** (0.001)	
<i>Sample</i>												
Observations	230,322	230,322	180,948	189,557	189,557	144,684	1,719,505	1,719,505	1,681,069	1,720,661	1,720,661	1,682,210
R ²	0.210	0.211	0.398	0.214	0.215	0.393	0.331	0.331	0.475	0.331	0.331	0.475
Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Student FE	No	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grade FE	No	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: This table presents the estimates of regressing each service delivery model (SDM) on observables for students without disabilities. The sample is restricted to students enrolled in grades three through eight for ELA and Math, and one through twelve for attendance rate and disciplinary incidents dataset. Students included in the regression attended at least one co-taught classroom during the sample period. The dependent variable is always co-taught, and the reference group is general education. The dependent variable is self-contained. We used three approximations: (1) lagged measures and school fixed effects, (2) lagged measures and school fixed effects with controls, and (3) lagged measures, as well as student and school fixed effects with controls. For non-test outcomes, we also add grade fixed effects for all the models. Standard errors are clustered at the student level and reported in the parenthesis. "Lagged Effects" includes the values for each variable in the previous year. co-taught takes a value of 1 if a student received that instruction type for that subject area in the previous year. "Elementary" takes a value of 1 if a student is below grade six. We designate a student as not having a disability if they have never had an Individualized Education Plan (IEP) at any point in our dataset.

Table C3: Service Delivery Model Effects on Test Scores by Disability Type

(a) Autism Spectrum Disorder

	English Language Arts		Mathematics		Attendance Rate		Disciplinary Incidents	
<i>General education vs. Co-taught</i>								
GENED	-0.005 (0.025)	-0.031 (0.040)	0.029 (0.030)	0.054 (0.049)	0.003* (0.002)	0.006*** (0.002)	-0.050* (0.030)	-0.091*** (0.028)
GENED × Elementary	-0.016 (0.039)	0.019 (0.064)	0.001 (0.041)	0.062 ^{††} (0.069)	-0.003 (0.002)	-0.004 (0.002)	0.050 (0.037)	0.109*** (0.036)
GENED × Middle					-0.004 (0.002)	-0.005** (0.002)	-0.075 ^{†††} (0.051)	-0.070 ^{†††} (0.054)
Observations	3,977	2,266	3,475	1,950	42,904	34,352	43,039	34,485
R^2	0.861	0.702	0.865	0.710	0.544	0.335	0.466	0.233
<i>General education vs. Pull-out</i>								
GENED	0.039 (0.024)	0.034 (0.042)	0.001 (0.030)	0.033 (0.048)	0.007** (0.003)	0.007*** (0.003)	-0.037 (0.040)	-0.063* (0.037)
GENED × Elementary	-0.053 (0.044)	-0.000 (0.076)	0.047 (0.049)	-0.010 (0.093)	-0.004 [†] (0.003)	-0.003 ^{††} (0.003)	-0.137 [†] (0.100)	-0.138 [†] (0.128)
GENED × Middle					-0.002 [†] (0.004)	-0.000 ^{†††} (0.004)	-0.140 ^{††} (0.094)	-0.061 [†] (0.079)
Observations	3,296	1,833	2,593	1,407	19,276	15,379	19,329	15,435
R^2	0.871	0.720	0.868	0.717	0.587	0.383	0.523	0.279
Lagged Outcome	No	Yes	No	Yes	No	Yes	No	Yes
Student FE	Yes	No	Yes	No	Yes	No	Yes	No
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grade FE	No	No	No	No	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; $H_0 : \beta = 0$

[†] $p < 0.10$, ^{††} $p < 0.05$, ^{†††} $p < 0.01$; $H_0 : \beta + \lambda = 0$

Notes: This table presents the effect of attending a general education (GENED) classroom versus a co-taught (Top) and pull-out (Bottom) setting on students' standardized ELA and Math scores, attendance rate, and the number of disciplinary incidents. The sample is restricted to students who were ever classified with Autism Spectrum Disorder during the sample period and includes students in grades 3–8 for test scores and grades 1–12 for non-test outcomes, with students observed in each service delivery model pair (e.g., GENED vs. co-taught) at least once. For test scores, "GENED" equals 1 if a student was in a GENED classroom for that subject. For non-test outcomes, "GENED" equals 1 if the student received instruction for ELA, Mathematics, Social Studies, and Science in GENED classrooms. "Elementary" equals 1 for students below sixth grade, and "Middle" equals 1 for grades 6–8. Each outcome includes two columns, the first of which (left column) reflects results from the school, and student fixed effects model, and the second of which (right column) reflects results from the lagged dependent model with school fixed effects. The lagged outcome used for the second approach is the student's ELA, Math, attendance rate, and number of disciplinary incidents during the previous year. Both models include student demographics as controls: gender, race, SPED classification, disability type, English Learner status, and free or reduced meal eligibility. For non-test outcomes, we also add grade fixed effects for all the models. Standard errors are clustered at the student level and reported in parentheses.

	English Language Arts		Mathematics		Attendance Rate		Disciplinary Incidents	
<i>General education vs. Self-contained</i>								
GENED	0.111 (0.071)	-0.010 (0.134)	-0.003 (0.083)	0.178 (0.138)	0.004* (0.002)	0.002 (0.002)	-0.053* (0.028)	-0.004 (0.024)
GENED × Elementary	0.085 ^{††} (0.110)	0.392 [†] (0.261)	0.213* ^{†††} (0.117)	0.087 [†] (0.210)	0.002 ^{†††} (0.003)	0.002 ^{††} (0.003)	0.092** (0.041)	-0.003 (0.041)
GENED × Middle					-0.002 (0.003)	0.001 (0.003)	0.124*** [†] (0.046)	0.046 (0.046)
Observations	640	287	797	371	26,012	19,607	26,098	19,682
R^2	0.843	0.770	0.808	0.691	0.577	0.371	0.509	0.275
<i>Pull-out vs. Self-contained</i>								
Pull-out	0.062 (0.068)	0.071 (0.153)	-0.058 (0.095)	0.047 (0.173)	0.004 (0.005)	0.002 (0.007)	0.010 (0.088)	-0.010 (0.078)
Pull-out × Elementary	-0.112 (0.101)	-0.091 (0.227)	0.048 (0.120)	0.130 (0.226)	-0.006 (0.006)	0.001 (0.008)	0.284 (0.209)	0.388 (0.259)
Pull-out × Middle					-0.000 (0.008)	-0.004 (0.009)	0.317* [†] (0.191)	0.466*** ^{††} (0.207)
Observations	648	282	473	198	5,857	4,633	5,872	4,646
R^2	0.842	0.693	0.848	0.753	0.676	0.394	0.664	0.310
Lagged Test	No	Yes	No	Yes	No	Yes	No	Yes
Student FE	Yes	No	Yes	No	Yes	No	Yes	No
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grade FE	No	No	No	No	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; $H_0: \beta = 0$

[†] $p < 0.10$, ^{††} $p < 0.05$, ^{†††} $p < 0.01$; $H_0: \beta + \lambda = 0$

Notes: This table presents the effect of attending a GENED (Top) and a pull-out (Bottom) classroom versus a self-contained one on students' standardized ELA and Math scores, attendance rate, and the number of disciplinary incidents. The sample is restricted to students who were ever classified with Autism Spectrum Disorder during the sample period and includes students in grades 3–8 for test scores and grades 1–12 for non-test outcomes, with students observed in each service delivery model pair (e.g., GENED vs. self-contained) at least once. For test scores, "GENED" equals 1 if a student was in a GENED classroom for that subject. For non-test outcomes, "GENED" equals 1 if the student received instruction for ELA, Mathematics, Social Studies, and Science in GENED classrooms. "Elementary" equals 1 for students below sixth grade, and "Middle" equals 1 for grades 6–8. Each outcome includes two columns, the first of which (left column) reflects results from the school, and student fixed effects model, and the second of which (right column) reflects results from the lagged dependent model with school fixed effects. The lagged outcome used for the second approach is the student's ELA, Math, attendance rate, and number of disciplinary incidents during the previous year. Both models include student demographics as controls: gender, race, SPED classification, disability type, English Learner status, and free or reduced meal eligibility. For non-test outcomes, we also add grade fixed effects for all the models. Standard errors are clustered at the student level and reported in parentheses.

	English Language Arts		Mathematics		Attendance Rate		Disciplinary Incidents	
<i>Co-taught vs. Pull-out</i>								
Co-taught	0.046 (0.046)	0.010 (0.081)	0.076 (0.063)	0.197* (0.113)	-0.006 (0.009)	0.004 (0.009)	0.149* (0.090)	0.094 (0.113)
Co-taught × Elementary	-0.039 (0.083)	0.011 (0.134)	-0.090 (0.107)	-0.296 (0.184)	0.002 (0.010)	-0.008 (0.010)	-0.126 (0.109)	-0.089 (0.132)
Co-taught × Middle					0.006 (0.011)	-0.004 (0.010)	-0.128 (0.188)	0.092 (0.246)
Observations	892	515	583	325	4,553	3,826	4,567	3,839
R^2	0.871	0.764	0.871	0.704	0.656	0.405	0.663	0.372
<i>Co-taught vs. Self-contained</i>								
Co-taught	0.029 (0.092)	0.053 (0.171)	0.274*** (0.095)	0.432*** (0.129)	0.001 (0.004)	0.001 (0.005)	0.024 (0.059)	-0.040 (0.065)
Co-taught × Elementary	0.007 (0.140)	0.190 (0.281)	-0.470***† (0.138)	-0.606* (0.316)	-0.005 (0.005)	-0.005 (0.006)	-0.049 (0.080)	-0.011 (0.093)
Co-taught × Middle					-0.003 (0.005)	-0.003 (0.006)	0.110 (0.123)	0.134 (0.116)
Observations	371	172	341	170	9,152	7,106	9,181	7,134
R^2	0.784	0.649	0.822	0.740	0.635	0.399	0.612	0.344
Lagged Outcome	No	Yes	No	Yes	No	Yes	No	Yes
Student FE	Yes	No	Yes	No	Yes	No	Yes	No
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grade FE	No	No	No	No	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; $H_0: \beta = 0$

† $p < 0.10$, †† $p < 0.05$, ††† $p < 0.01$; $H_0: \beta + \lambda = 0$

Notes: This table presents the effect of attending a co-taught classroom versus a pull-out (Top) and self-contained (Bottom) setting on students' standardized ELA and Math scores, attendance rate, and the number of disciplinary incidents. The sample is restricted to students who were ever classified with Autism Spectrum Disorder during the sample period and includes students in grades 3–8 for test scores and grades 1–12 for non-test outcomes, with students observed in each service delivery model pair (e.g., co-taught vs. pull-out) at least once. For test scores, "co-taught" equals 1 if a student was in a co-taught classroom for that subject. For non-test outcomes, "co-taught" equals 1 if the student received instruction for ELA, Mathematics, Social Studies, and Science in co-taught classrooms. "Elementary" equals 1 for students below sixth grade, and "Middle" equals 1 for grades 6–8. Each outcome includes two columns, the first of which (left column) reflects results from the school, and student fixed effects model, and the second of which (right column) reflects results from the lagged dependent model with school fixed effects. The lagged outcome used for the second approach is the student's ELA, Math, attendance rate, and number of disciplinary incidents during the previous year. Both models include student demographics as controls: gender, race, SPED classification, disability type, English Learner status, and free or reduced meal eligibility. For non-test outcomes, we also add grade fixed effects for all the models. Standard errors are clustered at the student level and reported in parentheses.

(b) Different Disability

	English Language Arts		Mathematics		Attendance Rate		Disciplinary Incidents	
<i>General education vs. Co-taught</i>								
GENED	0.026 (0.040)	-0.039 (0.070)	0.061 (0.044)	0.028 (0.073)	0.008** (0.004)	0.007* (0.004)	0.051 (0.049)	-0.013 (0.047)
GENED × Elementary	-0.057 (0.056)	0.092 (0.108)	-0.138**†† (0.059)	-0.188**†† (0.103)	-0.012** (0.005)	-0.010** (0.005)	-0.026 (0.052)	0.059†† (0.051)
GENED × Middle					-0.013**† (0.005)	-0.010* (0.005)	0.031 (0.071)	0.033 (0.072)
Observations	1,688	890	1,486	768	19,118	15,000	19,207	15,075
R ²	0.875	0.746	0.862	0.745	0.578	0.378	0.524	0.315
<i>General education vs. Pull-out</i>								
GENED	0.007 (0.042)	-0.053 (0.076)	-0.106* (0.061)	-0.095 (0.103)	0.005 (0.006)	-0.000 (0.007)	0.085 (0.085)	0.002 (0.073)
GENED × Elementary	-0.023 (0.073)	0.050 (0.149)	0.164* (0.098)	0.098 (0.148)	-0.002 (0.007)	0.009†† (0.008)	-0.079 (0.113)	0.169***††† (0.097)
GENED × Middle					-0.002 (0.008)	0.002 (0.009)	-0.337** (0.164)	-0.148 (0.148)
Observations	1,110	573	830	443	7,686	6,017	7,709	6,035
R ²	0.861	0.731	0.831	0.735	0.646	0.440	0.536	0.329
Lagged Outcome	No	Yes	No	Yes	No	Yes	No	Yes
Student FE	Yes	No	Yes	No	Yes	No	Yes	No
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grade FE	No	No	No	No	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; $H_0: \beta = 0$

† $p < 0.10$, †† $p < 0.05$, ††† $p < 0.01$; $H_0: \beta + \lambda = 0$

Notes: This table presents the effect of attending a general education (GENED) classroom versus a co-taught (Top) and pull-out (Bottom) setting on students' standardized ELA and Math scores, attendance rate, and the number of disciplinary incidents. The sample is restricted to students who were ever classified with Blind, Brain Injury, Deaf, Deaf-Blind, Developmental Delay, Multiple Disabilities, and Orthopedic Impairment and includes students in grades 3–8 for test scores and grades 1–12 for non-test outcomes, with students observed in each service delivery model pair (e.g., GENED vs. co-taught) at least once. For test scores, "GENED" equals 1 if a student was in a GENED classroom for that subject. For non-test outcomes, "GENED" equals 1 if the student received instruction for ELA, Mathematics, Social Studies, and Science in GENED classrooms. "Elementary" equals 1 for students below sixth grade, and "Middle" equals 1 for grades 6–8. Each outcome includes two columns, the first of which (left column) reflects results from the school, and student fixed effects model, and the second of which (right column) reflects results from the lagged dependent model with school fixed effects. The lagged outcome used for the second approach is the student's ELA, Math, attendance rate, and number of disciplinary incidents during the previous year. Both models include student demographics as controls: gender, race, SPED classification, disability type, English Learner status, and free or reduced meal eligibility. For non-test outcomes, we also add grade fixed effects for all the models. Standard errors are clustered at the student level and reported in parentheses.

	English Language Arts		Mathematics		Attendance Rate		Disciplinary Incidents	
<i>General education vs. Self-contained</i>								
GENED	-0.059 (0.107)	-0.044 (0.163)	0.221* (0.125)	0.150 (0.189)	-0.000 (0.005)	0.004 (0.005)	0.120 (0.095)	-0.036 (0.050)
GENED × Elementary	0.206 (0.192)	0.311 [†] (0.236)	-0.349* (0.209)	-0.074 (0.292)	0.004 (0.007)	0.006 ^{††} (0.007)	-0.086 (0.103)	0.113 (0.076)
GENED × Middle					0.005 (0.006)	0.000 (0.006)	-0.145 (0.099)	-0.027 (0.090)
Observations	239	106	298	133	9,700	7,221	9,753	7,263
R^2	0.800	0.806	0.696	0.750	0.636	0.413	0.526	0.302
<i>Pull-out vs. Self-contained</i>								
Pull-out	0.173 (0.111)	0.177 (0.221)	0.096 (0.260)	0.371 (0.413)	-0.001 (0.009)	-0.007 (0.009)	0.150 (0.141)	0.052 (0.104)
Pull-out × Elementary	-0.177 (0.175)	-0.212 (0.304)	0.057 (0.313)	-0.243 (0.490)	-0.001 (0.011)	0.001 (0.013)	-0.116 (0.148)	-0.118 (0.122)
Pull-out × Middle					0.007 (0.018)	0.002 (0.019)	-0.741* [†] (0.385)	-0.170 (0.341)
Observations	252	106	170	67	2,131	1,661	2,140	1,668
R^2	0.765	0.750	0.727	0.890	0.723	0.496	0.754	0.508
Lagged Outcome	No	Yes	No	Yes	No	Yes	No	Yes
Student FE	Yes	No	Yes	No	Yes	No	Yes	No
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grade FE	No	No	No	No	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; $H_0 : \beta = 0$

[†] $p < 0.10$, ^{††} $p < 0.05$, ^{†††} $p < 0.01$; $H_0 : \beta + \lambda = 0$

Notes: This table presents the effect of attending a GENED (Top) and a pull-out (Bottom) classroom versus a self-contained one on students' standardized ELA and Math scores, attendance rate, and the number of disciplinary incidents. The sample is restricted to students who were ever classified with Blind, Brain Injury, Deaf, Deaf-Blind, Developmental Delay, Multiple Disabilities, and Orthopedic Impairment and includes students in grades 3–8 for test scores and grades 1–12 for non-test outcomes, with students observed in each service delivery model pair (e.g., GENED vs. self-contained) at least once. For test scores, "GENED" equals 1 if a student was in a GENED classroom for that subject. For non-test outcomes, "GENED" equals 1 if the student received instruction for ELA, Mathematics, Social Studies, and Science in GENED classrooms. "Elementary" equals 1 for students below sixth grade, and "Middle" equals 1 for grades 6–8. Each outcome includes two columns, the first of which (left column) reflects results from the school, and student fixed effects model, and the second of which (right column) reflects results from the lagged dependent model with school fixed effects. The lagged outcome used for the second approach is the student's ELA, Math, attendance rate, and number of disciplinary incidents during the previous year. Both models include student demographics as controls: gender, race, SPED classification, disability type, English Learner status, and free or reduced meal eligibility. For non-test outcomes, we also add grade fixed effects for all the models. Standard errors are clustered at the student level and reported in parentheses.

	English Language Arts		Mathematics		Attendance Rate		Disciplinary Incidents	
<i>Co-taught vs. Pull-out</i>								
Co-taught	0.115 (0.118)	-0.071 (0.255)	0.103 (0.164)	0.209 (0.311)	-0.007 (0.009)	-0.010 (0.010)	-0.065 (0.124)	0.015 (0.116)
Co-taught × Elementary	0.033 (0.181)	0.173 (0.349)	-0.008 (0.199)	-0.788* [†] (0.428)	0.006 (0.011)	0.017 (0.013)	0.042 (0.135)	-0.101 (0.133)
Co-taught × Middle					-0.001 (0.014)	-0.008 (0.017)	0.009 (0.227)	-0.197 (0.251)
Observations	253	117	177	91	1,638	1,361	1,645	1,369
R^2	0.858	0.771	0.827	0.754	0.751	0.525	0.613	0.310
<i>Co-taught vs. Self-contained</i>								
Co-taught	0.103 (0.177)	0.053 (0.304)	0.324* (0.193)	0.362 (0.342)	-0.010 (0.010)	-0.012 (0.010)	0.017 (0.091)	0.059 (0.073)
Co-taught × Elementary	-0.314 (0.295)	0.275 (0.563)	-0.419 (0.272)	-0.380 (0.484)	0.024* (0.014)	0.015 (0.013)	-0.074 (0.111)	-0.145 (0.113)
Co-taught × Middle					0.015 (0.013)	0.003 (0.012)	-0.026 (0.149)	-0.096 (0.139)
Observations	131	59	130	65	3,288	2,581	3,312	2,598
R^2	0.873	0.830	0.798	0.748	0.726	0.541	0.693	0.435
Lagged Outcome	No	Yes	No	Yes	No	Yes	No	Yes
Student FE	Yes	No	Yes	No	Yes	No	Yes	No
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grade FE	No	No	No	No	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; $H_0 : \beta = 0$

[†] $p < 0.10$, ^{††} $p < 0.05$, ^{†††} $p < 0.01$; $H_0 : \beta + \lambda = 0$

Notes: This table presents the effect of attending a co-taught classroom versus a pull-out (Top) and self-contained (Bottom) setting on students' standardized ELA and Math scores, attendance rate, and the number of disciplinary incidents. The sample is restricted to students who were ever classified with Blind, Brain Injury, Deaf, Deaf-Blind, Developmental Delay, Multiple Disabilities, and Orthopedic Impairment and includes students in grades 3–8 for test scores and grades 1–12 for non-test outcomes, with students observed in each service delivery model pair (e.g., co-taught vs. pull-out) at least once. For test scores, "co-taught" equals 1 if a student was in a co-taught classroom for that subject. For non-test outcomes, "co-taught" equals 1 if the student received instruction for ELA, Mathematics, Social Studies, and Science in co-taught classrooms. "Elementary" equals 1 for students below sixth grade, and "Middle" equals 1 for grades 6–8. Each outcome includes two columns, the first of which (left column) reflects results from the school, and student fixed effects model, and the second of which (right column) reflects results from the lagged dependent model with school fixed effects. The lagged outcome used for the second approach is the student's ELA, Math, attendance rate, and number of disciplinary incidents during the previous year. Both models include student demographics as controls: gender, race, SPED classification, disability type, English Learner status, and free or reduced meal eligibility. For non-test outcomes, we also add grade fixed effects for all the models. Standard errors are clustered at the student level and reported in parentheses.

(c) Emotional Disability

	English Language Arts		Mathematics		Attendance Rate		Disciplinary Incidents	
<i>General education vs. Co-taught</i>								
GENED	0.017 (0.029)	-0.012 (0.047)	0.003 (0.031)	0.013 (0.043)	0.003 (0.002)	0.002 (0.003)	-0.028 (0.048)	-0.248*** (0.049)
GENED × Elementary	-0.052 (0.041)	-0.083 [†] (0.071)	-0.012 (0.043)	0.018 (0.069)	-0.003 (0.003)	0.000 (0.003)	0.070 (0.067)	0.272*** (0.073)
GENED × Middle					-0.004 (0.003)	-0.001 (0.003)	-0.196** ^{†††} (0.085)	-0.120 ^{†††} (0.087)
Observations	3,659	2,032	3,362	1,853	54,125	40,415	54,354	40,620
R ²	0.829	0.642	0.834	0.642	0.556	0.367	0.513	0.287
<i>General education vs. Pull-out</i>								
GENED	-0.008 (0.023)	-0.020 (0.041)	0.041 (0.030)	0.034 (0.051)	0.002 (0.003)	0.005 (0.003)	-0.062 (0.061)	-0.222*** (0.065)
GENED × Elementary	-0.012 (0.045)	-0.053 (0.079)	0.067 ^{†††} (0.051)	0.089 (0.099)	0.002 ^{††} (0.004)	-0.001 [†] (0.004)	-0.127 ^{††} (0.095)	-0.010 ^{†††} (0.107)
GENED × Middle					0.003 [†] (0.004)	-0.001 (0.004)	-0.348*** ^{†††} (0.111)	-0.455*** ^{†††} (0.126)
Observations	3,800	2,048	2,833	1,505	28,885	21,649	29,005	21,760
R ²	0.839	0.639	0.823	0.625	0.559	0.357	0.553	0.323
Lagged Outcome	No	Yes	No	Yes	No	Yes	No	Yes
Student FE	Yes	No	Yes	No	Yes	No	Yes	No
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grade FE	No	No	No	No	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; $H_0 : \beta = 0$

[†] $p < 0.10$, ^{††} $p < 0.05$, ^{†††} $p < 0.01$; $H_0 : \beta + \lambda = 0$

Notes: This table presents the effect of attending a general education (GENED) classroom versus a co-taught (Top) and pull-out (Bottom) setting on students' standardized ELA and Math scores, attendance rate, and the number of disciplinary incidents. The sample is restricted to students who were ever classified with Emotional disability and includes students in grades 3–8 for test scores and grades 1–12 for non-test outcomes, with students observed in each service delivery model pair (e.g., GENED vs. co-taught) at least once. For test scores, "GENED" equals 1 if a student was in a GENED classroom for that subject. For non-test outcomes, "GENED" equals 1 if the student received instruction for ELA, Mathematics, Social Studies, and Science in GENED classrooms. "Elementary" equals 1 for students below sixth grade, and "Middle" equals 1 for grades 6–8. Each outcome includes two columns, the first of which (left column) reflects results from the school, and student fixed effects model, and the second of which (right column) reflects results from the lagged dependent model with school fixed effects. The lagged outcome used for the second approach is the student's ELA, Math, attendance rate, and number of disciplinary incidents during the previous year. Both models include student demographics as controls: gender, race, SPED classification, disability type, English Learner status, and free or reduced meal eligibility. For non-test outcomes, we also add grade fixed effects for all the models. Standard errors are clustered at the student level and reported in parentheses.

	English Language Arts		Mathematics		Attendance Rate		Disciplinary Incidents	
<i>General education vs. Self-contained</i>								
GENED	0.107* (0.059)	0.104 (0.106)	0.110* (0.065)	0.108 (0.103)	0.002 (0.003)	0.005 (0.003)	-0.045 (0.059)	-0.208*** (0.059)
GENED × Elementary	-0.155* (0.088)	-0.191 (0.173)	-0.120 (0.090)	-0.097 (0.161)	0.001†† (0.003)	-0.000†† (0.003)	0.137* (0.082)	0.309***† (0.083)
GENED × Middle					-0.003 (0.004)	-0.002 (0.004)	0.017 (0.089)	-0.008††† (0.095)
Observations	1,031	454	1,295	643	36,272	25,595	36,428	25,735
R^2	0.823	0.657	0.784	0.598	0.553	0.367	0.525	0.310
<i>Pull-out vs. Self-contained</i>								
Pull-out	0.064 (0.060)	0.104 (0.103)	0.136* (0.072)	0.102 (0.125)	-0.000 (0.006)	0.000 (0.007)	0.236 (0.172)	0.072 (0.147)
Pull-out × Elementary	-0.176*† (0.090)	-0.299* (0.165)	-0.087 (0.106)	-0.204 (0.187)	-0.002 (0.007)	-0.004 (0.008)	0.145†† (0.231)	0.305†† (0.225)
Pull-out × Middle					-0.003 (0.008)	-0.009 (0.010)	0.587***††† (0.286)	1.061***††† (0.288)
Observations	943	484	752	381	8,138	6,179	8,174	6,210
R^2	0.799	0.594	0.819	0.608	0.662	0.395	0.657	0.351
Lagged Outcome	No	Yes	No	Yes	No	Yes	No	Yes
Student FE	Yes	No	Yes	No	Yes	No	Yes	No
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grade FE	No	No	No	No	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; $H_0 : \beta = 0$

† $p < 0.10$, †† $p < 0.05$, ††† $p < 0.01$; $H_0 : \beta + \lambda = 0$

Notes: This table presents the effect of attending a GENED (Top) and a pull-out (Bottom) classroom versus a self-contained one on students' standardized ELA and Math scores, attendance rate, and the number of disciplinary incidents. The sample is restricted to students who were ever classified with Emotional disability and includes students in grades 3–8 for test scores and grades 1–12 for non-test outcomes, with students observed in each service delivery model pair (e.g., GENED vs. self-contained) at least once. For test scores, "GENED" equals 1 if a student was in a GENED classroom for that subject. For non-test outcomes, "GENED" equals 1 if the student received instruction for ELA, Mathematics, Social Studies, and Science in GENED classrooms. "Elementary" equals 1 for students below sixth grade, and "Middle" equals 1 for grades 6–8. Each outcome includes two columns, the first of which (left column) reflects results from the school, and student fixed effects model, and the second of which (right column) reflects results from the lagged dependent model with school fixed effects. The lagged outcome used for the second approach is the student's ELA, Math, attendance rate, and number of disciplinary incidents during the previous year. Both models include student demographics as controls: gender, race, SPED classification, disability type, English Learner status, and free or reduced meal eligibility. For non-test outcomes, we also add grade fixed effects for all the models. Standard errors are clustered at the student level and reported in parentheses.

	English Language Arts		Mathematics		Attendance Rate		Disciplinary Incidents	
<i>Co-taught vs. Pull-out</i>								
Co-taught	-0.001 (0.048)	-0.057 (0.079)	0.066 (0.063)	0.108 (0.116)	0.012** (0.006)	0.012* (0.007)	-0.187 (0.138)	-0.057 (0.127)
Co-taught × Elementary	0.126 (0.097)	0.097 (0.169)	-0.128 (0.134)	-0.090 (0.251)	-0.014* (0.008)	-0.014* (0.009)	-0.297††† (0.223)	-0.536**††† (0.221)
Co-taught × Middle					-0.010 (0.010)	-0.001† (0.009)	-0.163 (0.274)	-0.263 (0.281)
Observations	797	451	555	290	6,867	5,423	6,909	5,458
R^2	0.844	0.636	0.810	0.642	0.681	0.369	0.670	0.344
<i>Co-taught vs. Self-contained</i>								
Co-taught	0.127 (0.078)	0.055 (0.119)	0.108 (0.086)	0.200 (0.143)	-0.002 (0.005)	0.003 (0.006)	-0.076 (0.123)	-0.122 (0.127)
Co-taught × Elementary	0.006 (0.121)	0.011 (0.215)	0.041 (0.129)	-0.020 (0.227)	0.004 (0.006)	0.002 (0.007)	-0.054 (0.168)	0.095 (0.179)
Co-taught × Middle					0.007 (0.007)	-0.003 (0.007)	0.147 (0.204)	0.185 (0.213)
Observations	604	299	591	298	12,446	9,090	12,511	9,139
R^2	0.776	0.587	0.770	0.540	0.605	0.376	0.592	0.334
Lagged Outcome	No	Yes	No	Yes	No	Yes	No	Yes
Student FE	Yes	No	Yes	No	Yes	No	Yes	No
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grade FE	No	No	No	No	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; $H_0: \beta = 0$

† $p < 0.10$, †† $p < 0.05$, ††† $p < 0.01$; $H_0: \beta + \lambda = 0$

Notes: This table presents the effect of attending a co-taught classroom versus a pull-out (Top) and self-contained (Bottom) setting on students' standardized ELA and Math scores, attendance rate, and the number of disciplinary incidents. The sample is restricted to students who were ever classified with Emotional disability and includes students in grades 3–8 for test scores and grades 1–12 for non-test outcomes, with students observed in each service delivery model pair (e.g., co-taught vs. pull-out) at least once. For test scores, "co-taught" equals 1 if a student was in a co-taught classroom for that subject. For non-test outcomes, "co-taught" equals 1 if the student received instruction for ELA, Mathematics, Social Studies, and Science in co-taught classrooms. "Elementary" equals 1 for students below sixth grade, and "Middle" equals 1 for grades 6–8. Each outcome includes two columns, the first of which (left column) reflects results from the school, and student fixed effects model, and the second of which (right column) reflects results from the lagged dependent model with school fixed effects. The lagged outcome used for the second approach is the student's ELA, Math, attendance rate, and number of disciplinary incidents during the previous year. Both models include student demographics as controls: gender, race, SPED classification, disability type, English Learner status, and free or reduced meal eligibility. For non-test outcomes, we also add grade fixed effects for all the models. Standard errors are clustered at the student level and reported in parentheses.

(d) Intellectual Disability

	English Language Arts		Mathematics		Attendance Rate		Disciplinary Incidents	
<i>General education vs. Co-taught</i>								
GENED	0.041 (0.044)	-0.012 (0.064)	-0.043 (0.061)	0.042 (0.090)	0.003 (0.003)	0.004 (0.003)	-0.063 (0.045)	-0.055 (0.049)
GENED × Elementary	-0.064 (0.062)	-0.043 (0.100)	0.005 (0.084)	0.011 (0.134)	-0.006* [†] (0.003)	-0.007** [†] (0.003)	0.140*** ^{†††} (0.053)	0.115*** ^{††} (0.055)
GENED × Middle					-0.004 (0.003)	-0.006 (0.003)	0.034 (0.067)	-0.030 (0.072)
Observations	1,870	950	1,713	865	39,352	29,669	39,599	29,905
R ²	0.707	0.514	0.701	0.545	0.566	0.382	0.529	0.292
<i>General education vs. Pull-out</i>								
GENED	0.055 (0.039)	0.090 (0.064)	0.031 (0.069)	-0.018 (0.109)	0.006 (0.004)	0.008** (0.004)	0.071 (0.054)	-0.009 (0.055)
GENED × Elementary	-0.049 (0.067)	-0.185 (0.115)	0.157 ^{†††} (0.096)	0.119 (0.172)	-0.006 (0.004)	-0.008* (0.005)	-0.120* (0.065)	-0.020 (0.066)
GENED × Middle					-0.005 (0.005)	-0.006 (0.005)	-0.251*** ^{††} (0.101)	-0.109 (0.091)
Observations	1,722	904	1,145	570	23,797	18,308	23,882	18,399
R ²	0.742	0.564	0.689	0.517	0.589	0.388	0.588	0.323
Lagged Outcome	No	Yes	No	Yes	No	Yes	No	Yes
Student FE	Yes	No	Yes	No	Yes	No	Yes	No
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grade FE	No	No	No	No	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; $H_0 : \beta = 0$

[†] $p < 0.10$, ^{††} $p < 0.05$, ^{†††} $p < 0.01$; $H_0 : \beta + \lambda = 0$

Notes: This table presents the effect of attending a general education (GENED) classroom versus a co-taught (Top) and pull-out (Bottom) setting on students' standardized ELA and Math scores, attendance rate, and the number of disciplinary incidents. The sample is restricted to students who were ever classified with Intellectual disability and includes students in grades 3–8 for test scores and grades 1–12 for non-test outcomes, with students observed in each service delivery model pair (e.g., GENED vs. co-taught) at least once. For test scores, "GENED" equals 1 if a student was in a GENED classroom for that subject. For non-test outcomes, "GENED" equals 1 if the student received instruction for ELA, Mathematics, Social Studies, and Science in GENED classrooms. "Elementary" equals 1 for students below sixth grade, and "Middle" equals 1 for grades 6–8. Each outcome includes two columns, the first of which (left column) reflects results from the school, and student fixed effects model, and the second of which (right column) reflects results from the lagged dependent model with school fixed effects. The lagged outcome used for the second approach is the student's ELA, Math, attendance rate, and number of disciplinary incidents during the previous year. Both models include student demographics as controls: gender, race, SPED classification, disability type, English Learner status, and free or reduced meal eligibility. For non-test outcomes, we also add grade fixed effects for all the models. Standard errors are clustered at the student level and reported in parentheses.

	English Language Arts		Mathematics		Attendance Rate		Disciplinary Incidents	
<i>General education vs. Self-contained</i>								
GENED	0.120*	0.090	0.128*	0.085	-0.011***	-0.005**	0.082**	0.030
	(0.062)	(0.119)	(0.076)	(0.108)	(0.002)	(0.002)	(0.036)	(0.034)
GENED × Elementary	-0.219*	-0.533**††	-0.044	0.241††	0.015***†††	0.007**	-0.120***	-0.030
	(0.127)	(0.248)	(0.112)	(0.180)	(0.003)	(0.003)	(0.045)	(0.042)
GENED × Middle					0.010***	0.006**	-0.031	0.023
					(0.003)	(0.003)	(0.049)	(0.051)
Observations	733	361	927	463	46,556	33,899	46,783	34,113
R ²	0.681	0.498	0.628	0.596	0.567	0.376	0.562	0.312
<i>Pull-out vs. Self-contained</i>								
Pull-out	-0.015	0.074	0.031	0.130	-0.013**	-0.015**	-0.001	0.007
	(0.067)	(0.120)	(0.092)	(0.171)	(0.006)	(0.006)	(0.123)	(0.090)
Pull-out × Elementary	0.025	-0.083	0.070	-0.180	0.018***	0.013*	0.090	0.106
	(0.100)	(0.172)	(0.130)	(0.230)	(0.006)	(0.007)	(0.131)	(0.104)
Pull-out × Middle					0.013*	0.015**	0.332*	0.226
					(0.007)	(0.008)	(0.175)	(0.157)
Observations	899	435	668	323	11,863	9,236	11,900	9,276
R ²	0.640	0.458	0.662	0.419	0.659	0.407	0.698	0.409
Lagged Outcome	No	Yes	No	Yes	No	Yes	No	Yes
Student FE	Yes	No	Yes	No	Yes	No	Yes	No
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grade FE	No	No	No	No	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; $H_0 : \beta = 0$

† $p < 0.10$, †† $p < 0.05$, ††† $p < 0.01$; $H_0 : \beta + \lambda = 0$

Notes: This table presents the effect of attending a GENED (Top) and a pull-out (Bottom) classroom versus a self-contained one on students' standardized ELA and Math scores, attendance rate, and the number of disciplinary incidents. The sample is restricted to students who were ever classified with Intellectual disability and includes students in grades 3–8 for test scores and grades 1–12 for non-test outcomes, with students observed in each service delivery model pair (e.g., GENED vs. self-contained) at least once. For test scores, "GENED" equals 1 if a student was in a GENED classroom for that subject. For non-test outcomes, "GENED" equals 1 if the student received instruction for ELA, Mathematics, Social Studies, and Science in GENED classrooms. "Elementary" equals 1 for students below sixth grade, and "Middle" equals 1 for grades 6–8. Each outcome includes two columns, the first of which (left column) reflects results from the school, and student fixed effects model, and the second of which (right column) reflects results from the lagged dependent model with school fixed effects. The lagged outcome used for the second approach is the student's ELA, Math, attendance rate, and number of disciplinary incidents during the previous year. Both models include student demographics as controls: gender, race, SPED classification, disability type, English Learner status, and free or reduced meal eligibility. For non-test outcomes, we also add grade fixed effects for all the models. Standard errors are clustered at the student level and reported in parentheses.

	English Language Arts		Mathematics		Attendance Rate		Disciplinary Incidents	
<i>Co-taught vs. Pull-out</i>								
Co-taught	0.031 (0.076)	0.086 (0.104)	-0.081 (0.099)	-0.169 (0.128)	0.008 (0.006)	0.016** (0.006)	-0.051 (0.110)	-0.053 (0.091)
Co-taught × Elementary	0.080 (0.120)	0.014 (0.206)	0.037 (0.163)	0.455 (0.280)	-0.002† (0.008)	-0.006†† (0.008)	-0.161††† (0.136)	-0.150†† (0.123)
Co-taught × Middle					-0.006 (0.009)	-0.009 (0.008)	-0.059 (0.179)	0.014 (0.165)
Observations	659	358	388	184	6,791	5,419	6,827	5,449
R^2	0.688	0.431	0.700	0.596	0.665	0.434	0.683	0.394
<i>Co-taught vs. Self-contained</i>								
Co-taught	0.107 (0.072)	0.179 (0.122)	0.097 (0.091)	0.212 (0.131)	-0.002 (0.004)	0.005 (0.005)	0.039 (0.068)	0.008 (0.073)
Co-taught × Elementary	-0.269 (0.178)	-0.292 (0.192)	-0.201 (0.162)	-0.368 (0.299)	0.011***††† (0.004)	-0.006 (0.006)	-0.034 (0.077)	-0.036 (0.086)
Co-taught × Middle					0.003 (0.005)	-0.003 (0.005)	0.077† (0.094)	-0.032 (0.108)
Observations	516	270	558	298	18,348	13,785	18,455	13,879
R^2	0.653	0.447	0.642	0.509	0.594	0.385	0.643	0.375
Lagged Outcome	No	Yes	No	Yes	No	Yes	No	Yes
Student FE	Yes	No	Yes	No	Yes	No	Yes	No
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grade FE	No	No	No	No	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; $H_0 : \beta = 0$

† $p < 0.10$, †† $p < 0.05$, ††† $p < 0.01$; $H_0 : \beta + \lambda = 0$

Notes: This table presents the effect of attending a co-taught classroom versus a pull-out (Top) and self-contained (Bottom) setting on students' standardized ELA and Math scores, attendance rate, and the number of disciplinary incidents. The sample is restricted to students who were ever classified with Intellectual disability and includes students in grades 3–8 for test scores and grades 1–12 for non-test outcomes, with students observed in each service delivery model pair (e.g., co-taught vs. pull-out) at least once. For test scores, “co-taught” equals 1 if a student was in a co-taught classroom for that subject. For non-test outcomes, “co-taught” equals 1 if the student received instruction for ELA, Mathematics, Social Studies, and Science in co-taught classrooms. “Elementary” equals 1 for students below sixth grade, and “Middle” equals 1 for grades 6–8. Each outcome includes two columns, the first of which (left column) reflects results from the school, and student fixed effects model, and the second of which (right column) reflects results from the lagged dependent model with school fixed effects. The lagged outcome used for the second approach is the student's ELA, Math, attendance rate, and number of disciplinary incidents during the previous year. Both models include student demographics as controls: gender, race, SPED classification, disability type, English Learner status, and free or reduced meal eligibility. For non-test outcomes, we also add grade fixed effects for all the models. Standard errors are clustered at the student level and reported in parentheses.

(e) Language or Speech Impairment

	English Language Arts		Mathematics		Attendance Rate		Disciplinary Incidents	
<i>General education vs. Co-taught</i>								
GENED	-0.007 (0.012)	0.012 (0.018)	-0.024 (0.015)	-0.012 (0.020)	0.004** (0.002)	0.006*** (0.002)	-0.095*** (0.023)	-0.127*** (0.024)
GENED × Elementary	-0.005 (0.015)	-0.008† (0.023)	0.008 (0.017)	0.011 (0.025)	-0.006***††† (0.002)	-0.006*** (0.002)	0.101*** (0.024)	0.130*** (0.024)
GENED × Middle					-0.002††† (0.002)	-0.004*††† (0.002)	-0.008††† (0.029)	-0.000††† (0.030)
Observations	20,057	12,065	17,875	10,468	202,865	164,850	203,703	165,647
R ²	0.851	0.653	0.865	0.689	0.489	0.357	0.423	0.223
<i>General education vs. Pull-out</i>								
GENED	-0.005 (0.038)	-0.001 (0.057)	-0.013 (0.053)	-0.041 (0.075)	0.010* (0.006)	0.013** (0.006)	-0.053 (0.067)	-0.023 (0.074)
GENED × Elementary	0.015 (0.050)	0.056 (0.079)	-0.042 (0.064)	-0.073 (0.099)	-0.010 (0.006)	-0.010 (0.006)	-0.048†† (0.078)	-0.022 (0.085)
GENED × Middle					-0.004† (0.007)	-0.003††† (0.007)	-0.158†† (0.123)	-0.268**††† (0.128)
Observations	2421	1,426	1,578	910	22,976	18,624	23,055	18,705
R ²	0.831	0.653	0.850	0.694	0.563	0.423	0.523	0.310
Lagged Outcome	No	Yes	No	Yes	No	Yes	No	Yes
Student FE	Yes	No	Yes	No	Yes	No	Yes	No
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grade FE	No	No	No	No	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; $H_0 : \beta = 0$

† $p < 0.10$, †† $p < 0.05$, ††† $p < 0.01$; $H_0 : \beta + \lambda = 0$

Notes: This table presents the effect of attending a general education (GENED) classroom versus a co-taught (Top) and pull-out (Bottom) setting on students' standardized ELA and Math scores, attendance rate, and the number of disciplinary incidents. The sample is restricted to students who were ever classified with Language or Speech Impairment and includes students in grades 3–8 for test scores and grades 1–12 for non-test outcomes, with students observed in each service delivery model pair (e.g., GENED vs. co-taught) at least once. For test scores, "GENED" equals 1 if a student was in a GENED classroom for that subject. For non-test outcomes, "GENED" equals 1 if the student received instruction for ELA, Mathematics, Social Studies, and Science in GENED classrooms. "Elementary" equals 1 for students below sixth grade, and "Middle" equals 1 for grades 6–8. Each outcome includes two columns, the first of which (left column) reflects results from the school, and student fixed effects model, and the second of which (right column) reflects results from the lagged dependent model with school fixed effects. The lagged outcome used for the second approach is the student's ELA, Math, attendance rate, and number of disciplinary incidents during the previous year. Both models include student demographics as controls: gender, race, SPED classification, disability type, English Learner status, and free or reduced meal eligibility. For non-test outcomes, we also add grade fixed effects for all the models. Standard errors are clustered at the student level and reported in parentheses.

	English Language Arts		Mathematics		Attendance Rate		Disciplinary Incidents	
<i>General education vs. Self-contained</i>								
GENED	0.105 (0.095)	0.159 (0.139)	0.012 (0.113)	0.213 (0.165)	0.005 (0.005)	0.008 (0.005)	-0.128* (0.076)	-0.171** (0.078)
GENED × Elementary	-0.180 (0.127)	-0.051 (0.215)	-0.135 (0.142)	-0.069 (0.218)	-0.002†† (0.006)	-0.005† (0.006)	0.115 (0.086)	0.150* (0.088)
GENED × Middle					-0.003 (0.006)	-0.004 (0.006)	0.032 (0.105)	0.041† (0.111)
Observations	461	255	529	294	21,643	17,544	21,749	17,651
R^2	0.802	0.726	0.755	0.687	0.528	0.382	0.531	0.300
<i>Pull-out vs. Self-contained</i>								
Pull-out	0.025 (0.130)	0.045 (0.189)	-0.445 (0.284)	-0.761*** (0.271)	-0.025 (0.018)	-0.014 (0.017)	-0.204 (0.283)	-0.010 (0.237)
Pull-out × Elementary	-0.292†† (0.183)	-0.200 (0.270)	0.292 (0.343)	0.717* (0.374)	0.020 (0.019)	0.013 (0.017)	0.591*†† (0.326)	0.394††† (0.277)
Pull-out × Middle					0.023 (0.019)	0.009 (0.018)	0.800**† (0.398)	0.885**††† (0.381)
Observations	176	122	130	88	3,764	3,539	3,776	3,554
R^2	0.802	0.750	0.733	0.799	0.719	0.464	0.705	0.405
Lagged Outcome	No	Yes	No	Yes	No	Yes	No	Yes
Student FE	Yes	No	Yes	No	Yes	No	Yes	No
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grade FE	No	No	No	No	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; $H_0: \beta = 0$

† $p < 0.10$, †† $p < 0.05$, ††† $p < 0.01$; $H_0: \beta + \lambda = 0$

Notes: This table presents the effect of attending a GENED (Top) and a pull-out (Bottom) classroom versus a self-contained one on students' standardized ELA and Math scores, attendance rate, and the number of disciplinary incidents. The sample is restricted to students who were ever classified with Language or Speech Impairment and includes students in grades 3–8 for test scores and grades 1–12 for non-test outcomes, with students observed in each service delivery model pair (e.g., GENED vs. self-contained) at least once. For test scores, "GENED" equals 1 if a student was in a GENED classroom for that subject. For non-test outcomes, "GENED" equals 1 if the student received instruction for ELA, Mathematics, Social Studies, and Science in GENED classrooms, "Elementary" equals 1 for students below sixth grade, and "Middle" equals 1 for grades 6–8. Each outcome includes two columns, the first of which (left column) reflects results from the school, and student fixed effects model, and the second of which (right column) reflects results from the lagged dependent model with school fixed effects. The lagged outcome used for the second approach is the student's ELA, Math, attendance rate, and number of disciplinary incidents during the previous year. Both models include student demographics as controls: gender, race, SPED classification, disability type, English Learner status, and free or reduced meal eligibility. For non-test outcomes, we also add grade fixed effects for all the models. Standard errors are clustered at the student level and reported in parentheses.

	English Language Arts		Mathematics		Attendance Rate		Disciplinary Incidents	
<i>Co-taught vs. Pull-out</i>								
Co-taught	-0.106 (0.088)	-0.158 (0.124)	0.043 (0.132)	0.122 (0.197)	0.015 (0.011)	0.009 (0.011)	-0.110 (0.122)	-0.078 (0.116)
Co-taught × Elementary	-0.025 (0.149)	0.077 (0.247)	0.160 (0.179)	0.088 (0.333)	-0.016 (0.011)	-0.006 (0.011)	0.026 (0.139)	0.004 (0.133)
Co-taught × Middle					-0.008 (0.012)	-0.004 (0.012)	0.088 (0.210)	0.039 (0.207)
Observations	303	195	205	135	4,624	4,159	4,644	4,177
R^2	0.859	0.760	0.843	0.754	0.687	0.462	0.714	0.404
<i>Co-taught vs. Self-contained</i>								
Co-taught	-0.092 (0.211)	0.099 (0.182)	0.025 (0.233)	0.022 (0.297)	0.015** (0.007)	0.021*** (0.007)	0.039 (0.116)	0.101 (0.118)
Co-taught × Elementary	0.422 (0.315)	0.313 ^{††} (0.261)	-0.234 (0.421)	0.011 (0.697)	-0.020*** (0.007)	-0.021*** (0.008)	0.001 (0.131)	-0.093 (0.136)
Co-taught × Middle					-0.012 (0.008)	-0.022*** (0.008)	-0.144 (0.175)	-0.149 (0.185)
Observations	94	63	74	48	6,550	5,888	6,585	5,924
R^2	0.821	0.920	0.753	0.842	0.634	0.421	0.648	0.363
Lagged Outcome	No	Yes	No	Yes	No	Yes	No	Yes
Student FE	Yes	No	Yes	No	Yes	No	Yes	No
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grade FE	No	No	No	No	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; $H_0 : \beta = 0$

[†] $p < 0.10$, ^{††} $p < 0.05$, ^{†††} $p < 0.01$; $H_0 : \beta + \lambda = 0$

Notes: This table presents the effect of attending a co-taught classroom versus a pull-out (Top) and self-contained (Bottom) setting on students' standardized ELA and Math scores, attendance rate, and the number of disciplinary incidents. The sample is restricted to students who were ever classified with Language or Speech Impairment and includes students in grades 3–8 for test scores and grades 1–12 for non-test outcomes, with students observed in each service delivery model pair (e.g., co-taught vs. pull-out) at least once. For test scores, "co-taught" equals 1 if a student was in a co-taught classroom for that subject. For non-test outcomes, "co-taught" equals 1 if the student received instruction for ELA, Mathematics, Social Studies, and Science in co-taught classrooms. "Elementary" equals 1 for students below sixth grade, and "Middle" equals 1 for grades 6–8. Each outcome includes two columns, the first of which (left column) reflects results from the school, and student fixed effects model, and the second of which (right column) reflects results from the lagged dependent model with school fixed effects. The lagged outcome used for the second approach is the student's ELA, Math, attendance rate, and number of disciplinary incidents during the previous year. Both models include student demographics as controls: gender, race, SPED classification, disability type, English Learner status, and free or reduced meal eligibility. For non-test outcomes, we also add grade fixed effects for all the models. Standard errors are clustered at the student level and reported in parentheses.

(f) Learning Disability

	English Language Arts		Mathematics		Attendance Rate		Disciplinary Incidents	
<i>General education vs. Co-taught</i>								
GENED	-0.005 (0.008)	0.020* (0.012)	-0.023** (0.010)	0.004 (0.015)	0.003*** (0.001)	0.004*** (0.001)	-0.025 (0.016)	-0.080*** (0.015)
GENED × Elementary	-0.006 (0.012)	-0.037* (0.019)	-0.004††† (0.014)	-0.028 (0.021)	-0.004*** (0.001)	-0.002**††† (0.001)	0.044**†† (0.018)	0.085*** (0.017)
GENED × Middle					-0.004*** (0.001)	-0.004*** (0.001)	-0.071***††† (0.025)	-0.043*††† (0.025)
Observations	27,732	16,338	25,638	14,580	292,651	232,379	293,692	233,435
R^2	0.767	0.483	0.789	0.525	0.518	0.360	0.502	0.269
<i>General education vs. Pull-out</i>								
GENED	0.006 (0.008)	0.026** (0.013)	-0.010 (0.012)	-0.008 (0.017)	0.006*** (0.002)	0.007*** (0.002)	-0.056*** (0.022)	-0.127*** (0.022)
GENED × Elementary	0.010 (0.014)	0.036††† (0.022)	0.053*** (0.018)	0.086*** (0.029)	-0.006*** (0.002)	-0.004**††† (0.002)	0.044 (0.030)	0.097*** (0.029)
GENED × Middle					-0.006*** (0.002)	-0.004*†† (0.002)	-0.133***††† (0.043)	-0.121***††† (0.047)
Observations	23,590	13,942	16,260	9,583	132,209	105,774	132,600	106,195
R^2	0.766	0.487	0.780	0.511	0.520	0.357	0.512	0.286
Lagged Outcome	No	Yes	No	Yes	No	Yes	No	Yes
Student FE	Yes	No	Yes	No	Yes	No	Yes	No
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grade FE	No	No	No	No	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; $H_0 : \beta = 0$

† $p < 0.10$, †† $p < 0.05$, ††† $p < 0.01$; $H_0 : \beta + \lambda = 0$

Notes: This table presents the effect of attending a general education (GENED) classroom versus a co-taught (Top) and pull-out (Bottom) setting on students' standardized ELA and Math scores, attendance rate, and the number of disciplinary incidents. The sample is restricted to students who were ever classified with Learning disability and includes students in grades 3–8 for test scores and grades 1–12 for non-test outcomes, with students observed in each service delivery model pair (e.g., GENED vs. co-taught) at least once. For test scores, "GENED" equals 1 if a student was in a GENED classroom for that subject. For non-test outcomes, "GENED" equals 1 if the student received instruction for ELA, Mathematics, Social Studies, and Science in GENED classrooms. "Elementary" equals 1 for students below sixth grade, and "Middle" equals 1 for grades 6–8. Each outcome includes two columns, the first of which (left column) reflects results from the school, and student fixed effects model, and the second of which (right column) reflects results from the lagged dependent model with school fixed effects. The lagged outcome used for the second approach is the student's ELA, Math, attendance rate, and number of disciplinary incidents during the previous year. Both models include student demographics as controls: gender, race, SPED classification, disability type, English Learner status, and free or reduced meal eligibility. For non-test outcomes, we also add grade fixed effects for all the models. Standard errors are clustered at the student level and reported in parentheses.

	English Language Arts		Mathematics		Attendance Rate		Disciplinary Incidents	
<i>General education vs. Self-contained</i>								
GENED	0.064*** (0.019)	0.091*** (0.030)	0.081*** (0.023)	0.103*** (0.034)	0.003** (0.001)	0.003** (0.001)	-0.056** (0.023)	-0.139*** (0.022)
GENED × Elementary	-0.026 (0.034)	-0.037 (0.053)	-0.042 (0.034)	-0.024†† (0.050)	0.002††† (0.002)	0.001††† (0.002)	0.074** (0.031)	0.112*** (0.029)
GENED × Middle					-0.003** (0.002)	-0.002 (0.002)	-0.031††† (0.038)	-0.016††† (0.039)
Observations	5,414	3,196	6,332	3,659	142,068	112,713	142,526	113,207
R^2	0.741	0.472	0.705	0.459	0.532	0.372	0.525	0.301
<i>Pull-out vs. Self-contained</i>								
Pull-out	0.036 (0.024)	0.010 (0.037)	0.022 (0.037)	0.043 (0.053)	-0.004 (0.003)	-0.007** (0.003)	-0.010 (0.053)	-0.003 (0.050)
Pull-out × Elementary	0.005† (0.034)	0.055† (0.054)	-0.020 (0.050)	-0.026 (0.071)	0.005 (0.004)	0.009** (0.004)	0.091†† (0.064)	0.030 (0.058)
Pull-out × Middle					0.005 (0.005)	0.002† (0.005)	0.369***††† (0.123)	0.515***††† (0.132)
Observations	4,422	2,609	2,917	1,772	29,582	24,615	29,653	24,681
R^2	0.763	0.501	0.701	0.447	0.654	0.404	0.669	0.373
Lagged Outcome	No	Yes	No	Yes	No	Yes	No	Yes
Student FE	Yes	No	Yes	No	Yes	No	Yes	No
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grade FE	No	No	No	No	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; $H_0: \beta = 0$

† $p < 0.10$, †† $p < 0.05$, ††† $p < 0.01$; $H_0: \beta + \lambda = 0$

Notes: This table presents the effect of attending a GENED (Top) and a pull-out (Bottom) classroom versus a self-contained one on students' standardized ELA and Math scores, attendance rate, and the number of disciplinary incidents. The sample is restricted to students who were ever classified with Learning disability and includes students in grades 3–8 for test scores and grades 1–12 for non-test outcomes, with students observed in each service delivery model pair (e.g., GENED vs. self-contained) at least once. For test scores, "GENED" equals 1 if a student was in a GENED classroom for that subject. For non-test outcomes, "GENED" equals 1 if the student received instruction for ELA, Mathematics, Social Studies, and Science in GENED classrooms. "Elementary" equals 1 for students below sixth grade, and "Middle" equals 1 for grades 6–8. Each outcome includes two columns, the first of which (left column) reflects results from the school, and student fixed effects model, and the second of which (right column) reflects results from the lagged dependent model with school fixed effects. The lagged outcome used for the second approach is the student's ELA, Math, attendance rate, and number of disciplinary incidents during the previous year. Both models include student demographics as controls: gender, race, SPED classification, disability type, English Learner status, and free or reduced meal eligibility. For non-test outcomes, we also add grade fixed effects for all the models. Standard errors are clustered at the student level and reported in parentheses.

	English Language Arts		Mathematics		Attendance Rate		Disciplinary Incidents	
<i>Co-taught vs. Pull-out</i>								
Co-taught	0.033** (0.015)	0.015 (0.023)	0.070*** (0.022)	0.042 (0.034)	0.004 (0.003)	0.006** (0.003)	0.020 (0.046)	0.019 (0.045)
Co-taught × Elementary	-0.011 (0.029)	-0.030 (0.049)	0.064††† (0.039)	0.088†† (0.070)	-0.002 (0.003)	-0.006* (0.003)	-0.037 (0.060)	0.034 (0.056)
Co-taught × Middle					0.000†† (0.004)	-0.001†† (0.003)	0.065 (0.093)	0.057 (0.094)
Observations	6,352	3,923	3,969	2,471	33,990	28,427	34,122	28,555
R^2	0.774	0.482	0.779	0.516	0.611	0.377	0.604	0.296
<i>Co-taught vs. Self-contained</i>								
Co-taught	0.058** (0.028)	0.038 (0.042)	0.257*** (0.038)	0.237*** (0.055)	0.009*** (0.002)	0.006*** (0.002)	-0.053 (0.038)	-0.095** (0.038)
Co-taught × Elementary	0.001 (0.048)	0.111†† (0.075)	-0.129**††† (0.062)	0.010††† (0.105)	-0.002††† (0.003)	-0.007*** (0.003)	0.024 (0.050)	0.112** (0.047)
Co-taught × Middle					-0.006** (0.003)	-0.005** (0.002)	0.109* (0.064)	0.120* (0.067)
Observations	2,490	1,541	2,180	1,306	49,544	40,893	49,755	41,095
R^2	0.740	0.427	0.700	0.447	0.592	0.383	0.626	0.332
Lagged Outcome	No	Yes	No	Yes	No	Yes	No	Yes
Student FE	Yes	No	Yes	No	Yes	No	Yes	No
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grade FE	No	No	No	No	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; $H_0 : \beta = 0$

† $p < 0.10$, †† $p < 0.05$, ††† $p < 0.01$; $H_0 : \beta + \lambda = 0$

Notes: This table presents the effect of attending a co-taught classroom versus a pull-out (Top) and self-contained (Bottom) setting on students' standardized ELA and Math scores, attendance rate, and the number of disciplinary incidents. The sample is restricted to students who were ever classified with Learning disability and includes students in grades 3–8 for test scores and grades 1–12 for non-test outcomes, with students observed in each service delivery model pair (e.g., co-taught vs. pull-out) at least once. For test scores, "Co-taught" equals 1 if a student was in a co-taught classroom for that subject. For non-test outcomes, "Co-taught" equals 1 if the student received instruction for ELA, Mathematics, Social Studies, and Science in co-taught classrooms. "Elementary" equals 1 for students below sixth grade, and "Middle" equals 1 for grades 6–8. Each outcome includes two columns, the first of which (left column) reflects results from the school, and student fixed effects model, and the second of which (right column) reflects results from the lagged dependent model with school fixed effects. The lagged outcome used for the second approach is the student's ELA, Math, attendance rate, and number of disciplinary incidents during the previous year. Both models include student demographics as controls: gender, race, SPED classification, disability type, English Learner status, and free or reduced meal eligibility. For non-test outcomes, we also add grade fixed effects for all the models. Standard errors are clustered at the student level and reported in parentheses.

(g) Other Health Impairment

	English Language Arts		Mathematics		Attendance Rate		Disciplinary Incidents	
<i>General education vs. Co-taught</i>								
GENED	0.007 (0.015)	0.010 (0.023)	-0.014 (0.018)	-0.009 (0.027)	0.003** (0.002)	0.003** (0.002)	-0.070** (0.031)	-0.151*** (0.030)
GENED × Elementary	-0.004 (0.022)	-0.010 (0.035)	-0.008 (0.024)	0.012 (0.037)	-0.005** (0.002)	-0.002 (0.002)	0.046 (0.035)	0.120*** (0.036)
GENED × Middle					-0.002 (0.002)	-0.002 (0.002)	-0.004†† (0.046)	0.072†† (0.048)
Observations	10,916	6,279	10,165	5,665	115,808	90,946	116,233	91,374
R^2	0.800	0.556	0.806	0.562	0.526	0.363	0.504	0.285
<i>General education vs. Pull-out</i>								
GENED	0.010 (0.015)	-0.006 (0.024)	0.002 (0.019)	-0.012 (0.028)	0.005* (0.002)	0.006** (0.002)	-0.068 (0.042)	-0.089** (0.043)
GENED × Elementary	-0.000 (0.025)	0.035 (0.043)	0.046† (0.031)	0.076 (0.052)	-0.005 (0.003)	-0.002††† (0.003)	0.023 (0.061)	-0.057††† (0.060)
GENED × Middle					-0.001† (0.003)	0.002††† (0.003)	-0.117††† (0.079)	-0.112††† (0.084)
Observations	9,033	5,092	6,326	3,547	50,245	39,668	50,421	39,852
R^2	0.799	0.550	0.803	0.580	0.550	0.394	0.538	0.318
Lagged Outcome	No	Yes	No	Yes	No	Yes	No	Yes
Student FE	Yes	No	Yes	No	Yes	No	Yes	No
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grade FE	No	No	No	No	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; $H_0 : \beta = 0$

† $p < 0.10$, †† $p < 0.05$, ††† $p < 0.01$; $H_0 : \beta + \lambda = 0$

Notes: This table presents the effect of attending a general education (GENED) classroom versus a co-taught (Top) and pull-out (Bottom) setting on students' standardized ELA and Math scores, attendance rate, and the number of disciplinary incidents. The sample is restricted to students who were ever classified with Other Health Impairment and includes students in grades 3–8 for test scores and grades 1–12 for non-test outcomes, with students observed in each service delivery model pair (e.g., GENED vs. co-taught) at least once. For test scores, "GENED" equals 1 if a student was in a GENED classroom for that subject. For non-test outcomes, "GENED" equals 1 if the student received instruction for ELA, Mathematics, Social Studies, and Science in GENED classrooms. "Elementary" equals 1 for students below sixth grade, and "Middle" equals 1 for grades 6–8. Each outcome includes two columns, the first of which (left column) reflects results from the school, and student fixed effects model, and the second of which (right column) reflects results from the lagged dependent model with school fixed effects. The lagged outcome used for the second approach is the student's ELA, Math, attendance rate, and number of disciplinary incidents during the previous year. Both models include student demographics as controls: gender, race, SPED classification, disability type, English Learner status, and free or reduced meal eligibility. For non-test outcomes, we also add grade fixed effects for all the models. Standard errors are clustered at the student level and reported in parentheses.

	English Language Arts		Mathematics		Attendance Rate		Disciplinary Incidents	
<i>General education vs. Self-contained</i>								
GENED	0.009 (0.042)	0.056 (0.064)	0.114*** (0.043)	0.186*** (0.065)	0.001 (0.002)	0.003 (0.002)	-0.055 (0.042)	-0.097** (0.039)
GENED × Elementary	0.074 (0.070)	-0.016 (0.120)	-0.022† (0.063)	0.029†† (0.097)	0.005*††† (0.003)	0.004*††† (0.003)	0.097 (0.060)	0.040 (0.055)
GENED × Middle					0.001 (0.003)	0.001†† (0.003)	0.072 (0.064)	0.026 (0.067)
Observations	1,545	765	2,144	1,116	51,403	39,861	51,598	40,049
R^2	0.734	0.587	0.720	0.551	0.547	0.388	0.552	0.325
<i>Pull-out vs. Self-contained</i>								
Pull-out	-0.021 (0.046)	0.017 (0.082)	0.027 (0.058)	-0.001 (0.098)	-0.006 (0.005)	-0.012** (0.006)	-0.042 (0.092)	-0.044 (0.094)
Pull-out × Elementary	-0.077† (0.071)	-0.157 (0.132)	0.001 (0.087)	-0.025 (0.158)	0.008 (0.006)	0.003†† (0.007)	0.222*†† (0.124)	0.318***†† (0.145)
Pull-out × Middle					0.009 (0.008)	0.004 (0.008)	0.651***††† (0.223)	0.724***††† (0.226)
Observations	1,236	616	1,129	598	10,673	8,885	10,700	8,904
R^2	0.785	0.579	0.761	0.544	0.669	0.392	0.687	0.371
Lagged Outcome	No	Yes	No	Yes	No	Yes	No	Yes
Student FE	Yes	No	Yes	No	Yes	No	Yes	No
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grade FE	No	No	No	No	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; $H_0 : \beta = 0$

† $p < 0.10$, †† $p < 0.05$, ††† $p < 0.01$; $H_0 : \beta + \lambda = 0$

Notes: This table presents the effect of attending a GENED (Top) and a pull-out (Bottom) classroom versus a self-contained one on students' standardized ELA and Math scores, attendance rate, and the number of disciplinary incidents. The sample is restricted to students who were ever classified with Other Health Impairment and includes students in grades 3–8 for test scores and grades 1–12 for non-test outcomes, with students observed in each service delivery model pair (e.g., GENED vs. self-contained) at least once. For test scores, "GENED" equals 1 if a student was in a GENED classroom for that subject. For non-test outcomes, "GENED" equals 1 if the student received instruction for ELA, Mathematics, Social Studies, and Science in GENED classrooms. "Elementary" equals 1 for students below sixth grade, and "Middle" equals 1 for grades 6–8. Each outcome includes two columns, the first of which (left column) reflects results from the school, and student fixed effects model, and the second of which (right column) reflects results from the lagged dependent model with school fixed effects. The lagged outcome used for the second approach is the student's ELA, Math, attendance rate, and number of disciplinary incidents during the previous year. Both models include student demographics as controls: gender, race, SPED classification, disability type, English Learner status, and free or reduced meal eligibility. For non-test outcomes, we also add grade fixed effects for all the models. Standard errors are clustered at the student level and reported in parentheses.

	English Language Arts		Mathematics		Attendance Rate		Disciplinary Incidents	
<i>Co-taught vs. Pull-out</i>								
Co-taught	0.044 (0.029)	0.038 (0.047)	0.031 (0.038)	-0.011 (0.054)	0.003 (0.005)	0.007 (0.005)	-0.002 (0.109)	0.052 (0.087)
Co-taught × Elementary	-0.080 (0.049)	-0.070 (0.087)	0.069 [†] (0.065)	0.059 (0.110)	-0.002 (0.005)	-0.002 (0.006)	-0.148 [†] (0.137)	-0.100 (0.126)
Co-taught × Middle					-0.005 (0.006)	-0.005 (0.006)	-0.115 (0.192)	-0.091 (0.173)
Observations	2,175	1,343	1,515	948	12,216	10,270	12,278	10,323
R^2	0.788	0.518	0.801	0.550	0.639	0.385	0.641	0.368
<i>Co-taught vs. Self-contained</i>								
Co-taught	0.067 (0.061)	0.181* (0.106)	0.192** (0.081)	0.159 (0.127)	0.001 (0.004)	0.002 (0.004)	-0.029 (0.084)	-0.038 (0.084)
Co-taught × Elementary	-0.072 (0.090)	-0.143 (0.159)	-0.234* (0.120)	-0.226 (0.188)	-0.000 (0.005)	-0.003 (0.006)	0.090 (0.104)	0.088 (0.109)
Co-taught × Middle					-0.002 (0.005)	-0.003 (0.005)	0.193 (0.131)	0.082 (0.127)
Observations	718	361	692	371	17,330	14,168	17,405	14,235
R^2	0.759	0.510	0.694	0.478	0.599	0.383	0.635	0.368
Lagged Outcome	No	Yes	No	Yes	No	Yes	No	Yes
Student FE	Yes	No	Yes	No	Yes	No	Yes	No
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grade FE	No	No	No	No	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; $H_0 : \beta = 0$

[†] $p < 0.10$, ^{††} $p < 0.05$, ^{†††} $p < 0.01$; $H_0 : \beta + \lambda = 0$

Notes: This table presents the effect of attending a co-taught classroom versus a pull-out (Top) and self-contained (Bottom) setting on students' standardized ELA and Math scores, attendance rate, and the number of disciplinary incidents. The sample is restricted to students who were ever classified with Other Health Impairment and includes students in grades 3–8 for test scores and grades 1–12 for non-test outcomes, with students observed in each service delivery model pair (e.g., co-taught vs. pull-out) at least once. For test scores, “Co-taught” equals 1 if a student was in a co-taught classroom for that subject. For non-test outcomes, “Co-taught” equals 1 if the student received instruction for ELA, Mathematics, Social Studies, and Science in co-taught classrooms. “Elementary” equals 1 for students below sixth grade, and “Middle” equals 1 for grades 6–8. Each outcome includes two columns, the first of which (left column) reflects results from the school, and student fixed effects model, and the second of which (right column) reflects results from the lagged dependent model with school fixed effects. The lagged outcome used for the second approach is the student's ELA, Math, attendance rate, and number of disciplinary incidents during the previous year. Both models include student demographics as controls: gender, race, SPED classification, disability type, English Learner status, and free or reduced meal eligibility. For non-test outcomes, we also add grade fixed effects for all the models. Standard errors are clustered at the student level and reported in parentheses.

Table C4: Service Delivery Model Student-School Fixed Effects. Students with Disabilities

	English Language Arts		Mathematics		Attendance Rate		Disciplinary Incidents			
<i>General education vs. Co-taught</i>										
GENED	0.005 (0.006)	0.012 (0.009)	0.006 (0.006)	-0.014* (0.008)	0.001 (0.011)	-0.015* (0.008)	0.003*** (0.001)	0.004*** (0.001)	-0.033*** (0.012)	-0.102*** (0.013)
GENED × Elementary	-0.018*** (0.009)	-0.023* (0.014)	-0.019*** (0.009)	-0.005*** (0.010)	-0.003 (0.015)	-0.004*** (0.010)	-0.004*** (0.001)	-0.003*** (0.001)	0.048*** (0.014)	0.115*** (0.015)
GENED × Middle							-0.004*** (0.001)	-0.004*** (0.001)	-0.072*** (0.019)	-0.045*** (0.021)
Observations	51,150	32,199	50,485	46,859	28,817	46,226	498,278	407,372	500,211	414,772
R ²	0.849	0.615	0.853	0.846	0.607	0.850	0.549	0.357	0.540	0.279
<i>General education vs. Pull-out</i>										
GENED	0.009 (0.007)	0.009 (0.010)	0.009 (0.007)	-0.003 (0.009)	-0.012 (0.013)	-0.003 (0.009)	0.005*** (0.001)	0.006*** (0.001)	-0.053*** (0.017)	-0.121*** (0.018)
GENED × Elementary	0.003 (0.012)	0.038*** (0.018)	0.005 (0.012)	0.063*** (0.014)	0.085*** (0.022)	0.064*** (0.015)	-0.005 (0.001)	-0.005*** (0.001)	0.021 (0.028)	0.049*** (0.028)
GENED × Middle							-0.003** (0.001)	-0.003*** (0.001)	-0.169*** (0.033)	-0.172*** (0.035)
Observations	37,449	23,686	37,063	26,335	16,623	26,096	216,793	177,747	217,480	182,300
R ²	0.818	0.549	0.821	0.817	0.567	0.820	0.537	0.353	0.540	0.694
Lagged Outcome	No	Yes	No	No	Yes	No	No	Yes	No	No
Student FE	Yes	No	No	Yes	No	No	Yes	No	Yes	No
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Student-School FE	No	No	Yes	No	No	Yes	No	No	No	No
Grade FE	No	No	No	No	No	No	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; $H_0: \beta = 0$

† $p < 0.10$, †† $p < 0.05$, ††† $p < 0.01$; $H_0: \beta + \lambda = 0$

Notes: This table presents the effect of attending a general education (GENED) classroom versus a co-taught (Panel A) or pull-out (Panel B) setting on SWDs' standardized ELA and Math scores, attendance rate, and the number of disciplinary incidents. The sample includes SWDs in grades 3–8 for test scores and grades 1–12 for non-test outcomes, with students observed in each service delivery model pair (e.g., GENED vs. co-taught) at least once. For test scores, "GENED" equals 1 if a student was in a GENED classroom for that subject. For non-test outcomes, "GENED" equals 1 if the student received instruction for ELA, Mathematics, Social Studies, and Science in GENED classrooms. "Elementary" equals 1 for students below sixth grade, and "Middle" equals 1 for grades 6–8. We used three modeling approaches: (1) school and student fixed effects, (2) school and student-school fixed effects, and (3) lagged outcomes and school fixed effects. For non-test outcomes, we also add grade fixed effects for all the models. The lagged outcome for the third approach is the student's previous year's attendance rate, number of disciplinary incidents, or ELA/Mathematics test scores. All models control for student demographics: gender, race, SPED classification, disability type, English Learner status, and free or reduced lunch status. Standard errors are reported in parenthesis and clustered at the student level for models (1) and (3) and at the student-school level for model (2).

	English Language Arts		Mathematics		Attendance Rate		Disciplinary Incidents		
<i>General education vs. Self-contained</i>									
GENED	0.058*** (0.015)	0.069*** (0.024)	0.104*** (0.018)	0.138*** (0.026)	0.103*** (0.018)	0.001 (0.001)	-0.022 (0.016)	-0.095*** (0.014)	-0.031 (0.020)
GENED × Elementary	-0.008†† (0.027)	-0.023 (0.042)	-0.041 (0.027)	-0.038 (0.039)	-0.040 (0.027)	0.002††† (0.001)	-0.000 (0.001)	0.062***†† (0.022)	0.115*** (0.021)
GENED × Middle				0.003** (0.001)	0.003** (0.001)	-0.002*†† (0.001)	0.001 (0.024)	0.013††† (0.024)	0.096***†† (0.034)
Observations	8,458	5,229	10,076	6,303	10,024	261,659	206,372	215,461	262,640
R ²	0.779	0.515	0.742	0.466	0.743	0.539	0.358	0.692	0.302
									0.703
<i>Pull-out vs. Self-contained</i>									
Pull-out	0.039** (0.018)	0.028 (0.029)	0.034 (0.025)	0.069* (0.039)	0.034 (0.025)	-0.006*** (0.002)	-0.008*** (0.002)	0.031 (0.041)	0.014 (0.036)
Pull-out × Elementary	-0.042 (0.026)	-0.035 (0.042)	-0.008 (0.035)	-0.035 (0.054)	-0.007 (0.035)	0.007*** (0.002)	0.007** (0.003)	0.121***††† (0.058)	0.178***††† (0.065)
Pull-out × Middle				0.004 (0.003)	0.004 (0.003)	0.000 (0.003)	0.000 (0.003)	0.489***††† (0.083)	0.595***††† (0.099)
Observations	8,096	4,718	5,838	3,466	5,820	62,114	50,362	47,274	62,280
R ²	0.782	0.519	0.783	0.474	0.759	0.631	0.373	0.711	0.647
Lagged Outcome	No	Yes	No	Yes	No	No	Yes	No	Yes
Student FE	Yes	No	Yes	No	No	Yes	No	Yes	No
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Student-School FE	No	No	No	No	Yes	No	No	No	Yes
Grade FE	No	No	No	No	No	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; $H_0: \beta = 0$

† $p < 0.10$, †† $p < 0.05$, ††† $p < 0.01$; $H_0: \beta + \lambda = 0$

Notes: This table presents the effect of attending a GENED (Top) and a pull-out (Bottom) classroom versus a self-contained one on SWDs' standardized ELA and Math scores, attendance rate, and the number of disciplinary incidents. The sample includes SWDs in grades 3–8 for test scores and grades 1–12 for non-test outcomes, with students observed in each service delivery model pair (e.g., GENED vs. self-contained) at least once. For test scores, "GENED" equals 1 if a student was in a GENED classroom for that subject. For non-test outcomes, "GENED" equals 1 if the student received instruction for ELA, Mathematics, Social Studies, and Science in GENED classrooms. "Elementary" equals 1 for students below sixth grade, and "Middle" equals 1 for grades 6–8. We used three modeling approaches: (1) school and student fixed effects, (2) school and student-school fixed effects, and (3) lagged outcomes and school fixed effects. For non-test outcomes, we also add grade fixed effects for all the models. The lagged outcome for the third approach is the student's previous year's attendance rate, number of disciplinary incidents, or ELA/Mathematics test scores. All models control for student demographics: gender, race, SPED classification, disability type, English Learner status, and free or reduced lunch status. Standard errors are reported in parenthesis and clustered at the student level for models (1) and (3) and at the student-school level for model (2).

	English Language Arts		Mathematics		Attendance Rate		Disciplinary Incidents			
<i>Co-taught vs. Pull-out</i>										
Co-taught	0.035*** (0.012)	0.022 (0.019)	0.051*** (0.017)	0.020 (0.026)	0.051*** (0.017)	0.007*** (0.002)	0.005*** (0.002)	0.003 (0.037)	0.026 (0.034)	-0.035 (0.039)
Co-taught × Elementary	-0.013 (0.022)	-0.023 (0.038)	0.031††† (0.030)	0.016 (0.051)	0.030††† (0.030)	-0.005* (0.002)	-0.004 (0.003)	-0.086*†† (0.050)	-0.086*† (0.047)	-0.031†† (0.050)
Co-taught × Middle						-0.003 (0.003)	-0.003 (0.003)	-0.086 (0.071)	-0.062 (0.070)	0.029 (0.083)
Observations	10,636	6,754	10,614	4,321	6,847	49,429	38,213	58,601	49,669	38,408
R ²	0.806	0.515	0.807	0.528	0.806	0.363	0.709	0.612	0.302	0.720
<i>Co-taught vs. Self-contained</i>										
Co-taught	0.071*** (0.021)	0.076** (0.034)	0.071*** (0.021)	0.226*** (0.040)	0.215*** (0.028)	0.004*** (0.001)	0.003** (0.001)	-0.009 (0.028)	-0.039 (0.030)	-0.070** (0.031)
Co-taught × Elementary	-0.036 (0.037)	0.041†† (0.058)	-0.034 (0.037)	-0.172*** (0.044)	-0.173*** (0.044)	-0.002 (0.002)	-0.002 (0.002)	0.017 (0.038)	0.063 (0.039)	0.057 (0.039)
Co-taught × Middle						-0.002 (0.002)	-0.002 (0.002)	0.085*†† (0.047)	0.080 (0.049)	0.109* (0.058)
Observations	4,609	2,801	4,591	2,592	4,264	78,426	70,735	97,557	78,825	71,074
R ²	0.759	0.445	0.760	0.440	0.731	0.361	0.706	0.617	0.334	0.741
Lagged Outcome	No	Yes	No	Yes	No	Yes	No	No	Yes	No
Student FE	Yes	No	No	No	No	No	No	Yes	No	No
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Student-School FE	No	No	Yes	No	Yes	No	Yes	No	No	Yes
Grade FE	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; $H_0 : \beta = 0$

† $p < 0.10$, †† $p < 0.05$, ††† $p < 0.01$; $H_0 : \beta + \lambda = 0$

Notes: This table presents the effect of attending a co-taught classroom versus a pull-out (Top) and self-contained (Bottom) setting on SWDs' standardized ELA and Math scores, attendance rate, and the number of disciplinary incidents. The sample includes SWDs in grades 3–8 for test scores and grades 1–12 for non-test outcomes, with students observed in each service delivery model pair (e.g., co-taught vs. pull-out) at least once. For test scores, “co-taught” equals 1 if a student was in a co-taught classroom for that subject. For non-test outcomes, “co-taught” equals 1 if the student received instruction for ELA, Mathematics, Social Studies, and Science in co-taught classrooms. “Elementary” equals 1 for students below sixth grade, and “Middle” equals 1 for grades 6–8. We used three modeling approaches: (1) school and student fixed effects, (2) school and student-school fixed effects, and (3) lagged outcomes and school fixed effects. For non-test outcomes, we also add grade fixed effects for all the models. The lagged outcome for the third approach is the student's previous year's attendance rate, number of disciplinary incidents, or ELA/Mathematics test scores. All models control for student demographics: gender, race, SPED classification, disability type, English Learner status, and free or reduced lunch status. Standard errors are reported in parenthesis and clustered at the student level for models (1) and (3) and at the student-school level for model (2).

Table C5: Service Delivery Model Effects. Students Without Disabilities

	English Language Arts		Mathematics		Attendance Rate		Disciplinary Incidents					
<i>General education vs. Co-taught</i>												
GENED	-0.006** (0.002)	0.003 (0.004)	-0.007*** (0.002)	-0.021*** (0.003)	-0.001 (0.004)	-0.021*** (0.003)	0.003*** (0.000)	0.005*** (0.000)	-0.001*** (0.000)	-0.032*** (0.006)	-0.111*** (0.006)	-0.005 (0.006)
GENED × Elementary	-0.006*** (0.003)	-0.022*** (0.005)	-0.005*** (0.003)	0.004*** (0.004)	-0.024*** (0.006)	0.005*** (0.004)	-0.004*** (0.000)	-0.003*** (0.000)	0.001*** (0.000)	0.040*** (0.007)	0.116*** (0.006)	0.014*** (0.007)
GENED × Middle							-0.002*** (0.000)	-0.003*** (0.000)	0.002*** (0.000)	0.015*** (0.008)	0.072*** (0.008)	-0.002 (0.008)
Observations	332,336	195,304	329,847	275,386	158,668	273,596	2,134,591	1,713,082	1,843,861	2,142,583	1,720,661	1,850,983
R ²	0.844	0.617	0.850	0.858	0.664	0.863	0.521	0.359	0.676	0.479	0.240	0.664
Lagged Outcome	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes	No
Student FE	Yes	No	No	Yes	No	No	Yes	No	No	Yes	No	No
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Student-School FE	No	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes
Grade FE	No	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; $H_0 : \beta = 0$

† $p < 0.10$, †† $p < 0.05$, ††† $p < 0.01$; $H_0 : \beta + \lambda = 0$

Notes: This table presents the effect of attending a general education (GENED) versus a co-taught classroom on standardized ELA and Math scores, attendance rate, and the number of disciplinary incidents for students without disabilities. The sample is restricted to students enrolled in grades 3–8 for test scores and 1–12 for non-test outcomes that we observe in a GENED and co-taught classroom at least once. For test scores, “GENED” equals 1 if a student was in that type of classroom in the subject area. For non-test outcomes, “GENED” equals 1 if the student took ELA, Mathematics, Social Studies, and Science in GENED classrooms. “Elementary” takes a value of 1 if a student is below grade six. “Middle” equals one if the student is above grade five but below grade 9. We used three modeling approaches: (1) school and student fixed effects, (2) school and student-school fixed effects, and (3) lagged outcomes and school fixed effects. For non-test outcomes, we also add grade fixed effects for all the models. The lagged outcome for the third approach is the student’s previous year’s attendance rate, number of disciplinary incidents, or ELA/Mathematics test scores. All models control for student demographics: gender, race, SPED classification, disability type, English Learner status, and free or reduced lunch status. Standard errors are reported in parenthesis and clustered at the student level for models (1) and (3) and at the student-school level for model (2).

Table C6: Service Delivery Model Effects. Students with Disabilities (Same students)

	English Language Arts		Mathematics		Attendance Rate		Disciplinary Incidents			
<i>General education vs. Co-taught</i>										
GENED	0.005 (0.006)	0.012 (0.009)	-0.003 (0.010)	-0.014* (0.008)	0.001 (0.011)	0.003 (0.012)	0.003*** (0.001)	0.004*** (0.001)	-0.033*** (0.012)	-0.102*** (0.013)
GENED × Elementary	-0.018***†† (0.009)	-0.023* (0.014)	0.001 (0.015)	-0.005††† (0.010)	-0.003 (0.015)	-0.021 (0.017)	-0.004***††† (0.001)	-0.003*** (0.001)	0.048***†† (0.014)	0.115***† (0.014)
GENED × Middle							-0.004*** (0.001)	-0.004*** (0.001)	-0.072***††† (0.019)	-0.069***††† (0.020)
Observations	51,150	32,199	22,616	46,859	28,817	19,396	498,278	407,372	500,211	409,276
R ²	0.849	0.615	0.862	0.846	0.607	0.859	0.549	0.357	0.540	0.279
							388,150	0.565		
<i>General education vs. Pull-out</i>										
GENED	0.009 (0.007)	0.009 (0.010)	0.001 (0.011)	-0.003 (0.009)	-0.012 (0.013)	0.001 (0.015)	0.005*** (0.001)	0.006*** (0.001)	-0.053*** (0.017)	-0.121*** (0.018)
GENED × Elementary	0.003 (0.012)	0.038***†† (0.018)	0.038***†† (0.020)	0.063***††† (0.014)	0.085***††† (0.022)	0.072***††† (0.025)	-0.005*** (0.001)	-0.005***†† (0.001)	0.021 (0.028)	0.049***†† (0.029)
GENED × Middle							-0.003** (0.001)	-0.003***††† (0.001)	-0.169***††† (0.033)	-0.155***††† (0.036)
Observations	37,449	23,686	15,878	26,335	16,623	10,852	216,793	177,747	217,480	178,488
R ²	0.818	0.549	0.838	0.817	0.567	0.836	0.537	0.353	0.540	0.290
Lagged Student FE	No	Yes	No	No	Yes	No	No	Yes	No	Yes
School FE	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes	No
Grade FE	No	Yes	No	No	Yes	No	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; $H_0: \beta = 0$

† $p < 0.10$, †† $p < 0.05$, ††† $p < 0.01$; $H_0: \beta + \lambda = 0$

Notes: This table presents the effect of attending a general education (GENED) classroom versus a co-taught (Panel A) or pull-out (Panel B) setting on SWDs' standardized ELA and Math scores, attendance rate, and the number of disciplinary incidents. The sample includes SWDs in grades 3–8 for test scores and grades 1–12 for non-test outcomes, with students observed in each service delivery model pair (e.g., GENED vs. co-taught) at least once. For test scores, "GENED" equals 1 if a student was in a GENED classroom for that subject. For non-test outcomes, "GENED" equals 1 if the student received instruction for ELA, Mathematics, Social Studies, and Science in GENED classrooms. "Elementary" equals 1 for students below sixth grade, and "Middle" equals 1 for grades 6–8. For each outcome, we present three estimates: (1) school and student fixed effects, (2) lagged outcomes with school fixed effects, and (3) school and student fixed effects but restricted to the sample of students included in the lagged outcome model (column 2). For non-test outcomes, we also add grade fixed effects for all the models. The lagged outcome used for the second approach is the student's previous year's ELA, Math, attendance rate, and number of disciplinary incidents during the previous year. All models control for student demographics: gender, race, SPED classification, disability type, English Learner status, and free or reduced lunch status. Standard errors are clustered at the student level and reported in parentheses.

	English Language Arts		Mathematics		Attendance Rate		Disciplinary Incidents				
<i>General education vs. Self-contained</i>											
GENED	0.058*** (0.015)	0.069*** (0.024)	0.079*** (0.028)	0.104*** (0.018)	0.138*** (0.026)	0.092*** (0.031)	-0.003*** (0.001)	0.001 (0.001)	-0.022 (0.016)	-0.095*** (0.014)	-0.029 (0.018)
GENED × Elementary	-0.008†† (0.027)	-0.023 (0.042)	-0.053 (0.049)	-0.041 (0.027)	-0.038 (0.039)	0.011 (0.047)	0.006***†† (0.001)	0.002††† (0.001)	0.062***†† (0.022)	0.115*** (0.021)	0.056** (0.028)
GENED × Middle							0.003** (0.001)	0.001††† (0.001)	0.001 (0.024)	0.013††† (0.024)	-0.004 (0.027)
Observations	8,458	5,229	3,271	10,076	6,303	4,037	261,659	206,372	262,640	207,334	200,561
R ²	0.779	0.515	0.809	0.742	0.466	0.772	0.539	0.358	0.548	0.302	0.567
<i>Pull-out vs. Self-contained</i>											
Pull-out	0.039** (0.018)	0.028 (0.029)	0.044 (0.034)	0.034 (0.025)	0.069* (0.039)	0.016 (0.047)	-0.006*** (0.002)	-0.008*** (0.002)	0.031 (0.041)	0.014 (0.036)	-0.038 (0.041)
Pull-out × Elementary	-0.042 (0.026)	-0.035 (0.042)	-0.087* (0.051)	-0.008 (0.035)	-0.035 (0.054)	-0.059 (0.065)	0.007*** (0.002)	0.007** (0.003)	0.121***††† (0.058)	0.178***††† (0.065)	0.164***†† (0.069)
Pull-out × Middle							0.004 (0.003)	0.003†† (0.003)	0.489***††† (0.083)	0.595***††† (0.086)	0.513***††† (0.093)
Observations	8,096	4,718	2,972	5,838	3,466	2,167	62,114	50,362	62,280	50,514	45,939
R ²	0.782	0.519	0.811	0.758	0.474	0.784	0.631	0.373	0.647	0.344	0.659
Lagged Outcome	No	Yes	No	No	Yes	No	No	Yes	No	Yes	No
Student FE	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes	No	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grade FE	No	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; $H_0: \beta = 0$

† $p < 0.10$, †† $p < 0.05$, ††† $p < 0.01$; $H_0: \beta + \lambda = 0$

Notes: This table presents the effect of attending a GENED (Top) and a pull-out (Bottom) classroom versus a self-contained one on SWDs' standardized ELA and Math scores, attendance rate, and the number of disciplinary incidents. The sample includes SWDs in grades 3–8 for test scores and grades 1–12 for non-test outcomes, with students observed in each service delivery model pair (e.g., GENED vs. self-contained) at least once. For test scores, "GENED" equals 1 if a student was in a GENED classroom for that subject. For non-test outcomes, "GENED" equals 1 if the student received instruction for ELA, Mathematics, Social Studies, and Science in GENED classrooms. "Elementary" equals 1 for students below sixth grade, and "Middle" equals 1 for grades 6–8. For each outcome, we present three estimates: (1) school and student fixed effects, (2) lagged outcomes with school fixed effects, and (3) school and student fixed effects but restricted to the sample of students included in the lagged outcome model (column 2). For non-test outcomes, we also add grade fixed effects for all the models. The lagged outcome used for the second approach is the student's previous year's ELA, Math, attendance rate, and number of disciplinary incidents during the previous year. All models control for student demographics: gender, race, SPED classification, disability type, English Learner status, and free or reduced lunch status. Standard errors are clustered at the student level and reported in parentheses.

	English Language Arts		Mathematics		Attendance Rate		Disciplinary Incidents					
<i>Co-taught vs. Pull-out</i>												
Co-taught	0.035*** (0.012)	0.022 (0.019)	-0.004 (0.019)	0.051*** (0.017)	0.020 (0.026)	0.052* (0.027)	0.005*** (0.002)	0.007*** (0.002)	0.003 (0.037)	0.026 (0.034)	0.011 (0.037)	
Co-taught × Elementary	-0.013 (0.022)	-0.023 (0.038)	0.011 (0.038)	0.031††† (0.030)	0.016 (0.051)	0.160***††† (0.052)	-0.004* (0.002)	-0.005* (0.002)	-0.086*†† (0.050)	-0.086*† (0.047)	-0.112***†† (0.055)	
Co-taught × Middle							-0.003 (0.003)	-0.002††† (0.003)	-0.086 (0.071)	-0.062 (0.070)	-0.052 (0.075)	
Observations	10,636	6,754	4,444	6,850	4,321	2,815	58,344	49,429	45,445	58,601	49,669	45,717
R ²	0.806	0.515	0.829	0.806	0.528	0.821	0.614	0.363	0.635	0.612	0.302	0.630
<i>Co-taught vs. Self-contained</i>												
Co-taught	0.071*** (0.021)	0.076** (0.034)	0.054 (0.038)	0.214*** (0.028)	0.226*** (0.040)	0.240*** (0.047)	0.003** (0.001)	0.004*** (0.001)	0.005*** (0.002)	-0.009 (0.028)	-0.039 (0.030)	-0.027 (0.031)
Co-taught × Elementary	-0.036 (0.037)	0.041†† (0.058)	0.054†† (0.066)	-0.172*** (0.044)	-0.171** (0.073)	-0.198** (0.085)	-0.002 (0.002)	-0.006*** (0.002)	-0.005** (0.002)	0.017 (0.038)	0.063 (0.039)	-0.007 (0.044)
Co-taught × Middle							-0.002 (0.002)	-0.004** (0.002)	-0.005** (0.002)	0.085*†† (0.047)	0.080 (0.049)	0.064 (0.050)
Observations	4,609	2,801	1,742	4,273	2,592	1,622	97,135	78,426	74,003	97,557	78,825	74,422
R ²	0.759	0.445	0.783	0.730	0.440	0.735	0.580	0.361	0.607	0.617	0.334	0.633
Lagged Outcome	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes	No
Student FE	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grade FE	No	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; $H_0: \beta = 0$

† $p < 0.10$, †† $p < 0.05$, ††† $p < 0.01$; $H_0: \beta + \lambda = 0$

Notes: This table presents the effect of attending a co-taught classroom versus a pull-out (Top) and self-contained (Bottom) setting on SWDs' standardized ELA and Math scores, attendance rate, and the number of disciplinary incidents. The sample includes SWDs in grades 3–8 for test scores and grades 1–12 for non-test outcomes, with students observed in each service delivery model pair (e.g., co-taught vs. pull-out) at least once. For test scores, "co-taught" equals 1 if a student was in a co-taught classroom for that subject. For non-test outcomes, "co-taught" equals 1 if the student received instruction for ELA, Mathematics, Social Studies, and Science in co-taught classrooms. "Elementary" equals 1 for students below sixth grade, and "Middle" equals 1 for grades 6–8. For each outcome, we present three estimates: (1) school and student fixed effects, (2) lagged outcomes with school fixed effects, and (3) school and student fixed effects but restricted to the sample of students included in the lagged outcome model (column 2). For non-test outcomes, we also add grade fixed effects for all the models. The lagged outcome used for the second approach is the student's previous year's ELA, Math, attendance rate, and number of disciplinary incidents during the previous year. All models control for student demographics: gender, race, SPED classification, disability type, English Learner status, and free or reduced lunch status. Standard errors are clustered at the student level and reported in parentheses.

Table C7: Service Delivery Model Effects. Students without Disabilities (Same students)

	English Language Arts		Mathematics		Attendance Rate		Disciplinary Incidents					
<i>General education vs. Co-taught</i>												
GENED	-0.006** (0.002)	0.003 (0.004)	-0.009** (0.004)	-0.021*** (0.003)	-0.001 (0.004)	-0.019*** (0.005)	0.003*** (0.000)	0.005*** (0.000)	0.002*** (0.000)	-0.032*** (0.006)	-0.111*** (0.006)	-0.029*** (0.006)
GENED × Elementary	-0.006*††† (0.003)	-0.022***††† (0.005)	-0.009*††† (0.005)	0.004††† (0.004)	-0.024***††† (0.006)	0.006††† (0.006)	-0.004***††† (0.000)	-0.003***††† (0.000)	-0.002***††† (0.000)	0.040***††† (0.007)	0.116***††† (0.006)	0.037***††† (0.007)
GENED × Middle							-0.002*** (0.000)	-0.003***††† (0.000)	-0.002*** (0.000)	0.015*††† (0.008)	0.072***††† (0.008)	0.010††† (0.008)
Observations	332,336	195,304	150,575	275,386	158,668	118,595	2,134,591	1,713,082	1,674,936	2,142,583	1,720,661	1,682,210
R ²	0.844	0.617	0.861	0.858	0.664	0.876	0.521	0.359	0.542	0.479	0.240	0.504
Lagged Outcome	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes	No
Student FE	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grade FE	No	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; $H_0: \beta = 0$

† $p < 0.10$, †† $p < 0.05$, ††† $p < 0.01$; $H_0: \beta + \lambda = 0$

Notes: This table presents the effect of attending a general education (GENED) versus a co-taught classroom on standardized ELA and Math scores, attendance rate, and the number of disciplinary incidents for students without disabilities. The sample is restricted to students enrolled in grades 3–8 for test scores and 1–12 for non-test outcomes that we observe in a GENED and co-taught classroom at least once. For test scores, “GENED” equals 1 if a student was in that type of classroom in the subject area. For non-test outcomes, “GENED” equals 1 if the student took ELA, Mathematics, Social Studies, and Science in GENED classrooms. “Elementary” takes a value of 1 if a student is below grade six. “Middle” equals one if the student is above grade five but below grade 9. For each outcome, we present three estimates: (1) school and student fixed effects, (2) lagged outcomes with school fixed effects, and (3) school and student fixed effects but restricted to the sample of students included in the lagged outcome model (column 2). The lagged outcome used for the second approach is the student’s previous year’s ELA, Math, attendance rate, and number of disciplinary incidents during the previous year. All models include student demographics as controls: gender, race, English Learner status, and free and reduced lunch. For non-test outcomes, we also add grade fixed effects for all the models. Standard errors are clustered at the student level and reported in parentheses.

Table C8: Service Delivery Model Effects by Move Types. Students with Disabilities

	English Language Arts		Mathematics		Attendance Rate		Disciplinary Incidents					
<i>General education vs. Co-taught</i>												
GENED	0.086*** (0.020)	-0.042* (0.023)	-0.035** (0.015)	0.154** (0.025)	-0.172*** (0.028)	-0.019 (0.019)	-0.009** (0.004)	0.015*** (0.002)	-0.002** (0.001)	-0.225*** (0.064)	0.074*** (0.028)	-0.017 (0.018)
GENED × Elementary	-0.025†† (0.033)	0.026 (0.030)	0.015 (0.024)	-0.150*** (0.036)	0.129***†† (0.033)	0.006 (0.027)	0.012*** (0.004)	-0.017***†† (0.002)	0.002* (0.001)	0.237*** (0.076)	-0.076** (0.032)	0.006 (0.021)
GENED × Middle							0.013*** (0.004)	-0.016*** (0.002)	0.002 (0.001)	-0.064††† (0.103)	-0.163***††† (0.041)	-0.052***†† (0.029)
Observations	6.833	8.195	8.234	5.641	7.588	7.118	41.116	114.847	130.520	41.434	115.321	130.866
R ²	0.873	0.874	0.854	0.880	0.863	0.849	0.730	0.656	0.653	0.727	0.649	0.646
<i>General education vs. Pull-out</i>												
GENED	0.097*** (0.025)	-0.042* (0.024)	-0.009 (0.016)	0.173*** (0.036)	-0.143*** (0.032)	0.025 (0.022)	-0.012 (0.012)	0.013*** (0.003)	0.001 (0.001)	-0.199 (0.204)	-0.004 (0.045)	-0.019 (0.024)
GENED × Elementary	-0.028† (0.044)	0.058 (0.038)	0.050 (0.031)	-0.053†† (0.061)	0.246***††† (0.045)	0.004 (0.038)	0.015 (0.013)	-0.013*** (0.003)	-0.001 (0.002)	0.180 (0.235)	-0.166* (0.089)	-0.013 (0.040)
GENED × Middle							0.016 (0.014)	-0.002††† (0.004)	-0.001 (0.002)	0.294 (0.267)	-0.266***††† (0.087)	-0.090†† (0.056)
Observations	5.354	5.425	5.920	3.130	4.130	3.809	19.194	40.883	67.809	19.263	41.039	67.998
R ²	0.845	0.848	0.826	0.848	0.841	0.826	0.735	0.685	0.646	0.744	0.700	0.648
Move	Less restrictive	More restrictive	Both moves	Less restrictive	More restrictive	Both moves	Less restrictive	More restrictive	Both moves	Less restrictive	More restrictive	Both moves
Student FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grade FE	No	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; $H_0: \beta = 0$

† $p < 0.10$, †† $p < 0.05$, ††† $p < 0.01$; $H_0: \beta + \lambda = 0$

Notes: This table presents the effect of attending a general education (GENED) classroom versus a co-taught (top panel) and pull-out (bottom panel) setting on SWDs' standardized ELA and Math scores, attendance rate, and the number of disciplinary incidents. The sample includes SWDs in grades 3–8 for test scores and grades 1–12 for non-test outcomes, with students observed in each service delivery model pair (e.g., GENED vs. co-taught) at least once. For test scores, "GENED" equals 1 if a student was in a GENED classroom for that subject. For non-test outcomes, "GENED" equals 1 if the student received instruction for in GENED classrooms for all subjects (ELA, Mathematics, Social Studies, and Science). "Elementary" equals 1 for students below sixth grade, and "Middle" equals 1 for students in grades 6–8. Each outcome includes three columns, the first of which (left column) reflects results from restricting the samples to students who exclusively transitioned from more restrictive to less restrictive settings (e.g., from co-taught to GENED). The second column restricts the sample to students who exclusively moved from less to more restrictive settings (e.g., from GENED to pull-out). The last column includes students who made a combination of more and less restrictive moves. This sampling method is modeled after Anderson (2021). All models include school and student fixed effects, as well as student demographics as controls: gender, race, SPED classification, disability type, English Learner status, and free or reduced meal eligibility. For non-test outcomes, we also add grade fixed effects. Standard errors are clustered at the student level and reported in parentheses.

	English Language Arts		Mathematics		Attendance Rate		Disciplinary Incidents					
<i>General education vs. Self-contained</i>												
GENED	0.124** (0.051)	-0.103 (0.101)	0.084* (0.046)	0.306** (0.062)	-0.133** (0.063)	0.074 (0.058)	-0.015** (0.006)	0.009** (0.003)	0.001 (0.001)	-0.600** (0.156)	0.112** (0.046)	-0.052* (0.027)
GENED × Elementary	0.024†† (0.083)	0.156 (0.134)	-0.112 (0.079)	-0.067††† (0.108)	0.190** (0.081)	-0.026 (0.088)	0.015** (0.007)	-0.005†† (0.003)	0.000 (0.002)	0.634** (0.171)	-0.135** (0.063)	0.046 (0.038)
GENED × Middle							0.022*** (0.008)	-0.002††† (0.003)	-0.002 (0.002)	0.509** (0.187)	-0.169** (0.075)	0.088* (0.046)
Observations	1,568	696	1,099	1,498	1,611	1,236	24,780	49,537	75,734	24,899	49,784	75,944
R ²	0.841	0.823	0.774	0.779	0.775	0.758	0.749	0.676	0.652	0.744	0.686	0.656
<i>Pull-out vs. Self-contained</i>												
Pull-out	0.179** (0.072)	-0.005 ()	-0.052 (0.055)	0.076 (0.115)	-0.019 (0.069)	0.200** (0.090)	-0.044 (0.069)	-0.003 (0.004)	-0.004 (0.003)	-0.656 (0.767)	-0.004 (0.065)	-0.080 (0.060)
Pull-out × Elementary	-0.316*** (0.113)	-0.010 ()	0.035 (0.081)	-0.099 (0.160)	-0.097† (0.092)	-0.187 (0.122)	0.046 (0.091)	0.003 (0.005)	0.009**†† (0.004)	-0.284 (1.516)	0.177 (0.112)	0.180* (0.094)
Pull-out × Middle							0.030 (0.080)	0.003 (0.006)	0.002 (0.005)	0.615 (1.507)	0.542*** (0.147)	0.501***††† (0.165)
Observations	909	1,273	963	410	1,178	683	847	19,470	18,026	847	19,539	18,048
R ²	0.791	0.823	0.793	0.824	0.779	0.782	0.800	0.716	0.701	0.717	0.721	0.701
Move	Less restrictive	More restrictive	Both moves	Less restrictive	More restrictive	Both moves	Less restrictive	More restrictive	Both moves	Less restrictive	More restrictive	Both moves
Student FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grade FE	No	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; $H_0: \beta = 0$

† $p < 0.10$, †† $p < 0.05$, ††† $p < 0.01$; $H_0: \beta + \lambda = 0$

Notes: This table presents the effect of attending a general education (GENED) classroom (top panel) and a pull-out setting (bottom panel) versus a self-contained one on SWDs' standardized ELA and Math scores, attendance rate, and the number of disciplinary incidents. The sample includes SWDs in grades 3–8 for test scores and grades 1–12 for non-test outcomes, with students observed in each service delivery model pair (e.g., GENED vs. self-contained) at least once. For test scores, "GENED" equals 1 if a student was in a GENED classroom for that subject. For non-test outcomes, "GENED" equals 1 if the student received instruction for in GENED classrooms for all subjects (ELA, Mathematics, Social Studies, and Science). "Elementary" equals 1 for students below sixth grade, and "Middle" equals 1 for grades 6–8. Each outcome includes three columns, the first of which (left column) reflects results from restricting the samples to students who exclusively transitioned from more restrictive to less restrictive settings (e.g., from self-contained to GENED). The second column restricts the sample to students who exclusively moved from less to more restrictive settings (e.g., from pull-out to self-contained). The last column includes students who made a combination of more and less restrictive moves. This sampling method is modeled after Anderson (2021). All models include school and student fixed effects, as well as student demographics as controls: gender, race, SPED classification, disability type, English Learner status, and free or reduced meal eligibility. For non-test outcomes, we also add grade fixed effects. Standard errors are clustered at the student level and reported in parentheses.

	English Language Arts			Mathematics			Attendance Rate			Disciplinary Incidents			
<i>Co-taught vs. Pull-out</i>													
Co-taught	0.048 (0.042)	-0.131*** (0.043)	0.019 (0.032)	0.151** (0.062)	-0.092 (0.060)	0.101* (0.051)	0.031 (0.051)	0.011** (0.005)	-0.666 (0.464)	0.004 (0.003)	0.000 (0.000)	-0.088 (0.080)	0.030 (0.044)
Co-taught × Elementary	0.039† (0.066)	0.148 (0.108)	-0.087 (0.054)	0.008†† (0.088)	0.408***†† (0.141)	0.088††† (0.087)	-0.028 (0.052)	-0.010* (0.006)	0.639 (0.520)	-0.006* (0.003)	-0.006* (0.003)	-0.082†† (0.106)	-0.069 (0.067)
Co-taught × Middle							-0.041 (0.053)	-0.002†† (0.006)	1.356** (0.637)	-0.001 (0.004)	0.006 (0.137)	0.006 (0.106)	0.111 (0.106)
Observations	1,731	903	1,653	971	604	944	1,939	14,375	1,944	19,993	14,483	20,074	20,074
R ²	0.832	0.852	0.820	0.843	0.865	0.782	0.777	0.714	0.791	0.689	0.710	0.688	0.688
<i>Co-taught vs. Self-contained</i>													
Co-taught	0.175*** (0.068)	-0.064 (0.124)	0.106 (0.065)	0.324*** (0.091)	-0.121 (0.134)	0.338*** (0.081)	0.001 (0.007)	0.014*** (0.004)	-0.176 (0.188)	0.003 (0.002)	0.060 (0.062)	-0.072* (0.040)	-0.072* (0.040)
Co-taught × Elementary	-0.173 (0.110)	0.167 (0.195)	0.008 (0.103)	-0.209 (0.145)	-0.067 (0.236)	-0.216* (0.124)	0.009†† (0.009)	-0.010** (0.005)	0.115 (0.219)	-0.004 (0.003)	-0.139* (0.079)	0.068 (0.058)	0.068 (0.058)
Co-taught × Middle							-0.013† (0.010)	-0.010** (0.005)	0.774***†† (0.344)	-0.002 (0.003)	-0.064 (0.093)	0.114 (0.075)	0.114 (0.075)
Observations	876	333	546	711	363	534	4,580	23,504	4,611	31,689	23,663	31,809	31,809
R ²	0.791	0.793	0.778	0.728	0.762	0.761	0.791	0.678	0.801	0.685	0.711	0.695	0.695
Move	Less restrictive	More restrictive	Both moves	Less restrictive	More restrictive	Both moves	Less restrictive	More restrictive	Less restrictive	Both moves	More restrictive	Both moves	Both moves
Student FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grade FE	No	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; $H_0: \beta = 0$

† $p < 0.10$, †† $p < 0.05$, ††† $p < 0.01$; $H_0: \beta + \lambda = 0$

Notes: This table presents the effect of attending a co-taught classroom versus a pull-out (top panel) and a self-contained (bottom panel) setting on SWDs' standardized ELA and Math scores, attendance rate, and the number of disciplinary incidents. The sample includes SWDs in grades 3–8 for test scores and grades 1–12 for non-test outcomes, with students observed in each service delivery model pair (e.g., co-taught vs. pull-out) at least once. For test scores, "GENED" equals 1 if a student was in a GENED classroom for that subject. For non-test outcomes, "GENED" equals 1 if the student received instruction for in GENED classrooms for all subjects (ELA, Mathematics, Social Studies, and Science). "Elementary" equals 1 for students below sixth grade, and "Middle" equals 1 for grades 6–8. Each outcome includes three columns, the first of which (left column) reflects results from restricting the samples to students who exclusively transitioned from more restrictive to less restrictive settings (e.g., from pull-out to co-taught). The second column restricts the sample to students who exclusively moved from less to more restrictive settings (e.g., from co-taught to self-contained). The last column includes students who made a combination of more and less restrictive moves. All models include school and student fixed effects, as well as student demographics as controls: gender, race, SPED classification, disability type, English Learner status, and free or reduced meal eligibility. For non-test outcomes, we also add grade fixed effects. Standard errors are clustered at the student level and reported in parentheses.

Table C9: Service Delivery Model Effects by Move Types. Students without Disabilities

	English Language Arts		Mathematics		Attendance Rate		Disciplinary Incidents		
<i>General education vs. Co-taught</i>									
GENED	0.012 (0.010)	-0.019*** (0.007)	0.240*** (0.012)	-0.271*** (0.009)	0.067*** (0.007)	-0.013*** (0.002)	0.012*** (0.002)	-0.151*** (0.031)	0.097*** (0.026)
GENED × Elementary	-0.029**† (0.013)	-0.005††† (0.009)	-0.250*** (0.015)	0.228***††† (0.010)	-0.063*** (0.009)	0.013*** (0.003)	-0.013***††† (0.002)	0.151*** (0.033)	-0.080***†† (0.027)
GENED × Middle						0.014*** (0.003)	-0.014***††† (0.002)	0.113*** (0.037)	-0.103*** (0.031)
Observations	41,495	57,542	54,838	49,437	42,835	269,994	240,337	271,832	240,892
R ²	0.874	0.870	0.863	0.888	0.874	0.683	0.699	0.673	0.633
Move	Less restrictive	More restrictive	Both moves	Less restrictive	More restrictive	Both moves	Less restrictive	More restrictive	Both moves
Student FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grade FE	No	No	No	No	No	No	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; $H_0: \beta = 0$

† $p < 0.10$, †† $p < 0.05$, ††† $p < 0.01$; $H_0: \beta + \lambda = 0$

Notes: This table presents the effect of attending a general education (GENED) classroom versus a co-taught on standardized ELA and Math scores, attendance rate, and the number of disciplinary incidents. The sample includes students without disabilities in grades 3–8 for test scores and grades 1–12 for non-test outcomes, with students observed GENED and co-taught classroom at least once. For test scores, “GENED” equals 1 if a student was in a GENED classroom for that subject. For non-test outcomes, “GENED” equals 1 if the student received instruction for in GENED classrooms for all subjects (ELA, Mathematics, Social Studies, and Science). “Elementary” equals 1 for students below sixth grade and “Middle” equals 1 for grades 6–8. Each outcome includes three columns, the first of which (left column) reflects results from restricting the samples to students who exclusively transitioned from more co-taught to classrooms. The second column restricts the sample to students who exclusively moved from GENED to co-taught settings. The last column includes students who made a combination of both move types. This sampling method is modeled after Anderson (2021). All models include school and student fixed effects, as well as student demographics as controls: gender, race, SPED classification, disability type, English Learner status, and free or reduced meal eligibility. For non-test outcomes, we also add grade fixed effects. Standard errors are clustered at the student level and reported in parentheses.

Suggested Citation: Jones, N., Kaler, L., Markham, J., Senese, J., & Winters, M. A. *Service Delivery Models: Impacts for Students with and without Disabilities* (Working Paper 2025-1). 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

