Research Article

Educational Evaluation and Policy Analysis
Month 202X, Vol. XX, No. X, pp. 1–26
DOI: 10.3102/01623737241257951
Article reuse guidelines: sagepub.com/journals-permissions

Whose IDEA Is This? An Examination of the Effectiveness of Inclusive Education

Katharine Parham Malhotra
Teachers College of Columbia University

The inclusion of students with disabilities in general education versus more restrictive settings has steadily increased since the 1990s. Yet little is known about inclusion's effectiveness for these students or their nondisabled peers. I examine the impacts of a district-wide inclusion policy, leveraging the staggered, school-level implementation to estimate the policy's causal effects on academic and behavioral outcomes. Elementary and middle school test scores and attendance rates were unaffected by the policy. High school graduation and ninth grade promotion rates increased by two and six percentage points, respectively, in the years following implementation. Findings suggest that inclusive education does not come at the expense of students' academic progress in the short term and may improve academic outcomes in the longer term.

Keywords: *special education, inclusion, policy evaluation, students with disabilities*

The Individuals with Disabilities Education Act (IDEA), originally passed in 1975 as the "Education for All Handicapped Children Act," came on the heels of a public awakening to issues of discrimination throughout the 1960s and helped transition U.S. education from two segregated forms of schooling to one in which students with disabilities (SWD) were not considered inherently different (Winzer, 2012). The federal law requires that states develop procedures to ensure that SWD are educated, to the greatest extent appropriate, alongside peers without disabilities in students' least restrictive environment (LRE). The practice of educating students with and without disabilities in the same learning environment has become increasingly prevalent in recent years (see Figure 1), and today more than 60% of all SWD nationwide spend 80% or more of their day in general education environments—up from just 30% in the early 1990s ([NCES], 2019a).

Though growing in popularity over the past three decades, the use of inclusive education is not supported by a robust or coherent evidence base, and even its proponents disagree on the merits. Some supporters argue on largely ideological grounds, seeking course correction from an insidious, segregated history of education for children with special needs (Crockett, 2020; Lindsay, 2007). Others argue inclusive settings benefit students with and without disabilities, emphasizing the cognitive and noncognitive benefits of time spent learning in a "diverse" environment (Peltier, 1997; Salend & Duhaney, 1999; Sanger, 2020). Empirical research on the effectiveness of inclusive settings for students from both subgroups is similarly conflicted. While some evidence suggests that SWD educated in inclusive settings are more likely to make academic progress and graduate on time (Dessemontet et al., 2012; Schifter, 2015), other work has found the impact of educating SWD in general education classrooms to have adverse effects both for them (Daniel & King, 1997) as well as their peers without special needs (Fletcher, 2010; Gottfried, 2014).

This study examines one anonymous U.S. school district that transitioned to a policy of inclusion in general education as the "default"

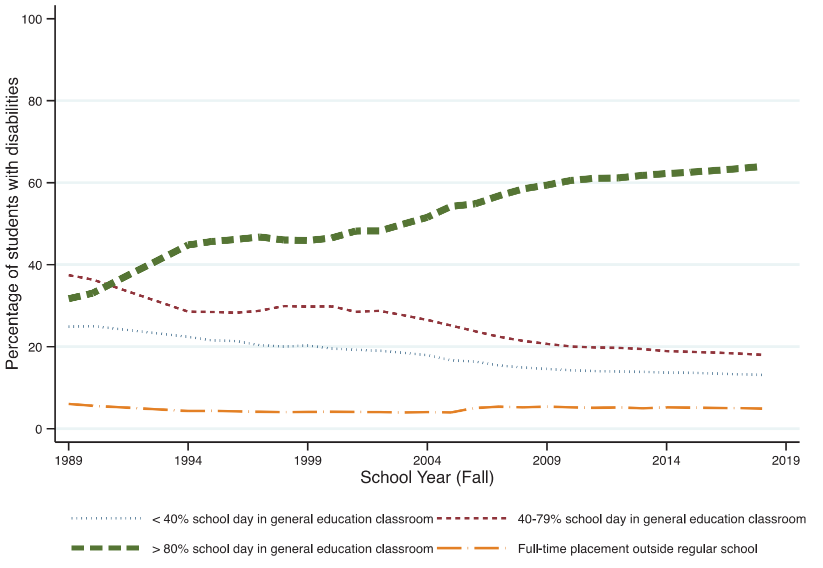

FIGURE 1. *Nationwide prevalence of inclusive education.*
Note. This figure illustrates national educational placement trends among students with disabilities ages 6 to 21 using historical data from the NCES Digest of Education Statistics (2019a). Data are grouped into four categories. The first three categories reflect the percentage of the school day students spend in general education settings across the three federal reporting categories—more than 80%, 40% to 79%, or less than 40%. Prior to 2008, these three reporting categories reflect time spent in general education at more than 60%, 21% to 60%, and less than 21%, respectively. The final category groups all placements where students spend 100% of the school day in a non-public-school setting (e.g., separate school, separate residential facility, private school, homebound, hospital, or correctional facility).

placement for SWD from along the full continuum of alternatives for special education service provision. The policy's implementation over an 8-year period in the early 2000s allows for drawing broader conclusions about one of the most pervasive challenges in education: teaching to meet the diverse, individualized needs of all students within a single classroom. The staggered policy adoption within the district is leveraged in an event study approach to estimate the policy's impacts on the academic and behavioral outcomes of both students with and without disabilities—critical given observed disparities in outcomes across the two subgroups in prior research. Contrary to this evidence base, results from this study show that the introduction of the district's inclusion policy was not associated with any negative impact on students from either group.

Background

What Is Inclusion?

IDEA was most recently reauthorized in 2004 and did not prescribe one path for all children with disabilities, but rather created a process by which a team of individuals who know a child can best determine what is appropriate for the child's education. The four basic provisions of IDEA ensure that, regardless of a child's unique needs (a) they are entitled to an appropriate education at the public expense; (b) a continuum of placements must be available to every student with a disability; (c) every student will be educated in their LRE; and (d) every student with special needs will have an individualized education program (IEP) providing for those needs (IDEA, 2004). The third provision describing the placement of SWD in the appropriate educational environment is the most relevant to the present study. While the law never uses the term "inclusion," advocates and practitioners have interpreted the motivation of the LRE mandate as including as many students as possible in their local community school, inside a regular, grade-level-appropriate classroom for as much of the day as possible (Dorn et al., 1996; Giordano, 2007).

Federal regulations mandate that states monitor the implementation of the LRE provision and annually report the proportion of time school-aged students are educated in the general education classroom across four main categories:[1] (a) more than 80% of the school day; (b) 40% to 79% of the school day; (c) less than 40% of the school day; or (d) all of the school day in a separate setting. The first of the four reporting categories is synonymous with the idea of inclusion, though no federal laws or regulations offer an explicit definition of the term and preferred terminology to describe the same concept has evolved over time.

Prior Research on Inclusion

Inclusion, while ill-defined, is also difficult to rigorously examine. The lack of consistent definition means inclusion may be implemented differently from one context to the next. There are also empirical challenges, consistent with those in the broader literature on special education effectiveness. SWD do not have an obvious comparison group among peers without disabilities, and examining SWD among themselves is limited by issues of selection into special education and the differences across individual students. Some studies have attempted analyses of the causal impacts of special education by examining within-student variation; that is, examining the academic performance of students who enter and exit special education over their educational careers (Hanushek et al., 2002), though this approach remains limited by selection issues. On average, both empirical and observational evidence suggests that students who are identified for and receive special education services have improved test scores (Hanushek et al., 2002; Rea et al., 2002; Schwartz et al., 2019) and long-term educational attainment (Ballis & Heath, 2019), though some evidence using matching methods has found special education generally to have a negative or insignificant impact on identified students' learning and behavior (Morgan et al., 2010).

Inclusion of Students with Disabilities

Research on inclusion as one form of special education service provision, specifically, is often limited to observational methods. On average, extant observational studies suggest that when SWD are included in general education, their outcomes improve even when controlling for peer, school, and district characteristics (McLeskey et al., 2018; Schifter & Hehir, 2018). This is true for both academic and noncognitive outcomes, as evidence of improved test scores is often observed alongside improved work habits, self-confidence, social competence, and attentive behavior (McLeskey et al., 2018). There is also, however, some evidence from older research that inclusive education results in null effects for SWD (Affleck et al., 1988; Jenkins et al., 1991). Two more recent meta-analyses further emphasize a more cautious interpretation of inclusion's positive effects for these students. Ruijs and Peetsma (2009) find inclusion's impacts for SWD to be neutral to positive and broadly comparable to education in noninclusive classrooms, while Lindsay (2007) argues that the balance of evidence in favor of inclusion for SWD is only "marginally positive" (p. 16).

Much of the conflicting evidence on inclusion can be attributed to the lack of clear definition and implementation differences across contexts. Issues of SWD's access to grade-level curriculum within the classroom, levels of individualized supports available, and differing teaching practices utilized further limit understanding of the specific mechanisms underlying the effectiveness of inclusive education. Evidence on co-teaching—a common practice for implementing inclusion—suggests that the staffing strategy has positive academic impacts on students with and without disabilities in inclusive settings (Jones & Winters, 2020; King-Sears et al., 2021; Tremblay, 2012), but more work is necessary to understand why. Further, not all SWD make progress in inclusive settings, even if performance improves on average, and students with different classifications cannot be treated interchangeably (Gilmour & Henry, 2018; Schulte & Stevens, 2015). Students with low-incidence, or severe, disabilities are disproportionately placed in more restrictive settings (Kurth et al., 2015; Smith, 2007), limiting knowledge of how students with the most significant needs may fare in inclusive environments.

Inclusion of Students without Disabilities

Studies on inclusion have also examined how students *without* disabilities (SWOD) in general education fare in inclusive settings. Research examining the peer effects associated with inclusive practices suggests largely negative effects on SWOD, with the caveat that many studies focus exclusively on the impacts of learning alongside students with significant behavioral problems—an attribute not representative of all SWD. Exposure to classmates with disruptive behaviors has been shown to have negative academic effects on other students in terms of both math and reading test scores (Fletcher, 2010). Peer behavior is similarly affected, with increases in the number of classmates with disabilities associated with lower levels of self-control and interpersonal skills among SWOD (Gottfried, 2014), as well as a potential reduction in lifetime earnings (Carrell et al., 2016).

Evidence from inclusion studies that do not focus on the behavior of SWD has found a mix of negative (Robinson, 2012), positive (Sharpe et al., 1994), and null (Brady, 2010; Brewton, 2005; McDonnell et al., 2003; Trabucco, 2011) impacts of inclusive education on the academic performance of SWOD. The variation in findings again suggests that the specifics of how inclusion is implemented matter significantly. Overall, the confluence of evidence when the behavior of SWD is not the primary independent variable suggests that the academic performance of SWOD is largely unaffected by the increased presence of peers with special needs in the same classroom (Brady, 2010; Brewton, 2005; McDonnell et al., 2003; Trabucco, 2011).

Method

Policy Details

Prior to implementation of the inclusion policy, SWD in the case study district were largely segregated from their nondisabled peers. While SWOD were enrolled almost exclusively in their neighborhood schools, 49% of all SWD were educated for the majority of their school day in separate classes, fully segregated settings, or regional centers. The case study district had one of the highest rates of SWD educated primarily in noninclusive settings across districts within the state in the years prior to the policy change. Two schools in the district served as "centers" in the pre-policy period, with targeted programs for specific student populations, including those with significant cognitive impairments, emotional disturbance/behavior disorders, visual or hearing impairments, and autism. Roughly 11% of all SWD in the district received special education services in one of these centers rather than their neighborhood schools, and nearly 100% of center-based students rode specialized school buses to and from these locations. Seven schools in the district also offered self-contained programs for students with severe cognitive impairments and emotional disabilities, while all district schools offered in-school resource classrooms for pull-out services.

Academic performance of SWD in the district was among the lowest statewide in the years preceding the inclusion policy. A separate, "parallel" curriculum was used in segregated classrooms, meaning that SWD received content distinct from their general education peers in the same grade levels. General education and special education teachers also received separate professional development (PD) programs, with special educator development focused on process and legal issues, while general educator development addressed content and student achievement indicators. Both student placement patterns and the separation of PD pathways in the pre-policy period contributed to a lack of collaboration across instructional staff and the continuation of segregated educational plans for SWD.

Additionally, in the years prior to the policy, the county in which the district is located experienced a steady influx of SWD from surrounding areas, both in and out of state, given its geographic location near the state border and a reputation for offering a large number of specialized services. This influx resulted in an associated increase in the costs of special education service provision. Transportation costs were a particular pain point, with a large proportion of SWD requiring specialized busing to non-neighborhood schools. While small in terms of population, the district is large by geographic area, increasing costs for transporting students to schools not in close proximity. Roughly 20% of the district's total annual transportation budget was allocated to special needs transportation in the years prior to the policy, with annual costs per special education student at nearly $4,000, compared to approximately $500 per general education student. These existing financial concerns, alongside concerns about low student performance, led district leaders to reexamine their broader approach to special education.

Implementation Process

The case study district sought implementation support from a nonprofit organization with experience facilitating whole-system transformation centered around inclusive practices. The organization had experience working with other districts in- and out-of-state and provided staffing, PD, and technical assistance for district staff throughout implementation.

The transition to inclusion at each district school followed a 4-year implementation arc. The district arranged cohorts of between four and eight schools into a predetermined order over the 8-year transition period (see Table 1). With minimal variation, district implementation began with elementary schools (grades K-5), followed by middle schools (grades 6–8), and then high schools (grades 9–12). Within school levels, the order of schools implementing the policy was close to random; that is, no specific criteria (e.g., test scores, stated willingness, and size of special education population) were used to determine the order of implementation among elementary, middle, or high schools. School-level transitions over the 4-year arc followed a consistent process, designed as a gradual-release model in which the capacity of each school slowly increased alongside a decrease in support from the external partner.

The first year of the policy focused on transitioning SWD from non-neighborhood schools into general education within their neighborhood community schools, and bringing students from more segregated settings within their community schools into general education. Schools worked to identify student-level needs and plan for individual students to transition. During the first year, schools also began participating in PD led by the nonprofit partner. In the second year, teams from both the receiving and sending schools met one-by-one with the families of each special education student, and the students themselves when age-appropriate, in a series of meetings to discuss the transition process and plan individualized support structures to ensure student success in the general education classroom in the receiving school.

The third year of implementation focused on developing and solidifying whole-school structures to support inclusive education, such as schedule revisions to allow for collaborative planning, and included additional PD on best practices for collaborative teaching. The fourth year of each school's transition emphasized improvements to the quality of instruction in classrooms and the meaningful participation of all students. PD during this year was designed to be responsive to the outstanding needs and challenges faced by schools in their final year of implementation.

Implementation Success

Figure 2 provides evidence of implementation success. Panel A demonstrates the district steadily increased the number of SWD spending 80% or more of their day in general education settings over the course of the 8-year policy transition period. The district average of 60% of SWD included just prior to the policy transition increased to more than 90% in post-implementation years. This high rate of inclusivity sustained over the subsequent decade. Panel B compares the inclusivity trends of the case study district to all other districts in the state over the same 20-year period, showing the same increase during the implementation period and sustained, high rates of inclusion over time. This figure also demonstrates that, while many districts in the state followed national trends of increased inclusion over this period, inclusion rates in the case study district moved beyond those observed elsewhere.

Data and Sample

Data

This study uses nonpublic, school-level data from the case study district along with data on academic outcomes from the district's associated state department of education and the National Center for Education Statistics' (NCES) Elementary and Secondary Information System. An original panel dataset is constructed, allowing observation of the key independent variable—placement into general education versus alternative educational environments as a student's LRE—along with academic and

TABLE 1

Implementation Plan for Case Study District Schools

Year	2002–03	2003–04	2004–05	2005–06	2006–07	2007–08	2008–09	2009–10
Year 1: Planning and training	A Elementary B Elementary C Elementary C Middle	D Elementary E Elementary F Elementary G Elementary	H Elementary I Elementary J Elementary K Elementary L Elementary M Elementary E Middle G Middle	Q Middle O Middle N Elementary O Elementary P Elementary	R Middle S Middle R High S High E High C High G High			
Year 2: Implementation and continued training support		A Elementary B Elementary C Elementary C Middle	D Elementary E Elementary F Elementary G Elementary	H Elementary I Elementary J Elementary K Elementary L Elementary M Elementary E Middle G Middle	Q Middle O Middle N Elementary O Elementary P Elementary	R Middle S Middle R High S High E High C High G High		
Year 3: Implementation and follow up			A Elementary B Elementary C Elementary C Middle	D Elementary E Elementary F Elementary G Elementary	H Elementary I Elementary J Elementary K Elementary L Elementary M Elementary E Middle G Middle	Q Middle O Middle N Elementary O Elementary P Elementary	R Middle S Middle R High S High E High C High G High	
Year 4: Continued training and support				A Elementary B Elementary C Elementary C Middle	D Elementary E Elementary F Elementary G Elementary	H Elementary I Elementary J Elementary K Elementary L Elementary M Elementary E Middle G Middle	Q Middle O Middle N Elementary O Elementary P Elementary	R Middle S Middle R High S High E High C High G High

Note. This chart reflects an artifact from the case study district outlining the plan for policy implementation across all district schools, but with the names of the schools redacted to preserve district anonymity. Each school in the district followed a 4-year implementation arc and all district schools completed the policy implementation between the 2002–03 and 2009–10 school years.

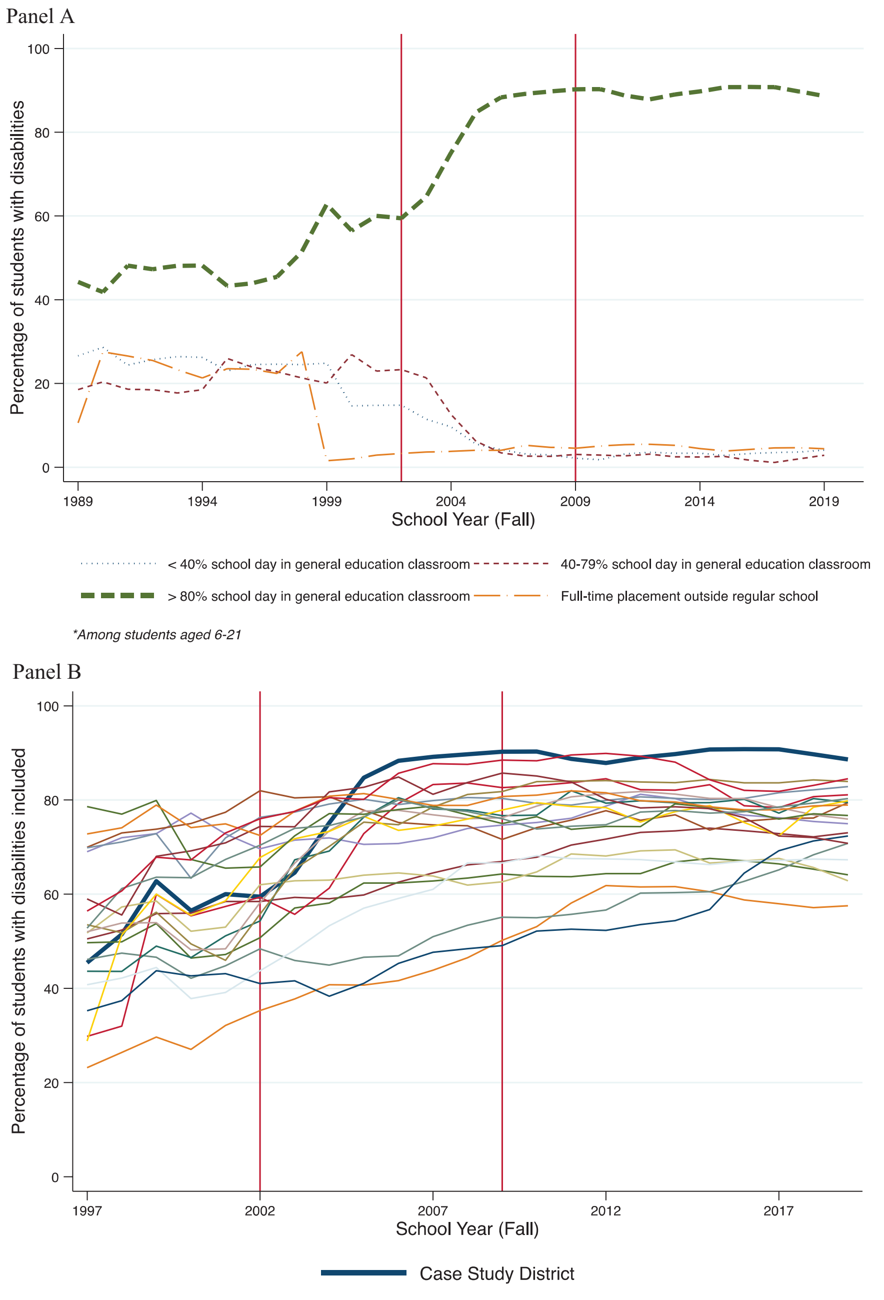

FIGURE 2. *Prevalence of inclusion.*

Note. (Panel A) Case study district. (Panel B) Statewide. (Panel A) This figure illustrates educational placement trends among students with disabilities ages 6 to 21 using historical data from the case study state department of education. Data are grouped into four categories. The first three categories reflect the percentage of the school day students spend in general education settings across the three federal reporting categories—more than 80%, 40% to 79%, or less than 40%. Prior to 2008, these three reporting categories reflect time spent in general education at more than 60%, 21% to 60%, and less than 21%, respectively. The final category groups all placements where students spend 100% of the school day in a non-public-school setting (e.g., separate school, separate residential facility, private school, homebound, hospital, or correctional facility). (Panel B) This figure illustrates the percentage of students with disabilities ages 6 to 21 whose primary educational placement is general education for more than 80% of the school day (i.e., an inclusive placement) for all school districts in the case study state. Data are from the case study state department of education. In both panels, vertical lines at 2002 and 2009 mark the beginning and end of the policy implementation period.

TABLE 2

Case Study District Demographics

Demographic	Mean
Race/Ethnicity	
% American Indian/Alaska Native	1.3
% Asian	1.7
% Black	10.4
% Native Hawaiian/Pacific Islander	0.4
% White	85.2
% Two or more races	6.4
% Hispanic	7.5
Gender	
% Male	52.3
% Female	47.7
Special education	
% Students with individualized education plans	16.3
% Students receiving 504 accommodations	3.8
Other demographics	
% English language learners	2.5
% Free or reduced-price lunch	49.3
% Title I	23.6
% Homeless	4.1
% Foster	0.5
% Gifted	6.7

Note. This table presents demographic averages for the case study district based on student-level records from the district for the 2019–20 school year.

behavioral measures of effectiveness consistent with prior literature (Fletcher, 2010; Hanushek et al., 2002; Morgan et al., 2010; Schifter & Hehir, 2018; Schwartz et al., 2019), including: high school graduation and dropout rates, rates of grade promotion and retention, attendance rates, and performance on state ELA and math assessments.

Sample and Context

The school district represented in this study is anonymous. Descriptive details on the district's student population are presented in Table 2. The majority of students enrolled are White, non-Hispanic. Approximately 17% of students within the district qualify for special education—three percentage points higher than the national average for public schools (NCES, 2019a). Just under half of all students qualify for free or reduced-priced lunch, and less than 3% of students are English-language learners. The district is in a rural area in the Northeast region of the United States, less than 5 miles from an urban area, with a population of less than 20,000 students. This setting is notable, as rural districts face particular special education challenges including teacher retention and recruitment and transportation issues (Berry & Gravelle, 2018), and more than half of all U.S. school districts are in rural environments (The School Superintendent's Association, 2017). Roughly 10% of district families fall below the poverty line, and 80% of households have internet access (NCES, 2019b).

Test Score Outcome Transformations

A key outcome measure in this study relies on standardized test score data reported by the case study state's department of education for all students and student subgroups in grades 3 through 8 as the percentage of students in a given grade in a given year performing at a proficient level. Raw school and district test score means are not reported, and under No Child Left Behind (NCLB), states were allowed to determine their own thresholds for proficiency which may change from year to year—presenting challenges for both long-term measurement and analysis.

Following the work of Reardon et al. (2016), homoskedastic ordered probit (HOMOP) is used to transform the reported frequencies of students scoring proficient or above into estimated means and standard deviations. An HOMOP model estimates a unique mean for each student group (all students, SWD, and SWOD) on each assessment within each school in each year. Each subgroup's subject-by-grade-by-school-by-year estimate is transformed from a frequency into an inference of that subgroup's *propensity* for proficiency. Under the assumption that test score distributions are normal (which they are in this instance), HOMOP allows for a transformation of percents-proficient into standard deviation units, which are implied differences in averages. This rescaling corrects for potential distortions that occur if proficiency thresholds are set near the extremes of normal distributions.

One constraint of this transformation approach is that it is limited in the cases of insufficient data or small sample sizes. Reardon et al. (2016) demonstrate that accurate estimations of means and standard deviations of test score distributions are possible when sample sizes are larger than 50. This primarily affects a key subgroup of interest—SWD—for whom within-school frequencies are naturally small. Between 85% and 98% of school-level proficiency counts (variation across subjects and grades) for SWD in this study are below this threshold and, therefore, have estimated test score means that are potentially slightly negatively biased. However, Reardon et al. (2016) note that, even when sample sizes are small, average bias is not sizable with respect to the true standard deviations in the underlying data.

Figures A.1 and A.2 in the Supplemental Appendix (available in the online version of the journal) illustrate the impact of these transformations on the underlying proficiency data for reading and math assessments. Both figures show average student performance by district over time, with pre-transformation trends in column 1 and transformed performance data in column 2. For both subjects, and across all student groups, in the pretransformed data, there is a steady increase for all districts over time up until 2012, when the introduction of the Partnership for Assessment of Readiness for College and Career (PARCC) assessment began influencing curriculum decisions in classrooms, followed by a subsequent decline. This same trend is not observed in the transformed data. These figures make clear that viewing student proficiency as a rate alone distorts actual trends in student performance. While the percentage of students performing proficient or higher on state exams steadily increases in the years prior to the PARCC rollout, actual student performance in terms of estimated means is more consistent. The retrieved test score means are used as the key indicator of academic performance in the event study analysis.

Research Design

A variation of the standard two-way fixed effects (TWFE) difference-in-differences (DiD) strategy is used to estimate the impact of the district's policy of inclusion on students' academic and behavioral outcomes. This strategy that draws on variation in the year in which a school began implementing the inclusion policy, allowing for examination of potentially dynamic treatment effects. The following model, accounting for this staggered adoption of treatment, reflects the main estimation of the policy's effectiveness:

$$y_{st} = \alpha_s + \delta_{tg} + \sum_{\substack{k=-3 \\ k\neq -1}}^{k=8} 1(t = t_s^* + k)\beta_k + \varepsilon_{st}.$$

In this specification, y_{st} reflects an outcome for a given school, s, in a given year, t. The parameters α_s and δ_{tg} indicate the inclusion of both school and year-by-school-level fixed effects, respectively, controlling for school-invariant and time-by-school-level-invariant differences across schools. The latter restricts within-year comparisons to schools at the same level (e.g., elementary, middle, high). The effect of the inclusion policy's implementation beginning in year t_s^* is reflected in the coefficient β_k, relative to outcomes k years later. The model traces out the comparison between treated and untreated schools from 3 years prior to the inclusion policy's implementation for a given school to 8 years after implementation began, omitting the year prior to the start of implementation as the excluded group.

The variation that identifies each β_k, therefore, comes from the interaction between within-school changes and time, as two comparisons of the outcome variable: (a) comparing to the years before the policy change began for a given school and (b) comparing treated and untreated schools within the same level and academic year. This estimation strategy allows for observation of the policy's potentially heterogeneous effects throughout the formal treatment period as well as after treatment has concluded, explicitly modeling the dynamic treatment effects across time. In all models, standard errors are clustered at the school level to address any potential bias resulting from serial correlation across outcome variables given that data span multiple years and variation occurs only at the group level (Angrist & Pischke, 2015).

Results from the event study are presented graphically and are also estimated in a piecewise spline function, decomposing a more traditional DiD estimate into an "implementation period" (years 0–4) and a "post-implementation" period (years 5+). This model captures the most important differences in policy response and allows for variation in impact over time, but has greater statistical power than the event study. Four design choices and assumptions underlie the causality of findings from this model: comparison group selections, parallel trends, exogenous assignment to treatment, and homogeneous treatment effects.

Comparison Group Selection

All other untreated schools within the state are used as the comparison group. The large number of untreated schools (1,723) offers power, and the high level of certainty about their having never received treatment lends substantive confidence to the choice. The limit of this approach is that the sample of schools statewide differs from the schools in the case study district in terms of both demographics and geography, as the larger sample necessarily includes a wider range of school sizes, compositions, and locations (see column 2 of Table 3).

Limiting the comparison group to only untreated schools in other rural districts within the state, as the case study district is in a rural setting, is also considered. Given that rural school districts face particular challenges with respect to special education, this approach seemingly has substantive merit. However, there remain both demographic and geographic differences between the populations of schools in these untreated, rural districts, and those in the case study district, despite their shared rurality (see column 3 of Table 3). Additionally, the diminished sample size (117 schools) results in a substantive loss of power.

A synthetic comparison group is a third option—a subset of schools (1,087) drawn from the full pool of untreated schools within the state based on a set of observable characteristics used to match to the set of treated schools.[2] Descriptive statistics for this third group (shown in column 4 of Table 3) are highly similar to those of the full population of untreated schools within the state. As such, the main results are based on a comparison to the broader comparison group, though to assess the sensitivity of findings to this choice, results compared to both comparison group alternatives are presented in an Supplemental Appendix (available in the online version of the journal).

Parallel Trends

Demographics of schools in the case study district are compared to those in all other untreated districts in the state to test for the presence of parallel trends. Data on key outcomes in the case study district are not available until the first year of policy implementation and render outcome-based pre-trends unobservable. Inability to observe pre-trends is a common limitation across DiD studies (Roth, 2019); however, historical demographic data are available to assess pre-trends and offer some evidence that there were no substantive changes to the case study district or comparison districts over the arc of the policy implementation period.

Demographic data from 1986 to 2019 offer 16 years of pre-trend information. As a proxy for outcome data, these data show that based on the composition of case study district schools compared to untreated schools in the state by race, gender, disability status, and the percent of students eligible for free or reduced-price lunch, parallel trends do exist in the period preceding treatment (see Figure A.3 in the Supplemental Appendix in the online version of the journal). This confirms that the composition of the case study district did not substantively change before and after the policy implementation and that the path of untreated schools in the state serves as a meaningful comparison for the case study district.

Exogenous Assignment to Treatment

Causal interpretation of DiD results relies on an assumption of exogenous assignment to treatment—that treated units' assignment to treatment is either random or as-good-as random. Implementation across district schools began with elementary schools, followed by middle and high schools. While this was not random, the selection of schools *within* levels was as-good-as random; that is, the order was not determined by schools' level of pre-policy effectiveness, openness to inclusion, or some other qualifying criterion.

TABLE 3

Summary Statistics—Case Study District and Comparison Groups

Statistic	Case study district	Comparison group A (untreated schools in state)	Comparison group B (untreated rural schools in state)	Comparison group C (synthetic comparison group)
Avg. district population	99,069	555,557*	61,993*	571,690*
Avg. school enrollment	566	646	545	662
% SPED	14.2	12.4	12.4	12.4
% FRPL	24.2	35.9*	32.6*	33.6*
% AIAN	0.31	0.39	0.26	0.42
% Asian	0.68	4.36*	0.99	4.99*
% Hispanic	2.12	5.87	1.55	6.71*
% Black	6.96	31.1*	18.9*	33.2*
% White	85.2	45.7*	72.2*	49.9*
Median household income	$65,079	$70,424	$65,441	$73,176*
% Poverty	7.9	8.7	9.3*	7.9
Number of districts	1	21	6	21
Number of schools	30	1,723	117	1,087

Note. This table presents summary statistics for the case study district and three potential comparison groups using data from the NCES Elementary and Secondary Information System (ELSI) from the 2002–03 school year (the year in which policy implementation began in the case study district). Data reflect averages across all schools in each group. Asterisks (*) indicate comparison group means that are statistically significantly different from those of the treated case study district ($p < .05$).

A threat to validity is if the anticipation of treatment led schools to begin policy implementation at a time other than their assigned start of treatment. Despite schools' knowing the order of implementation in advance, there is no evidence this happened in practice. A key component of the policy required transitioning SWD one-by-one from the schools in which they were located into their local community schools. This required a series of meetings among the staff at the receiving school, the staff at the sending school, and the families of each student. It is unlikely these stakeholder groups added additional meetings outside the prescribed order, or would have wanted to move individual students to inclusive schools and classrooms without the broader changes to school-wide structures already in place.

Heterogeneous Treatment Effects

A final threat to validity results from drawing on a longer panel of data, wherein early-treated schools become incorporated within the comparison group for later-treated schools, muddling the identification of average treatment effects (Goodman-Bacon, 2021). This is an issue if there are differences in the impact of the treatment over time. Two tests confirm the presence of heterogeneous treatment effects in the data.

First, the weights associated with each individual DiD estimator of average treatment effects underlying a standard TWFE regression where differential treatment timing exists are computed. Following De Chaisemartin and D'Haultfœuille (2018) and Goodman-Bacon (2021), identical weights or a lack of negative weights would indicate homogeneous treatment effects. This assessment shows negative weights to be associated with more than one individual TWFE regression and that the magnitude of the weights varies over time—evidence of heterogeneous treatment effects. Because schools were treated in cohorts of four to eight schools in each year of the implementation period, weights of each cohort-specific average treatment effect on the treated (CATT) underlying the TWFE regressions in the event study specification are also computed.[3] These weights better reflect the impact of heterogeneous treatment effects over time, in the appropriate context of dynamic treatment effects. Figure A.4 in the Supplemental Appendix (available in the online version of the journal) displays these weights for each of the five treated cohorts over their respective, 4-year policy implementation periods and shows differential weights across cohorts—offering further evidence of different policy impacts for each treated group.

Three steps are taken to address this. First, the nature of the policy implementation was such that the order of treated schools was almost perfectly correlated with the *type* of school (i.e., cohorts comprised groups of elementary, middle, or high schools). It is probable that the initial TWFE weights analysis conflates this strong correlation between treatment timing and school level in the data. In this case, heterogeneous treatment effects are logical given the reasonable expectation that different school levels might respond differently to the inclusion policy. This expectation is further corroborated by the CATT analysis illustrated in Figure A.4 in Supplemental Appendix (available in the online version of the journal). Given this, school-level-specific results are presented along with the main analyses aggregating the policy's impacts for all treated schools. Additionally, all models control for year-by-school-type fixed effects. As a final step, results are also estimated using an "interaction-weighted" estimator that reflects a weighted average of each cohort-average treatment-on-the-treated estimate.

Results

Results from the main event study are presented in Tables 4 to 7, with findings reflected for all students in all grades (3–12), elementary school students (grades 3–5), middle school students (grades 6–8), and high school students (grades 9–12), respectively. Results are split into three periods: the pre-policy period (all years prior to the start of implementation), the policy implementation period (0–4 years), and the post-implementation period (5–9 years after the policy implementation concluded).[4] The pre-policy period is omitted as a reference group. This piece-wise spline specification allows for different slopes of exposure across these three, meaningful time periods for the policy's implementation. Results estimate the policy's impact on each outcome: attendance rates, math and reading test score means, dropout rates, graduation rates, and promotion rates.

Attendance

As the policy implementation required significant shifts for SWD, many of whom were transitioned into entirely new school buildings as well as into general education classrooms, there is a reasonable expectation that there would be equally significant disruption to student attendance—both for the moving students as well as their peers in classrooms with new student compositions. However, the results do not bear this out. While there are some statistically significant findings observed across all grades and within the school-level-specific results, the magnitude of the findings is so small as to have no substantive significance: Both increases and decreases in attendance rates are within one percentage point in either direction. While the estimates across school level models skew slightly more positive, particularly for SWOD, overall, there were no substantive impacts on student attendance as a result of this policy. Figure 3 shows event study estimates for all students in grades 3 through 12; results for all other school levels and student groups are shown in Figure F.1 in the Supplemental Appendix (available in the online version of the journal).

Math and Reading Test Scores

Contrary to expectations, estimates from all event study specifications showed no statistically significant changes in test scores for either student subgroup in reading or math. In other words, students with and without disabilities did no better or worse academically as a result of this policy implementation. This finding is reflected in Figure 4 for all grades and student groups. Results further disaggregated by elementary and middle school are shown in Figures F.2 and F.3 in the Supplemental Appendix (available in the online version of the journal).

Reardon et al. (2016) note that HOMOP transformations are limited in the cases of insufficient data and that imprecise estimates are possible when sample sizes fall below 50. As this impacts a key subgroup of interest in this study—SWD,

TABLE 4

Event Study Results (All Grades)

	Implementation period (0–4 years)			Post-implementation (5–9 years)		
Outcome	All students	SWD	SWOD	All students	SWD	SWOD
Attendance rate	0.409**	−0.0670	0.429**	0.610***	0.0768	0.640***
	(0.179)	(0.355)	(0.191)	(0.196)	(0.314)	(0.211)
Pre-treatment mean	92.8	91.1	93	92.8	91.1	93
Number of schools	1,606	1,579	1,562	1,606	1,579	1,562
Math test score means	0.0432	0.0223	0.0867	0.0327	0.0209	0.0638
	(0.0429)	(0.0858)	(0.0536)	(0.0445)	(0.0848)	(0.0584)
Pre-treatment mean	0.020	−0.023	0.029	0.020	−0.023	0.029
Number of schools	1,057	1,043	1,050	1,057	1,043	1,050
Reading test score means	−0.0117	0.0530	0.0233	−0.0139	0.0228	0.0168
	(0.0331)	(0.0669)	(0.0376)	(0.0416)	(0.0748)	(0.0461)
Pre-treatment mean	−0.008	−0.069	−0.014	−0.008	−0.069	−0.014
Number of schools	1,057	1,055	1,050	1,057	1,055	1,050

Note. Coefficients are event study estimates of the impact of the inclusion policy's implementation on the indicated outcome for all students, SWD, and SWOD across grades 3 through 12 for attendance and 3 through 8 for math and reading test scores. Disaggregated outcome data are from the case study state department of education. All regressions control for school and year-by-school-level fixed effects. Results are split into three periods: the pre-policy period (all years prior to the start of implementation), the policy implementation period (0–4 years), and the post-implementation period (5–9 years after the policy implementation concluded). The pre-policy period is omitted as a reference group. Leads greater than 3 years prior to implementation and lags more than 9 years after the start of implementation were binned, respectively. Math and reading test score means are transformed from reported subgroup proficiency rates into recovered test score means using HOMOP. Standard errors are clustered at the school level. The sample includes all schools in the case study district, while the comparison group comprises all untreated schools in the state. Robust standard errors are in parentheses (***$p < .01$, **$p < .05$, *$p < .1$). SWD = students with disabilities; SWOD = students without disabilities.

TABLE 5

Event Study Results (Elementary)

	Implementation period (0–4 years)			Post-implementation (5+ years)		
Outcome	All students	SWD	SWOD	All students	SWD	SWOD
Attendance rate	−0.0152	−0.0658	0.0728	−0.100	−0.624***	−0.115**
	(0.0570)	(0.157)	(0.0594)	(0.0746)	(0.187)	(0.0575)
Pre-treatment mean	94.5	94.5	95	94.5	94.5	95
Number of schools	1,028	1,505	1,503	1,028	1,505	1,503
Math test score means	0.0657	0.0117	0.120	0.0511	0.00953	0.0917
	(0.0584)	(0.126)	(0.0745)	(0.0584)	(0.118)	(0.0812)
Pre-treatment mean	0.01	−0.041	−0.023	0.01	−0.041	−0.023
Number of schools	799	797	797	799	797	797
Reading test score means	−0.0135	0.0676	0.0311	−0.0149	0.0428	0.0333
	(0.0477)	(0.103)	(0.0556)	(0.0591)	(0.111)	(0.0682)
Pre-treatment mean	−0.030	−0.119	−0.035	−0.030	−0.119	−0.035
Number of schools	799	797	797	799	797	797

Note. Coefficients are event study estimates of the impact of the inclusion policy's implementation on the indicated outcome for all students, SWD, and SWOD across all grades 3 through 5. Disaggregated outcome data are from the case study state department of education. All regressions control for school and year-by-school-level fixed effects. Results are split into three periods: the pre-policy period (all years prior to the start of implementation), the policy implementation period (0–4 years), and the post-implementation period (5–9 years after the policy implementation concluded). The pre-policy period is omitted as a reference group. Leads greater than 3 years prior to implementation and lags more than 9 years after the start of implementation were binned, respectively. Math and reading test score means are transformed from reported subgroup proficiency rates into recovered test score means using homoskedastic ordered probit. Standard errors are clustered at the school level. The sample includes all schools in the case study district, while the comparison group comprises all untreated schools in the state. Robust standard errors are in parentheses (***$p<.01$, **$p<.05$, *$p<.1$). SWD = students with disabilities; SWOD = students without disabilities.

TABLE 6

Event Study Results (Middle)

	Implementation period (0–4 years)			Post-implementation (5 + years)		
Outcome	All students	SWD	SWOD	All students	SWD	SWOD
Attendance Rate	0.000920	−0.175	−0.0242	0.368**	0.290	0.353
	(0.134)	(0.314)	(0.153)	(0.170)	(0.456)	(0.245)
Pre-treatment mean	93.6	90.4	94	93.6	90.4	94
Number of schools	493	480	457	493	480	457
Math test score means	0.00590	0.0406	0.0321	0.00184	0.0455	0.0181
	(0.0588)	(0.0951)	(0.0737)	(0.0686)	(0.123)	(0.0759)
Pre-treatment mean	0.049	0.026	0.044	0.049	0.026	0.044
Number of schools	451	374	446	451	374	446
Reading test score means	−0.0129	0.0342	0.00752	−0.0197	−0.00277	0.00224
	(0.0421)	(0.0490)	(0.0483)	(0.0504)	(0.0693)	(0.0473)
Pre-treatment mean	0.053	0.072	0.047	0.053	0.072	0.047
Number of schools	451	447	446	451	447	446

Note. Coefficients are event study estimates of the impact of the inclusion policy's implementation on the indicated outcome for all students, SWD, and SWOD across all grades 6 through 8. Disaggregated outcome data are from the case study state department of education. All regressions control for school and year-by-school-level fixed effects. Results are split into three periods: the pre-policy period (all years prior to the start of implementation), the policy implementation period (0–4 years), and the post-implementation period (5–9 years after the policy implementation concluded). The pre-policy period is omitted as a reference group. Leads greater than 3 years prior to implementation and lags more than 9 years after the start of implementation were binned, respectively. Math and reading test score means are transformed from reported subgroup proficiency rates into recovered test score means using homoskedastic ordered probit. Standard errors are clustered at the school level. The sample includes all schools in the case study district, while the comparison group comprises all untreated schools in the state. Robust standard errors are in parentheses (***$p < .01$, **$p < .05$, *$p < .1$). SWD = students with disabilities; SWOD = students without disabilities.

TABLE 7

Event Study Results (High)

	Implementation period (0–4 years)			Post-implementation (5+ years)		
Outcome	All students	SWD	SWOD	All students	SWD	SWOD
Attendance rate	0.115	−0.993***	0.189	1.059***	0.0959	1.109***
	(0.255)	(0.337)	(0.253)	(0.322)	(0.377)	(0.329)
Pre-treatment mean	90.6	88.9	90.9	90.6	88.9	90.9
Number of schools	353	332	316	353	332	316
Dropout rate	0.224	2.505***	−0.383	−0.551	0.648	−0.826***
	(0.389)	(0.460)	(0.332)	(0.427)	(0.519)	(0.301)
Pre-treatment mean	4.2	9.9	3.3	4.2	9.9	3.3
Number of schools	369	347	341	369	347	341
Graduation rate	−1.755	—	—	2.631**	—	—
	(1.301)			(1.266)		
Pre-treatment mean	81.3			81.3		
Number of schools	325			325		
Promotion rate (9–12)	0.451	—	—	2.945***	—	—
	(0.920)			(1.020)		
Pre-treatment mean	91.3			91.3		
Number of schools	379			379		
Promotion rate (9th Grade)	1.650*	—	—	6.738***	—	—
	(0.959)			(1.298)		
Pre-treatment mean	87.7			87.7		
Number of schools	352			352		
Promotion rate (10th Grade)	−1.243	—	—	2.195***	—	—
	(0.973)			(0.838)		
Pre-treatment mean	89.3			89.3		
Number of schools	336			336		
Promotion rate (11th Grade)	−0.170	—	—	0.983	—	—
	(0.804)			(0.675)		
Pre-treatment mean	93.4			93.4		
Number of schools	340			340		
Promotion rate (11th Grade)	−0.249	—	—	0.0954	—	—
	(0.607)			(0.751)		
Pre-treatment mean	94.6			94.6		
Number of schools	332			332		

Note. Coefficients are event study estimates of the impact of the inclusion policy's implementation on the indicated outcome for all students, SWD, and SWOD across all grades 9 through 12. Outcome data are from the case study state department of education. All regressions control for school and year-by-school-level fixed effects. Results are split into three periods: the pre-policy period (all years prior to the start of implementation), the policy implementation period (0–4 years), and the post-implementation period (5–9 years after the policy implementation concluded). The pre-policy period is omitted as a reference group. Leads greater than 3 years prior to implementation and lags more than 9 years after the start of implementation were binned, respectively. Standard errors are clustered at the school level. The sample includes all schools in the case study district, while the comparison group comprises all untreated schools in the state. Robust standard errors are in parentheses (***$p < .01$, **$p < .05$, *$p < .1$). SWD = students with disabilities; SWOD = students without disabilities.

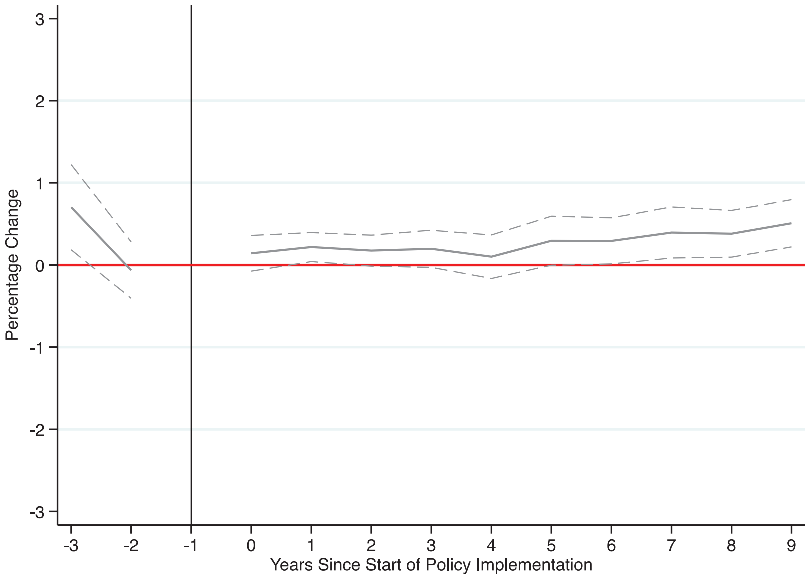

FIGURE 3 *Attendance rates—all students.*
Note. This figure presents results from the event study, illustrating the impact of the inclusion policy on attendance rates for all students in grades 3 through 12 from 3 years prior to the start of implementation to 9 years after implementation began. Dotted lines reflect confidence intervals around the main estimates. The year before the start of implementation is excluded as the reference group.

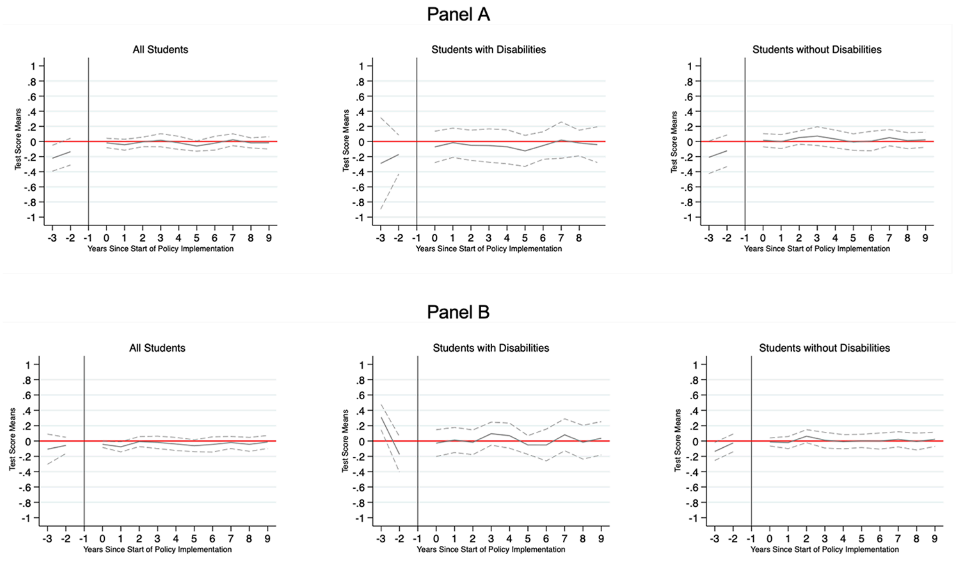

FIGURE 4. *Math and reading test score means.*
Note. This figure presents results from the event study, illustrating the impact of the inclusion policy on math (Panel A) and reading (Panel B) test score means for all students in grades 3 through 8 from 3 years prior to the start of implementation to 9 years after implementation began. Dotted lines reflect confidence intervals around the main estimates. The year before the start of implementation is excluded as the reference group.

for whom within-school frequencies are naturally small—analyses of impacts on test scores were run again using only cell counts greater than or equal to 50. These results, presented in Supplemental Appendix E (available in the online version of the journal), are consistent with

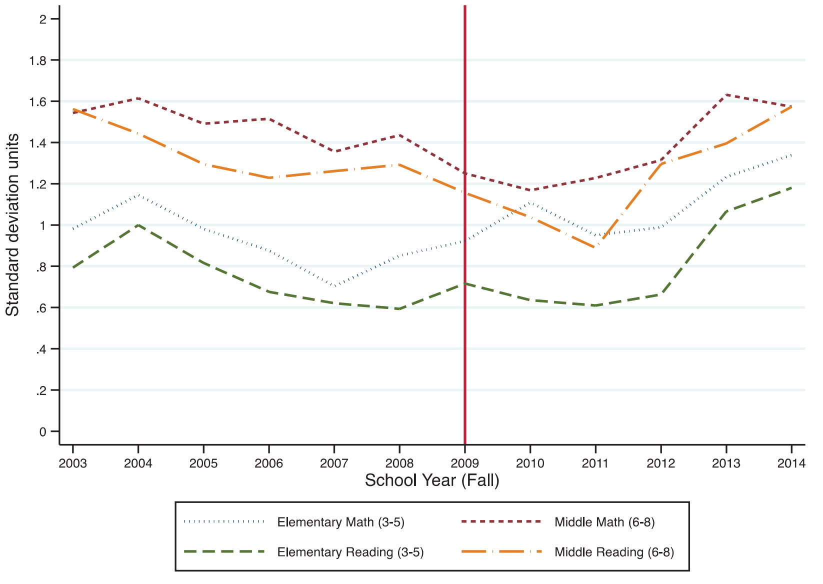

FIGURE 5. *Student achievement gap trends.*
Note. This figure presents average achievement gaps between students with and without disabilities on elementary and middle school reading and math standardized assessments between 2003 and 2014. The vertical line at 2009 marks the end of the implementation period. Elementary gap trends reflect averages across grades 3 through 5, and middle school gap trends reflect averages across grades 6 through 8. Achievement gaps represent average, group-level differences wherein each group's originally reported percent-proficient metric has been transformed into the group's average latent propensity for proficiency interpretable in a standard deviation-unit metric (Ho, 2009). Raw data are from the case study department of education.

initial findings; there were, again, no changes to either math or reading test score means for either student group.

The impact of the policy on achievement gaps *between* the two groups—students with and without disabilities—is also assessed. In the event study results, achievement reflects recovered test score means transformed from percents-proficient among each subgroup through HOMOP. These transformations make comparisons over time more reliable, but—as previously discussed—they are limited as they take place subgroup-by-subgroup, meaning the resulting test score means (and performance gaps between subgroups) are not directly interpretable from the previous results.

To speak to achievement gaps, following Ho (2009), the inverse normal of the original, state-reported percent-proficient for each subgroup is taken and used to calculate the resulting achievement gaps on this scale of standard deviation units. Figure 5 presents the path of achievement gaps between the two groups over time, again split by school level.[5] During the policy implementation years, the achievement gap between students with and without disabilities generally declines in both reading and math; however, post-implementation these trends reverse course and gaps between the two groups increase back to pre-policy levels or higher.

Dropout Rates

Absent consistent, standardized assessment data in high school subjects, alternative metrics are used to gage policy impacts on students' academics at this level, including 4-year adjusted cohort dropout rates.[6] During the implementation period, a two percentage-point increase in dropout rates is observed among SWD ($p < .01$). However, this increase did not sustain for SWD after implementation concluded (see Figure 6). SWOD saw a slight *decline* in dropout rates in the 5 to 9 years following implementation ($p < .01$), suggesting a longer-term positive effect for nondisabled students of time spent in classrooms with diverse learners.

Graduation Rates

The policy was associated with no impact on graduation rates during the implementation period, but resulted in a 2.6 percentage point

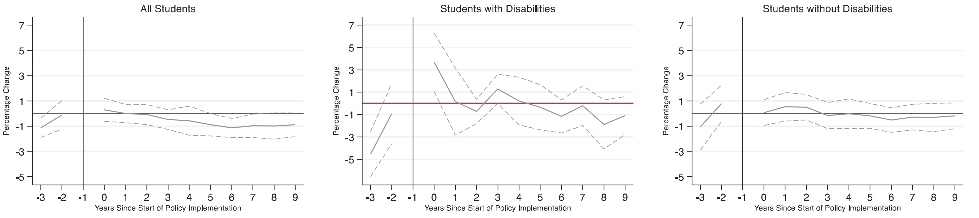

FIGURE 6. *High school dropout rates.*
Note. This figure presents results from the event study, illustrating the impact of the inclusion policy on high school dropout rates for all students in grades 9 through 12 from 3 years prior to the start of implementation to 9 years after implementation began. Dotted lines reflect confidence intervals around the main estimates. The year before the start of implementation is excluded as the reference group.

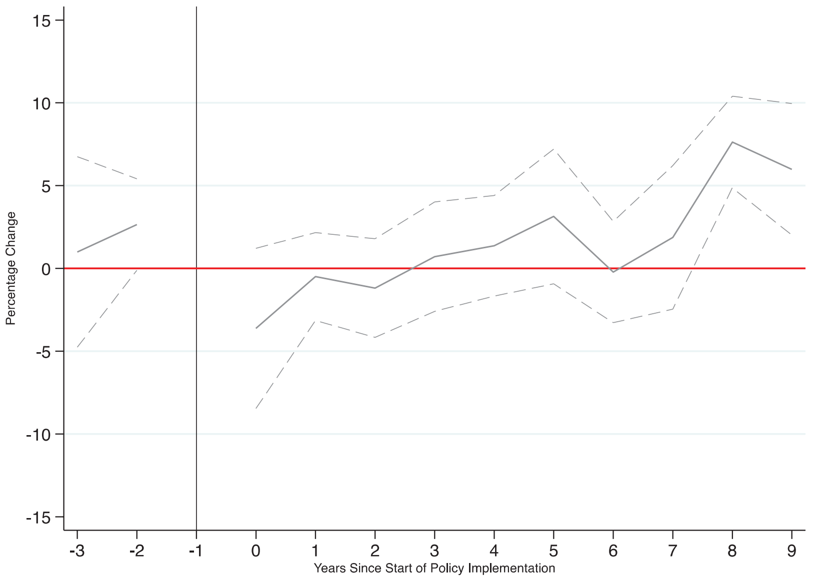

FIGURE 7. *High school graduation rates.*
Note. This figure presents results from the event study, illustrating the impact of the inclusion policy on high school graduation rates for all students in grades 9 through 12 from 3 years prior to the start of implementation to 9 years after implementation began. Dotted lines reflect confidence intervals around the main estimates. The year before the start of implementation is excluded as the reference group.

increase in the years following implementation ($p < .05$).[7] Figure 7 presents event study results illustrating a steady increase in graduation rates observed over the 9-year, post-implementation period. Figure F.4 in the Supplemental Appendix (available in the online version of the journal) further highlights this long-term positive trend over the subsequent decade. While graduation rate data are unavailable by subgroup from the pre-policy and implementation periods, disaggregated data are available from 2009 to 2019—the post-implementation years. High school graduation rates across the full population and both subgroups increase steadily in the decade following the conclusion of the policy implementation, offering additional, descriptive evidence of positive academic impacts for students in the case study district in the long term.

Promotion Rates

Results for student promotion rates from one grade to the next are consistent with results for both dropout and graduation rates. The policy did not affect students' likelihood of promotion to the next grade during the implementation period, but had a positive, statistically significant impact in the longer-term. Students in 9th and 10th grade

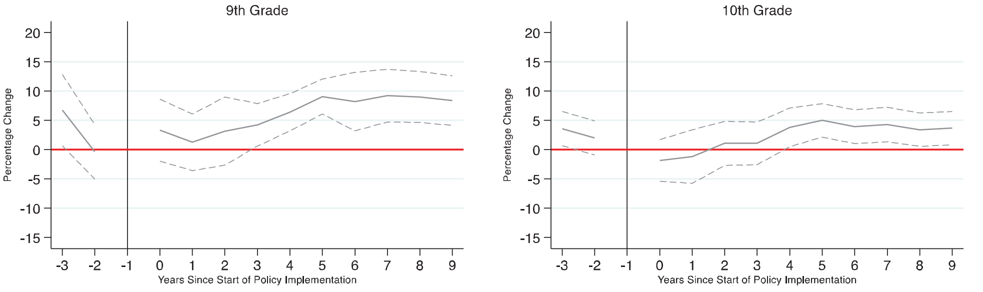

FIGURE 8. *High school promotion rates.*
Note. This figure presents results from the event study, illustrating the impact of the inclusion policy on high school promotion rates for all students in grades 9 through 12 from 3 years prior to the start of implementation to 9 years after implementation began. Dotted lines reflect confidence intervals around the main estimates. The year before the start of implementation is excluded as the reference group.

were 6.7 and 2.2 percentage points more likely, respectively, to be promoted to the next grade in the post-implementation years ($p < .01$). No impacts on student promotion were observed in the later high school grades. Results for 9th and 10th grade promotion rates are presented in Figure 8, and all other grades are in Figure F.5 in the Supplemental Appendix (available in the online version of the journal). This is a final piece of evidence suggesting a positive policy impact on high school students' academics. While promotion rates are not a direct measure of academic proficiency, they are by definition a measure of preparedness for the next grade level and are therefore a comprehensive measure, similar to graduation rates, of students' ability to succeed academically in inclusive learning environments.

Robustness Checks

A series of robustness checks assess the sensitivity of results to the choice of comparison group as well as the potential confounding influences of: heterogeneous treatment effects, changes to the tested student population, changes to the population of SWD, and differential policy impacts across disability classifications.

Comparison Group Choice

Results against both alternative comparison groups—only other schools in rural districts in the state and a synthetically generated comparison group—are presented in Supplemental Appendix B (available in the online version of the journal), and show some minor differences compared to the main results. Comparing against other rural schools in the state, attendance rates across all school levels remain substantively insignificant, with changes still inside one percentage point in either direction. Math and reading test scores also remain largely unchanged, though one coefficient—math scores for SWOD across grades 3 through 8 during the implementation period—becomes statistically significant, reflecting a slight increase for these students' test scores of 0.09 standard deviations as a result of the policy ($p < .10$). Among high school students, dropout rates for SWD are still shown to increase during the implementation period, but against the rural-only comparison group the increase in dropouts sustains in the post-implementation years—though at a lower rate than during the implementation years (1.3 vs. 2.3 percentage points). Graduation and promotion rate estimates also shift, with results suggesting small decreases for all students during implementation, but no changes to either over the longer term.

Measuring policy effects against the synthetic comparison group, findings are even closer to the main results. Attendance rates are not substantively impacted at any school level, and math and reading test scores are again shown to be unaffected by the inclusion policy. At the high school level, dropout rates for SWD increase during the implementation period to a similar degree as the main results (3 percentage points vs. 2.5), but again this increase does not sustain in the later years. Graduation and promotion rates increase post-implementation at rates similar to the primary findings. Overall, though there are some changes to findings across these six outcomes, none are significant enough to alter broad conclusions drawn about policy impact.

Heterogeneous Treatment Effects

To address heterogeneous treatment effects in the main results, findings are disaggregated by school level and include year-by-school-type fixed effects. As an additional check, the main model is also estimated using an interaction-weighted estimator, accounting for the weighted average of each cohort-average treatment-on-the-treated estimate (Sun & Abraham, 2021). Results from this estimation are in Supplemental Appendix C (available in the online version of the journal) and show small decreases in the magnitude of some coefficients but no changes to either substantive or statistical significance for any outcome. This confirms that the adjustments to the main model have accounted for the majority of this issue.

Testing

The years over which this policy took place overlap with significant federal changes to school accountability policies under NCLB, which required all states to test all students annually, and that results be disaggregated and reported for specific student subgroups including SWD. This increased accountability mechanism likely drew new students into the tested-students sample, which would bias estimates of the inclusion policy's impact if students who were less likely to perform well on standardized assessments (e.g., students with more severe disabilities) were increasingly included in the sample.

Figure G.1 in the Supplemental Appendix (available in the online version of the journal) shows the percentage of test takers over time for both the case study district and other districts in the state, for all students and SWD. There is a notable increase in the number of test takers across all groups between 2003 and 2005, likely a result of NCLB accountability mechanisms slowly changing district and school behavior. To assess whether this descriptive increase in test takers is biasing results, participation rates are regressed as an outcome using the main event study model. Results from this assessment are in Table G.1 in the Supplemental Appendix (available in the online version of the journal) and show that participation rates among SWD did not change substantively during either the policy's implementation period or the 5 to 9 years after it concluded. While data are not available on the test-taking population by disability classification, this suggests that results examining students' academic outcomes are not biased by changes in the overall number of SWD participating in testing.

While the passage of NCLB overlapped with the beginning of the inclusion policy's implementation, the introduction of the PARCC assessment overlapped with the end. In the 2014 to 2015 school year, all schools in the case study state formally transitioned from using the state standardized assessment of the previous decade to the nationally normed PARCC assessment as the statewide measure of students' academic performance. Pilot testing of the PARCC assessment began statewide in 2013 to 2014, and teaching transitions to curriculum addressing the Common Core State Standards (standards aligned to the PARCC assessment, but not aligned to the previously used state assessment) began as early as 2012 to 2013. While state test data are available from this period, student performance on state assessments is not a reliable indicator of academic achievement and is less comparable to data from previous years. Results from the main model eliminating the years after 2012 to 2013 are in Supplemental Appendix D (available in the online version of the journal), and reflect some small increases in the magnitude of some coefficients, but no substantive changes to overall results. This implies that the influence of PARCC in the later years slightly negatively biased the main results.

District Population Changes

An increase in the number of SWD being removed from or moving out of the case study district in response to the policy would similarly bias results and mask the policy's true impact. Figure G.2 in the Supplemental Appendix (available in the online version of the journal) displays the number of nonpublic placements in the case study district over time in two ways: as the raw number of nonpublic placements and as the number of nonpublic placements against the total number of SWD in the district. The data in Panel A show a slight increase in the number of SWD sent to private settings over the policy implementation period, followed by a steady decline back to pre-policy levels beginning in 2008. Panel B compares these numbers to the total number of SWD within the district and demonstrates that, while nonpublic placements increased over the policy implementation period, the change was not meaningful with respect to the total number of SWD who remained in the district's public schools (98% of all SWD) under the new policy of inclusion.

A related check on overall student mobility rates confirms that the case study district saw no significant change in student mobility over the policy implementation period, with an average mobility rate of around 30% each year between 2002 and 2009. Mobility rates measure the sum of student entrants and withdrawals over a total student population and are, therefore, not a perfect indicator of the number of students exiting a district voluntarily. However, the consistency of the case study district's mobility rate means either the inclusion policy did not spur an increase in voluntary student exits or there was an increase in student withdrawals but it was masked by a comparable influx of new entrants each year—a mathematical improbability.

Disability Classifications

By 2006, roughly 90% of SWD in the case study district were placed in general education as their primary learning environment or LRE—a rate which sustained until 2019.[8] Student-level data from post-implementation years enable a more granular analysis of whether students across disability types were equally likely to be placed in inclusive classrooms, with some expected variation relative to their level of need. While this information is only available for the 2020 to 2021 school year, given that the proportion of all SWD in general education remained high over a 13-year period and there is no evidence of SWD disproportionately exiting the district in response to the policy, it is probable that the underlying composition of students in inclusive classrooms also did not substantively change over this time.

Figure G.3 in the Supplemental Appendix (available in the online version of the journal) shows the distribution of disability classifications within inclusion settings as a proportion of all students with each classification in the population for the 2020 to 2021 school year. Panel A suggests that there is no systematic discrimination by disability classification in terms of likelihood of placement in general education settings. The majority of all SWD, regardless of classification, are spending 80% or more of their school day in general education. Panel B reaffirms this conclusion, but also demonstrates that the likelihood of placement in inclusion varies across classifications, as expected. Smaller percentages of students with more severe disabilities (i.e., emotional disturbance and intellectual disabilities) are placed in inclusive classrooms relative to peers with less severe disabilities. There are likely differential impacts of the inclusion policy among SWD by classification; however, understanding these impacts requires more granular data beyond the scope of this study.

Conclusion

This study estimated the impacts of a district-level policy of including SWD in general education as their default educational placement. The district in which this policy was implemented provides a unique opportunity to rigorously examine this issue, given the staggered, school-level policy implementation and the district's first-order implementation success in moving 90% of all students with IEPs into general education classrooms for the majority of the school day. This study adds to a small body of existing, quasi-experimental research examining inclusion as a form of special education service provision.

Results from this study, drawn from an event study approach, run contrary to existing evidence that when SWD are moved into general education settings, the academics of their peers without disabilities declines (Fletcher, 2010; Robinson, 2012). Findings show that SWOD did no worse on standardized assessments in grades 3 through 8 after the introduction of the inclusion policy than they did previously. For SWD, for whom some observational literature suggests the potential for academic improvement in inclusive classrooms (Dessemontet et al., 2012;

McLeskey et al., 2018), findings from this study are more consistent with other research that finds inclusion to have neutral impacts for these students (Lindsay, 2007; Ruijs & Peetsma, 2009). Similar to their peers without disabilities, findings from this district show the relocation to inclusive classrooms to have had a null effect on the standardized test scores of SWD in grades 3 through 8.

Ancillary measures of academic performance—attendance, dropout, graduation, and promotion rates—reinforce the broader conclusion that this policy did not negatively affect the academics of students in either subgroup in the case study district and may have beget some positive outcomes in the later grades. Attendance rates stayed largely consistent across both student groups all grade bands over time, with some minor (less than one percentage point) fluctuations. While SWD saw a slight increase in high school dropout rates (two percentage points) during the implementation period, this increase did not sustain beyond the four initial years of the policy. Notably, estimates for graduation and promotion rates suggest the potential for *positive*, long-term policy impacts of inclusion for all students after the implementation period concluded. District graduation rates rose nearly three percentage points following the introduction of this policy after implementation concluded ($p < .05$), and the likelihood of promotion from ninth grade rose nearly seven percentage points in the same time period ($p < .05$). More data are needed to understand these latter impacts for specific student subgroups, but both metrics are indicative of a positive influence of inclusion for all students.

There is more to understand about this policy's effectiveness than the available data can convey. One missing component is a better understanding of student behavior, as prior research suggests that it is the challenging behaviors of students with more severe disabilities that are the mechanism underlying their peers' negatively affected academic performance (Carrell et al., 2016; Gottfried, 2014). Information on student discipline referrals or indicators of socioemotional well-being would augment this analysis. Additionally, data limitations preclude observation of the differential impacts of this policy for students across disability classifications. This study does not address this important issue, but future work should offer additional assessments of inclusion's implementation and impact in other contexts.

Acknowledgments

I am grateful to a number of people for their feedback and suggestions throughout the development of this paper, including Sarah Cohodes, Doug Ready, Jeff Henig, Nora Gordon, Aaron Pallas, Shani Bretas, Anna Moyer, Elizabeth Huffaker, and seminar and conference participants at the American Educational Research Association (AERA) and Association of Education Finance and Policy (AEFP) annual conferences. Additionally, support from staff at the district on which this study was based was instrumental, in addition to staff at the nonprofit that supported the district's policy implementation.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by a grant from the American Educational Research Association which receives funds for its "AERA-NSF Grants Program" from the National Science Foundation under NSF award NSF-DRL #1749275. Opinions reflect those of the author and do not necessarily reflect those AERA or NSF.

ORCID iD

Katharine Parham Malhotra https://orcid.org/0009-0000-1723-9446

Notes

1. Prior to 2008, the first three categories were (a) more than 60%, (b) 21% to 60%, and (c) less than 21%. In this study, for analytic continuity, these categories in the years prior to 2008 have been recoded to reflect the newer percentage bands. This does not affect the number of students reported as included, but notably, the thresholds for inclusivity were lower prior to this date.

2. The synthetic comparison group is generated using a propensity-score matching technique—radius matching with replacement, with a radius of

.01—allowing for multiple matches for each treated unit. Estimates from this method are more precise than nearest neighbor alternatives given the larger resulting sample size (Somers et al., 2013). Covariates for propensity score prediction include: school location (e.g., urban or rural), school size, proportion of students in special education, proportion of students receiving free or reduced-price lunch, and proportion of students in each major racial/ethnic subgroup.

3. Following Sun and Abraham (2021), these weights are calculated through an auxiliary regression depending only on the distribution of cohorts and indicators of relative time, using the *eventstudyweights* command in Stata (Sun, 2020).

4. Leads greater than 3 years prior to implementation and lags more than 9 years after the start of implementation were binned. The time horizon is intentionally restricted to 9 years after implementation to allow for the observation of long-term impacts while also limiting the potential influence of additional factors that could occur much later and skew estimates.

5. Because this is a different method of transformation than that used to transform the raw data used in the event study analyses, the magnitude of gaps reported in this figure are not comparable to those in the event study results.

6. This measure is the number of dropouts (students terminating formal education for any reason other than death and not known to enroll in another school or state-approved program) divided by the adjusted student cohort (the number of first-time ninth graders, plus any students who transfer in, minus any who transfer out, emigrate or die during the 4-year period).

7. Graduation rate reflects the number of high school graduates divided by the sum of dropouts for grades 9 to 12 plus the number of high school graduates. This calculation, which is distinct from the calculation of an adjusted 4-year cohort graduation rate, accounts for dropout rates within the measure itself. As such, estimates of policy impacts on dropout rates cannot be directly compared to the estimates of impacts on high school graduation rates overall.

8. Of the 10% of remaining SWD for whom general education was not their LRE, 7% were in general education for a smaller proportion of the school day (less than 80%), in combination with time in resource rooms. Two percent of SWD were in nonpublic placements, and the remaining 1% were in full-time homebound or hospital settings.

References

Affleck, J. Q., Adams, A., Lowenbraum, S., & Madge, S. (1988). Integrated classroom versus resource model: Academic viability and effectiveness. *Exceptional Children*, *54*(4), 339–349.

Angrist, J. D., & Pischke, J. (2015). *Mastering 'metrics: The path from cause to effect*. Princeton University Press.

Ballis, B., & Heath, K. (2019). *The long-run impacts of special education* (EdWorkingPaper No. 19-151). Annenberg Institute at Brown University.

Berry, A., & Gravelle, M. (2018). The benefits and challenges of special education positions in rural settings: Listening to teachers. *The Rural Educator*, *34*(2), 1–13.

Brady, F. (2010). *The influence of inclusion on language arts literacy and math achievement of non-disabled middle school students* [Doctoral dissertation]. Seton Hall University.

Brewton, S. (2005). *The effects of inclusion on mathematics achievement of general education students in middle school* [Doctoral dissertation]. Seton Hall University.

Carrell, S. E., Hoekstra, M., & Kuka, E. (2016). *The long-run effects of disruptive peers* (Working Paper No. 22042). National Bureau of Economic Research.

Crockett, J. B. (2020). Inclusion as an idea and its justification in law. In J. M. Kauffman (Ed.), *On educational inclusion: Meanings, history, issues and international perspectives* (pp. 25–46). Routledge.

Daniel, L., & King, D. (1997). Impact of inclusion education on academic achievement, student behavior and self-esteem, and parental attitudes. *Journal of Educational Research*, *91*, 67–80.

De Chaisemartin, C., & D'Haultfœuille, X. (2018). Two-way fixed effects estimators with heterogenous treatment effects. *American Economic Review*, *110*(9), 2964–2996.

Dessemontet, R. S., Bless, G., & Morin, D. (2012). Effects of inclusion on the academic achievement and adaptive behaviour of children with intellectual disabilities. *Journal of Intellectual Disability Research*, *56*(6), 579–587.

Dorn, S., Fuchs, D., & Fuchs, L. S. (1996). A historical perspective on special education reform. *Theory into Practice*, *35*(1), 12–19.

Fletcher, D. (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.

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

Giordano, G. (2007). *American special education: A history of early political advocacy*. Peter Lang.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, *225*(2), 254.

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

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? *The Review of Economics and Statistics*, *84*(4), 584–599.

Ho, A. D. (2009). Nonparametric framework for comparing trends and gaps across tests. *Journal of Educational and Behavioral Statistics*, *34*(2), 201–228.

Individuals with Disabilities Education Act [IDEA]. (2004). 20 U.S.C. § 1400 et seq.

Jenkins, J. R., Jewell, M., Leicester, N., Jenkins, L., & Troutner, N. M. (1991). Developing a school building model for educating students with handicaps and at-risk students in general education classrooms. *Journal of Learning Disabilities*, *24*(5), 311–320.

Jones, N., & Winters, M. A. (2020). *Are two teachers better than one? The effect of co-teaching on students with and without disabilities* (Working Paper 2020-1). Wheelock Educational Policy Center, Boston University.

King-Sears, M. E., Stefanidis, A., Berkeley, S., & Strogilos, V. (2021). Does co-teaching improve academic achievement for students with disabilities? A meta-analysis. *Educational Research Review*, *34*(2021), 100405.

Kurth, J. A., Morningstar, M. E., & Kozleski, E. B. (2015). The persistence of highly restrictive special education placements for students with low-incidence disabilities. *Research and Practice for Persons With Severe Disabilities*, *39*(3), 227–239.

Lindsay, G. (2007). Educational psychology and the effectiveness of inclusive education/mainstreaming. *British Journal of Educational Psychology*, *77*(1), 1–24.

McDonnell, J., Thorson, N., Disher, S., Mathot-Buckner, C., Mendel, J., & Ray, L. (2003). The achievement of students with developmental disabilities and their peers without disabilities in inclusive settings: An exploratory study. *Education and Treatment of Children*, *26*(3), 224–236.

McLeskey, J., Rosenberg, M. S., & Westling, D. L. (2018). *Inclusion: Effective practices for all students* (3rd ed.). Pearson.

Morgan, P. L., Frisco, M. L., Farkas, G., & Hibel, J. (2010). A propensity score matching analysis of the effects of special education services. *The Journal of Special Education*, *43*(4), 236–254.

National Center for Educational Statistics (NCES). (2019a). *The condition of education: Children and youth with disabilities*. U.S. Department of Education, Institute of Education Sciences. https://nces.ed.gov/programs/coe/indicator/cgg/students-with-disabilities.

National Center for Educational Statistics (NCES). (2019b). *Education demographic and geographic estimates (EDGE)*. ACS-ED District Demographic Dashboard.

Peltier, G. L. (1997). The effect of inclusion of non-disabled children: A review of the research. *Contemporary Education*, *68*(4), 234–239.

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.

Reardon, S. F., Shear, B. R., Castellano, K. E., & Ho, A. D. (2016). *Using heteroskedastic ordered probit models to recover moments of continuous test score distributions from coarsened data* (Working Paper No. 16-02). Stanford Center for Education Policy Analysis.

Robinson, C. M. (2012). *The influence of inclusion on the academic performance of general education students on the New Jersey assessment of skills and knowledge in grades 6, 7, and 8* [Doctoral dissertation]. Seton Hall University.

Roth, J. (2019). *Pre-test with caution: Event study estimates after testing for parallel trends* [Unpublished manuscript]. Department of Economics, Harvard University.

Ruijs, N. M., & Peetsma, T. T. D. (2009). Effects of inclusion on students with and without special educational needs reviewed. *Educational Research Review*, *4*(2009), 67–79.

Salend, S. J., & Duhaney, L. G. (1999). The impact of inclusion on students with and without disabilities and their educators. *Remedial & Special Education*, *20*(2), 114–127.

Sanger, C. S. (2020). Inclusive pedagogy and universal design approaches for diverse learning environments. In C. S. Sanger & N. W. Gleason (Eds.), *Diversity and inclusion in global higher education: Lessons from across Asia*. Palgrave Macmillan.

Schifter, L. A. (2015). Using survival analysis to understand graduation of students with disabilities. *Exceptional Children*, *82*(4), 479–496.

Schifter, L. A., & Hehir, T. (2018). The better question: How can we improve inclusive education? *Education Next*. https://www.educationnext.org/better-question-how-can-we-improve-inclusion-education-response-has-inclusion-gone-too-far/.

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

Schwartz, A. E., Hopkins, B. G., & Stiefel, L. (2019). *The effects of special education on the academic performance of students with learning disabilities* (EdWorkingPaper No. 19-86). Annenberg Institute at Brown University.

Sharpe, M. N., York, J. L., & Knight, J. (1994). Effects of inclusion on the academic performance of classmates without disabilities: A preliminary study. *Remedial and Special Education*, *15*(5), 281–287.

Smith, P. (2007). Have we made any progress? Including students with intellectual disabilities in regular education classrooms. *Intellectual and Developmental Disabilities*, *45*(5), 297–309.

Somers, M., Zhu, P., Jacob, R., & Bloom, H. (2013). *The validity and precision of the comparative interrupted time series and the difference-in-difference design in educational evaluation*. MDRC Working Paper on Research Methodology. https://www.mdrc.org/sites/default/files/validity_precision_comparative_interrupted_time_series_design.pdf

Sun, L. (2020). *EVENTSTUDYWEIGHTS: Stata module to estimate the implied weights on the cohort-specific average treatment effects on the treated (CATTs) (event study specifications)*. Statistical Software Components, Boston College Department of Economics.

Sun, L., & Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, *225*(2021), 175–199.

The School Superintendent's Association. (2017). *Leveling the playing field for rural students*. https://www.aasa.org/docs/default-source/resources/reports/aasa-rural-equity-report-final.pdf.

Trabucco, M. (2011). *The influence of co-taught inclusion on the academic achievement of third grade non-disabled students in mathematics* [Doctoral dissertation]. Seton Hall University.

Tremblay, P. (2012). Comparative outcomes of two instructional models for students with learning disabilities: Inclusion with co-teaching and solo taught special education. *Journal of Research in Special Educational Needs*, *13*(4), 251–258.

Winzer, M. A. (2012). *History of special education: From isolation to integration*. Gallaudet University Press.

Author

KATHARINE PARHAM MALHOTRA, PhD, is a postdoctoral research associate at the University of Virginia School of Education and Human Development. Her research focuses on issues of equity in PK-12 education policies, particularly those affecting the educational experiences of students with disabilities.

Manuscript received May 7, 2023
First revision received November 21, 2023
Second revision received January 25, 2024
Third revision received March 22, 2024
Accepted May 8, 2024