Discussion Paper Series

IZA DP No. 18531

April 2026

Special Education Substantially Improves Learning: Evidence from Three States

Stephanie Coffey
Saint Anselm College

Joshua Goodman
Boston University and IZA@LISER

Amy Ellen Schwartz
University of Delaware

Leanna Stiefel
NYU Wagner Graduate School of Public Service

Marcus A. Winters
Boston University

Yunee H. Yoon
Boston University

IZA@LISER NETWORK
ISSN: 2365-9793

IZA DP No. 18531 April 2026

Special Education Substantially Improves Learning: Evidence from Three States*

Abstract

Special education serves more than one in seven U.S. students yet its causal impact remains understudied. Using longitudinal data from Massachusetts, Indiana, and Connecticut, we estimate the effect of individualized supports with an event-study design that tracks achievement around initial classification. Students' scores decline prior to placement and rise sharply afterward, yielding a consistent V-shaped pattern. Within three years, achievement is 0.2–0.4σ higher than counterfactual trends imply. Gains are similar across disability categories and subgroups, are not driven by testing accommodations, and remain under conservative assumptions. Individualized supports substantially increase learning productivity.

JEL classification

I21, I28, H52, J24

Keywords

special education, human capital, treatment effects, education policy

Corresponding author

Joshua Goodman
edecjosh@bu.edu

* *Acknowledgments*: We thank the Connecticut State Department of Education (CSDE), the Indiana Department of Education (IDOE), and the Massachusetts Department of Elementary and Secondary Education (MADESE) for supplying the necessary data and providing valuable feedback. Meagan Comb and Sidrah Baloch provided essential support for the research-practice partnerships within which this research was conducted. This research was supported in part by a generous grant from the William T. Grant Foundation. We received valuable feedback from participants at the BU Wheelock PREREQ seminar, and the annual conferences of the Association for Policy Analysis and Management and the Association for Education Finance and Policy.

A major challenge in designing public policies to enhance human capital is that individuals differ in their capabilities to access, process, and use information. Special education, the set of individualized supports that public schools provide to students with disabilities so that they can fully access instruction, represents one of the largest and most intensive public investments aimed at building human capital. It serves more than one in seven U.S. students, costs roughly twice as much per-pupil as general education, and accounts for about one-third of federal appropriations to elementary and secondary education (U.S. Department of Education, 2021).

The causal impact of these services on learning remains, however, understudied. Many studies reporting mixed associations between special education participation and achievement do not sufficiently address non-random selection into services, leading to persistent questions about whether special education "works" (Woods et al., 2023; O'Hagan and Stiefel, 2024). While individualized instruction and supports should enhance learning, negative effects could arise if classification exposes students to lower-performing or more disruptive peers (Müller, 2010), if services are implemented ineffectively (Shifrer et al., 2013; Shifrer, 2013), or if students are misclassified.

This paper examines how providing individualized supports to students with disabilities affects skill formation and educational progress. Using longitudinal administrative data from Massachusetts, Indiana, and Connecticut, we estimate the causal effect of entering special education on academic achievement. Our event-study and sensitivity-analysis framework addresses the central empirical challenge that assignment to special education is endogenous to a student's educational trajectory. We exploit within-student variation in the timing of disability classification to compare outcomes in years before and after students first receive services, applying the Rambachan and Roth (2023) framework to relax standard parallel-trends assumptions.

Across all three states, students' achievement trajectories are declining prior to classification and then rise sharply after they begin receiving special education services, with cumulative gains of 0.2–0.4σ within three years. These effects are similar across disability types and student demographics. In Massachusetts, where testing accommodation data are available, results remain nearly identical on assessments taken without accommodations, indicating that the measured gains reflect genuine learning rather than byproducts of accommodations. A complementary difference-in-differences design comparing students entering special education in different grades produces nearly identical estimates. These findings reveal that individualized instruction can substantially improve learning outcomes, offering new evidence on the productivity of targeted human-capital investments and the returns to inclusive educational policy.

This paper makes three key contributions. First, we provide the most comprehensive evidence to date on the causal impact of individualized educational supports on learning for students with disabilities. Recent studies using student fixed-effects designs often find that special education services modestly improve test scores, including improvements of 0.11σ in New York City (Schwartz et al., 2021) and 0.04σ in Kentucky (Hurwitz et al., 2020).[1] These studies assume student outcomes would have remained flat absent services. Hanushek et al. (2002) apply a value-added design to a student fixed effects model, imposing a constant marginal effect in all post-classification years and yielding an average annual improvement of roughly 0.04σ. Our event-study framework relaxes both of these restrictions by explicitly modeling pre-classification trends and estimating dynamic treatment effects in each post-classification year. We reveal small immediate gains that compound over time, yielding substantially larger cumulative effects consistent with prior positive estimates once pre-treatment trajectories and time-varying treatment effects are properly accounted for. Our key contribution is demonstrating that the dynamics of skill recovery after classification are central to understanding the full impact of special education services.

Our findings also align with Ballis and Heath (2021), who exploit exogenous policy variation in Texas to show that withdrawing special-education services reduces educational attainment. Recent work on charter schools provides complementary evidence on service delivery models. Setren (2015) and Imberman and Johnson (2025) show that charter schools are less likely to provide specialized services or time in specialized environments but nonetheless substantially raise the achievement of students with disabilities. These studies suggest that effectiveness may depend on service quality and delivery rather than classification alone. Our results demonstrate that in traditional public school settings where specialized services are maintained, the positive effects of special education classification are substantial, sustained, and grow over time.

[1] An exception is Curran et al. (2021), who found that students previously served for an Emotional Disability scored about 0.07σ higher after being reclassified out of special education.

Second, our findings contribute to broader economic debates about how targeted public investments can enhance human capital formation and reduce inequality. By showing that individualized educational supports produce large and sustained improvements in achievement, the analysis demonstrates that substantial returns can arise from tailoring instructional inputs to diverse learning needs. In this sense, our work is closely related to other literature on educational inputs that successfully tailor instruction to students, such as tutoring and computer-assisted learning. Extensive evidence summarized in meta-analyses (Nickow et al., 2024; Fryer Jr, 2017) and exemplified by recent experimental evaluations (Carlana and La Ferrara, 2025; Guryan et al., 2023; Kraft et al., 2022) demonstrates that tutoring is a particularly effective educational intervention in large part because small group or one-on-one settings allow tutors to customize instruction to the student. Computer-assisted learning also appears effective in meta-analyses (Escueta et al., 2020) and recent interventions (Muralidharan et al., 2019; Bhatt et al., 2024; Oreopoulos et al., 2024) largely because of its ability to tailor curriculum to individual students' needs. Our results suggest that special education services should be thought of similarly, with the highest marginal gains from education spending accruing where instruction is adapted to learners' access needs. More broadly, the evidence implies that inclusive educational policies - those that remove barriers to effective learning - can generate high social returns.

Our final key contribution is to advance empirical methods for evaluating policies with endogenous treatment timing and non-parallel pre-trends. We apply a modern event-study framework that extends beyond conventional fixed-effects designs by allowing treatment effects to evolve flexibly over time. To assess sensitivity to assumptions about counterfactual trends, we implement the Rambachan and Roth (2023) bounding approach, which quantifies the range of treatment effects consistent with alternative trajectories that could have occurred absent special-education placement. This design addresses the central identification concern in earlier studies that students were experiencing systematic declines before classification, making transparent how estimates vary under different plausible scenarios. This approach provides a general template for researchers studying interventions where treatment occurs endogenously along a trending outcome path, a common feature in education, health, and labor-market settings.

1 Conceptual Framework

This section outlines a simple framework for interpreting how individualized educational supports influence human capital accumulation. Let h_{it} denote the human capital of student i at time t, which evolves according to a production function

$$h_{it+1} = f(h_{it}, I_{it}; \theta_i) + \varepsilon_{it}, \tag{1}$$

where I_{it} represents instructional inputs and θ_i captures how effectively student i converts inputs into learning. Students differ in θ_i, reflecting heterogeneity in how they access and process instruction.

In a standard classroom, inputs are optimized for the average learner, so the marginal product of instruction

$$\frac{\partial f}{\partial I}(h_{it}, I_{it}; \theta_i) \tag{2}$$

is lower when θ_i diverges from that norm. Individualized supports (special education) adjust instruction to align with each student's learning needs. Let $\delta_i \in \{0, 1\}$ indicate receipt of supports. Then the productivity of instruction can be written as

$$\frac{\partial f}{\partial I}(h_{it}, I_{it}; \theta_i, \delta_i) = \begin{cases} \phi_0(\theta_i), & \text{if } \delta_i = 0, \\ \phi_1(\theta_i) > \phi_0(\theta_i), & \text{if } \delta_i = 1. \end{cases} \tag{3}$$

When a student's learning needs are unmet ($\delta_i = 0$), the effective return to instruction declines over time as cumulative mismatch grows, producing downward-sloping achievement trajectories. When supports are introduced ($\delta_i = 1$), the production function shifts upward, restoring higher productivity. This dynamic yields the V-shaped pattern observed empirically:

$$\frac{dh_{it}}{dt} = \begin{cases} \rho_0 < 0, & \text{for } t < t^*, \\ \rho_1 > 0, & \text{for } t \geq t^*, \end{cases} \tag{4}$$

where t^* denotes the time of classification.

Note that in a dynamic setting, the effects of individualized supports emerge gradually. Students accumulate instructional mismatch before classification, which depresses the productivity of learning. When supports are introduced, this mismatch decays over time, producing the observed V-shaped recovery rather than an instantaneous jump. In this sense, our empirical pattern reflects a gradual restoration of effective learning rather than a one-time level shift.

2 Special Education Policy

Special education refers to a continuum of services and supports designed to meet the individualized needs of students with disabilities (SWDs). These services address various areas such as academics, communication, and social-emotional development, ensuring students have meaningful access to education alongside their non-disabled peers. The Individuals with Disabilities Education Act (IDEA) establishes the federal framework for identifying, evaluating, and providing special education services. Students who qualify for special education are entitled to an Individualized Education Program (IEP), a legally binding document developed by educators, parents, and specialists, detailing the student's current performance, setting measurable goals, and defining the specific services or accommodations required. The goal of this process is to ensure that instructional strategies and resources are tailored to each student's unique challenges. The student's IEP team annually evaluates progress toward goals and assesses the appropriate level of services.

Students are typically first referred for assessment of a disability by a parent, teacher, or school administrator. A school psychologist or special education teacher then evaluates the student's needs. Eligibility is formally determined by a multidisciplinary team that includes parents and qualified professionals, with the evaluator's assessment as a key input. Special education encompasses thirteen qualifying disabilities. Some disability categories have objective physical eligibility criteria (e.g. Hearing Impairment, Orthopedic Impairment, Visual Impairment, and Traumatic Brain Injury), more than 70% of SWDs nationwide fall within categories of Emotional Disturbance, Specific Learning Disability, Other Health Impairment, or Speech or Language Impairment that depend more on the evaluator's subjective assessment (National Center for Education Statistics, 2023a).

Importantly, eligibility for special education requires not only a qualifying disability but also evidence that it adversely affects educational performance and necessitates specially designed instruction. Students entering special education across disability categories are thus unified by a common margin: a documented disability that materially constrains academic progress in the general education setting.

Though the services specified in each student's IEP vary across and within disability categories, they overlap along instructional margins that plausibly affect academic outcomes. Classification typically increases individualized instruction, reduces group size, introduces accommodations that improve access to grade-level material, and formalizes progress monitoring. Students with specific learning disabilities more often receive direct academic intervention, while those classified with communication, health, or emotional disabilities frequently receive services—such as language support, executive-function accommodations, or behavioral interventions—that improve access to instruction and increase effective learning time. Because standardized tests aggregate performance across multiple skills and access channels, this heterogeneity in services may nonetheless produce fairly consistent achievement impacts across classifications.

In addition to supplemental services, SWDs may be eligible to receive accommodations on standardized tests to ensure that the assessments measure their knowledge and skills rather than the impact of their disability. These accommodations can include extended time, alternate testing locations, additional breaks, the use of assistive technology, or alternative formats (e.g., Braille, large print). Their primary purpose is to create an equitable testing environment that acknowledges the individual's unique needs without altering the exam content or reducing its rigor.

About 17%, 18% and 20% of public school students in Connecticut, Indiana, and Massachusetts, respectively, are enrolled in special education – each above the national average of 15%. Since 2000-01, the number of students receiving special education services has grown by 20% nationwide, 17% in Connecticut, 12% in Massachusetts, and 20% in Indiana (National Center for Education Statistics, 2023b). All three states adhere to the federal framework for eligibility determination under the IDEA, with some implementation differences. The states also vary somewhat in their special education funding models.[2]

3 Data and Sample

3.1 Data Sources

We use statewide longitudinal administrative data on the universe of public school students in Indiana, Massachusetts, and Connecticut. Our Indiana data, from the Indiana Department of Education (IDOE), span the 2011-12 through 2017-18 school years. Our Connecticut data, from the Connecticut Department of Education (CTDOE), span the 2013-14 through 2019-20 school years. Our Massachusetts data, from the Massachusetts Department of Elementary and Secondary Education (MADESE), span the 2007-08 through 2017-18 school years. Unique student identifiers in each data system allow students to be tracked over time. In each state we observe student demographic characteristics including race/ethnicity, gender, and free or reduced price lunch eligibility (FRPL).

Our treatment variable is an indicator for whether a student is classified as entitled to special education services. Each state's data contain annual measures of disability classification, allowing us to identify the first year that a student is classified as having a disability. Such variables also allow us to observe each student's specific disability classification. We focus our analysis on four categories of disability for which we observe a sufficient number of new classifications during the observation period: Specific Learning Disability (such as dyslexia or auditory processing disorder), Emotional Disability (such as anxiety or depression), Other Health Impairment (such as ADHD or asthma), and Speech/Language Disability (such as stuttering or articulation disorder).

Our primary outcomes are Mathematics and English Language Arts scores from annual state standardized assessments taken in grades three through eight. For each state, we use the full statewide distribution to standardize scores by year, grade and subject, so that effect sizes can be interpreted as student-level standard deviations. In Massachusetts, we also observe whether the

[2]See Online Appendix B for more details.

student has received an accommodation on a given test, such as having the test read aloud or having access to a calculator. In both Massachusetts and Indiana, we also observe and use as an outcome the percentage of days the student attended during the school year.

3.2 Sample Construction

Each observation represents a student in an academic year. We include only observations in grades 3 through 8, when standardized testing occurs. The analytic sample includes observations only from students who we observe at least once with and once without an IEP. We further restrict the sample to include only observations within 4 years prior to or following a student's initial disability classification. Because we observe test scores in grades 3 through 8, we can only analyze the effect of initial classification in grades 4 through 8. Thus, the number of years that we could potentially observe the student's test scores as an outcome before and after initial placement into special education differs by the grade level in which they were first classified with a disability.[3]

Our analysis largely excludes three types of students: those with severe disabilities, those placed into private schools for support services, and those taking alternative assessments. Students with severe disabilities are more likely to have their needs identified and placed into special education in early elementary grades, so that initial classification happens too early for us to have pre-treatment test score outcomes to measure. We do not observe test scores and thus do not include in the test score analyses SWDs who take an alternative assessment because they are deemed unable to take the standard test even with accommodations. Federal law restricts eligibility for alternative assessments to students with the most significant cognitive disabilities, so such students represent a small, more severely disabled subset of the special education population and an even smaller share of students initially classified as late as the fourth grade. These students are, however, included when analyzing impacts on attendance rate. Finally, because our data does not contain private schools, our sample does not contain students with such severe disabilities that they have obtained a private placement.

[3] In the Online Appendix, we also describe the results from models that restrict the sample according to the grade level in which the student was first classified.

Overall, students whose classification status changes are observationally similar to those who we always observe with a disability, and these groups have lower test scores and are more likely to be eligible for FRPL than the full student population.[4] Test scores for SWDs generally lag those of their general education peers but students classified with a speech/language disability score roughly in line with the state average in both ELA and math. Students with a health impairment, emotional disability, speech/language disability or other IEP are disproportionately male, while students with a Specific Learning Disability are split roughly equally between male and female. The specific disability classifications also differ by the grade in which they are typically identified. For instance, students with a speech/language disability are more frequently identified in elementary grades, while those with an emotional disturbance or other health impairment are more likely to be identified in middle school grades. These differences suggest that the effect of special education services on student outcomes may also vary between disability classifications, which we explore in heterogeneity analysis that follows.

4 Measuring the Effect of Special Education on Student Outcomes

The central challenge with measuring the causal effect of special education services on student outcomes is that both the existence of a disability and the timing of a student's placement into special education are likely associated with unobserved factors that also impact educational outcomes. Our primary analyses use different strategies to address potential bias from fixed and dynamic sources of selection. We eliminate potential bias from fixed student attributes by controlling for student fixed-effects to make within-student comparisons that leverage variation in the timing of assignment into special education. We address potential for selection bias due to dynamic factors by generating confidence intervals under alternative transparent assumptions for the underlying relationship between pre- and post-classification counterfactual trend.

[4]Tables A.1, A.2 and A.3 in the Online Appendix report summary statistics for each respective state.

4.1 Student Outcomes Relative to Initial Classification

The foundation for our approach is an event-study model tracking the trajectory of within-student outcomes during years leading to and following their initial assignment into special education. We estimate an event study regression that includes a student fixed effect (θ_i) and indicators for each year relative to the student's initial placement into special education (D_j). The goal is to explain a given outcome for student i during school-year t (y_{it}). Time period $t = 0$ represents the first year that a student receives special education services, and we treat the period immediately preceding the student's classification ($t = -1$) as the omitted comparison group.

$$y_{it} = \sum_{t=-4}^{-2} \beta_j D_j + \sum_{t=0}^{3} \beta_j D_j + \theta_i + \epsilon_{it} \tag{5}$$

The pattern of student test scores relative to initial classification is strikingly similar across states and subjects. Figure 1 illustrates the coefficients and 95% confidence intervals resulting from estimating Equation 5. Students experience a clear downward trajectory in both their ELA and math test scores as they approach classification, with achievement dropping by roughly 0.02-0.04σ per year in the pre-period. Scores then turn sharply upwards beginning the first year they receive special education services, often with the largest increases in that first year but with consistent continuing growth through four years after classification.

Student attendance rates follow a somewhat similar pattern in Massachusetts, though not in Indiana. In Massachusetts, attendance rates decline substantially in the years leading up to classification, then plateau in the post-classification period. Attendance rates in Indiana are the only outcome that neither exhibits a declining pre-trend nor a substantial shift post-classification.

These negative pre-classification trends, particularly in test scores, occur in the opposite direction of the anticipated treatment effect. In other contexts, pre-trends can be concerning because their continuation into the post-treatment period can create the illusion of a treatment effect that does not exist. Here, however, continuation of the pre-trends into the counterfactual post-treatment period would lead a conventional event-study or difference-in-differences approach to understate an existing positive treatment effect. The case that placement into special education has a positive causal effect on student outcomes is strengthen by the “v-shaped” patterns of student achievement (or “l-shaped” in the case of attendance), hinged at the time a student is initially classified with a disability.

Believing that secular trends explain the entirety of the treatment effect suggested by the patterns illustrated on Figure 1 requires one to believe that school systems systematically classify students who are experiencing a downward trend in outcomes into special education precisely at the time when their outcome would have otherwise plateaued or even turned sharply upwards. There is no empirical or theoretical justification to support such an expectation. Rather, the observed downward trend in outcomes leading to initial classification is consistent with the idea, formalized in Section 1, that students are often referred to be assessed for a disability precisely because they are struggling in school, falling further behind each year that they do not receive the supports they require to fully access instruction in light of their yet-undiagnosed disability.

4.2 Accounting for Underlying Trends in the Counterfactual

Quantifying the causal effect of special education requires projecting the counterfactual outcome students would have achieved absent classification. Most prior work assumes no secular trend in post-classification outcomes — the dashed horizontal line at zero on Figure 1. Both theory and the empirical patterns in Figure 1 suggest, however, that this assumption is exceptionally conservative, as students are clearly on declining trajectories at the time of classification.

A more plausible approach allows that the pre-treatment trend contains information useful in predicting the likely trajectory for the counterfactual post-treatment trend. We decompose the reduced-form event-study coefficients β into the sum of treatment effects τ and underlying trends δ.

$$\beta = \tau + \delta \tag{6}$$

Assuming no treatment effect prior to classification ($\tau_{pre} = 0$), the key question is the relationship between pre- and post-classification trends. A standard event-study assumes $\delta_{pre} = \delta_{post} = 0$, yet here both theory and data suggest $\delta_{pre} < 0$. The dotted line on Figure 1 illustrates the counterfactual under the assumption that the linear pre-classification trend would have continued into the post-period.

Rather than imposing that $\delta_{pre} = \delta_{post}$ exactly — which may itself be a strong assumption given measurement error in the pre-trend and the possibility that factors beyond undiagnosed disability contribute to pre-classification declines (Wolfers, 2006; Kahn-Lang and Lang, 2020; Lee and Solon, 2011) — we adopt the more flexible bounding approach of Rambachan and Roth (2023). Formally, we impose $\delta \in \Delta^{SD}(M)$ for $M \leq 0$, where

$$\Delta^{SD}(M) := [\delta : |(\delta_{t+1} - \delta_t) - (\delta_t - \delta_{t-1}| \leq M, \forall t] \quad (7)$$

The parameter $M \geq 0$ bounds the amount by which the slope of δ can change across consecutive periods; $M = 0$ coincides with exact linearity, while larger values relax this assumption, producing wider confidence sets.

We take a "Goldilocks" approach to setting M. For test scores, we report results under $M = 0$ (linear trend), $M = 0.02$, and $M = 0.04$; for attendance rates (measured on a different scale), we use $M = 0$, $M = 0.01$, and $M = 0.02$. The most conservative test score assumption ($M = 0.04$) allows the secular trend to shift across consecutive periods by nearly as much as what Kraft (2020) classifies as a "medium" treatment effect (0.05σ) in education research.

5 Results

5.1 Full Sample Estimates

We show how alternative assumptions for the counterfactual trend change our estimates for the effect of receiving special education services on student test scores and attendance rates, for the initial classification year and for the two subsequent years. Figure 2 illustrates the range of estimates resulting from three different assumptions. The first (leftmost) set of estimates in each sub-figure assume a flat counterfactual trend, equivalent to the respective coefficient estimates illustrated on Figure 1. The second set of estimates assume that the linear approximation of the pre-trend would have continued into the post-period, equivalent to the difference between the coefficient estimates in Figure 1 and the illustrated gray post-trend. The third and fourth sets of estimates apply more conservative assumptions for the amount by which that trend may deviate across consecutive periods.

Assuming that pre-trends would have continued linearly into the post-period substantially raises the apparent impact of special education classification on student achievement, particularly as time passes. Comparing the first and second sets of estimates in each panel in the top two rows, we see that by two years after initial classification, ELA achievement improves 0.1-0.2σ relative to the last pre-classification year but by 0.2-0.3σ relative to pre-trend. Similarly, math achievement two years after classification improves 0.2-0.3σ relative to the last pre-classification year but by 0.3-0.4σ relative to pre-trend. All of these effects are economically meaningful, highly statistic significant, and estimated with high precision.

The confidence intervals grow as we further relax the assumption that the post-treatment trend would replicate the pre-treatment trend if the student were not classified in $t = 0$. In analyses where we allow $M > 0$, the confidence intervals also increase as we consider impacts in years further from the initial classification, reflecting the accumulating influence of the potential annual deviation from the trend over time. Nonetheless, we find significant positive impacts on ELA and math scores after three years even under the implausibly conservative assumption that the post-classification trend in test scores could differ by up to 0.04σ across consecutive periods. In both Indiana and Massachusetts our “middle” specification can reject an effect size in the third post-classification year ($t = +2$) that is smaller than 0.18σ in ELA and smaller than 0.20σ in math.

The base model ($M = 0$) that assumes a linear projection of the pre-classification trend into the post-classification period also finds a significant positive effect on student attendance rate that grows over time. In contrast to the estimates for student test scores, even small deviations from the pre-treatment trend leads to confidence intervals that can not reject zero impact on the attendance rate.

5.2 Investigating Heterogeneous Impacts

Meaningfully large, positive impacts of special education services on student achievement appear across nearly all disability classifications and state contexts. Figure 3 illustrates for ELA the effects of initial classification according to the particular nature of the student's disability, with focus on categories for which we observe a sufficiently large number of new classifications. We find qualitatively similar impacts for students classified with a Specific Learning Disability, Emotional Disability, Other Health Impairment, or Speech/Language Disability. Math effects also appear broadly across most disability classifications and state contexts.[5] Similarity in achievement impacts across disability categories is perhaps less surprising that it first appears, given both the information contained in test scores and the fact that eligibility depends on having a documented disability that impedes access to instruction. There may also be disability-specific differences in the trajectories that lead to such estimated treatment effects on achievement.[6] Positive attendance effects observed in Massachusetts appear driven by students with Emotional Disability.[7]

Special education services also appear beneficial regardless of student socioeconomic status or race/ethnicity. For both ELA and math achievement, estimated impacts of intial classification are large, positive, and generally statistically significant across state contexts for FRPL-eligible and FRPL-ineligible students, and for White, Black, and Hispanic students.[8] Though the magnitudes of estimated positive effects vary somewhat by demographic group and state context, they are rarely if ever statistically distinguishable across groups, suggesting that the benefits of special education services accrue to students of all types. Positive attendance effects in Massachusetts similarly appear across students of different socioeconomic status and race/ethnicity.[9]

The benefits of special education services also appear across students classified in different initial grades and across schools with different levels of urbanicity. Across nearly every state context, positive achievement effects in ELA and math appear for those first classified in fifth grade, sixth grade, seventh grade, and eighth grade, with the earlier grades showing somewhat

[5] See Figure A.1 for details.
[6] See Figures A.2 and A.3 for details.
[7] See Figure A.4 for details.
[8] See Figures A.5 and A.6 for details.
[9] See Figure A.7 for details.

larger effects.[10] Special education benefits also appear large and remarkably similar across schools in urban, suburban, and rural locations.[11]

5.3 Controlling for Test Accommodations

An important contribution of this paper over previous studies is our ability to control for the use of accommodations on statewide standardized tests within the Massachusetts data. In Massachusetts, allowable accommodations include having the test read aloud (either by text to speech or a human reader), the use of supplemental reference sheets or graphic organizers, and the use of a calculator, among others. Improvement in student test scores following disability classification may reflect the combined impacts of testing accommodations along with the causal benefits of special education services. In Massachusetts, we can address this issue by incorporating controls for whether the student received a particular accommodation on the test in a given year and subject.

Testing accommodations do not account for our estimated effects of special education services on achievement. Table 1 reports results from our primary regressions and from regressions that incorporate different controls for test accommodations for students in Massachusetts. Receipt of an accommodation is significantly and substantially related to student test scores. Controlling for receipt of a test accommodation does not, however, meaningfully alter the coefficients on years leading to or following disability classification. This pattern suggests that the timing of a student's receipt of accommodations is sufficiently independent from initial disability classification that failure to account for them does not meaningfully bias our primary estimate as the causal effect of special education on student test scores.

6 Alternative Comparison: Difference-in-Differences Approach

To assess the robustness of our results, we explore an alternative strategy that relies on across-student comparisons. Our preferred approach estimates the causal effect of special education by

[10] See Figures A.8 and A.9 for details.
[11] See Figures A.10 and A.11 for details.

comparing outcomes for the same student before and after initial classification. We prioritize this within-student design due to concerns about unobserved heterogeneity across students, which are especially pronounced when analyzing outcomes for SWDs. It is difficult to construct a credible comparison group from students who are never classified, as these students likely differ in meaningful but unobserved ways from those who enter special education. A potential alternative is to compare students newly classified in a given grade to those who will eventually be classified but have not yet entered special education. We do so in a difference-in-differences framework, following Callaway and Sant'Anna (2021).[12]

Results from this difference-in-differences approach comparing treated students and not-yet-treated students are remarkably similar to those from our within-student analyses. Figure 4 illustrates these results in an event-study framework. Because this analysis compares a treated cohort to not-yet treated students on a similar downward trajectory, we do not observe a downward trend in outcomes as the student approaches initial placement into special education. Consistent with our within-student comparisons, we observe a sharp increase in test scores for the treated students relative to not-yet treated students occurring in the first year they are enrolled in special education and continuing upward in the years after. Estimates are very similar in magnitude to the respective within-student comparison results by state and subject (in Figure 2) from models that incorporate continuation of the student's pre-treatment trend. As in the within-student model, effects also appear across students initially classified in different grades.[13] In the difference-in-differences model, attendance effects also appear positive and meaningful in both Massachusetts and Indiana, a result also consistent across students first classified in different grades.[14]

The difference-in-differences analyses yield two key takeaways. First, the consistency of results from another comparison group increases confidence that our primary within-student estimates reflect the causal effect of special education on student outcomes. Second, the similarity in the magnitude of the difference-in-differences estimates and those from the student fixed-effects models that incorporate trend continuation into the post-classification period strengthens the case

[12] See Online Appendix C for details.
[13] See Figures A.12 and A.13 for details.
[14] See Figure A.14 for details.

for incorporating prior outcome trends into the counterfactual trajectory in our primary analyses.

7 Conclusion

This paper provides new causal evidence that human capital accumulation is improved by individualized educational supports for students with disabilities. Using longitudinal administrative data from Massachusetts, Indiana, and Connecticut, we document that students' achievement trajectories decline in the years leading up to classification for special education and rise sharply thereafter. Estimates that allow for dynamic treatment effects and relax conventional parallel-trends assumptions reveal substantial and sustained learning gains, suggesting that the productivity of public spending on individualized supports is higher than previously believed. Our estimates reflect the causal effect of special education services for students initially classified in grades four through eight, a parameter highly relevant to both the research literature and public policy.

The results imply that heterogeneity in the returns to educational investment is large: students whose learning needs diverge from standard instruction experience low marginal productivity in the absence of individualized supports but large gains once instruction is adapted. In dynamic terms, the evidence points to a gradual restoration of learning productivity as instructional mismatch diminishes—an adjustment process that echoes other contexts where targeted public investments repair or re-optimize existing human capital. These findings speak to broader questions about how governments can design education systems that are both equitable and efficient. Policies that expand access to individualized instruction may yield high social returns by converting previously underproductive investment into effective learning.

The explicit goal of special education is to provide students with individualized supports tailored to meet their specific needs, and thus the treatment of special education services is by its nature heterogeneous. Nonetheless, it is likely that some common components of special education services are more effective than others. Though our data does not allow us to distinguish efficacy of different service components, future research might do so by taking advantage of information contained in students' IEPs. Though IEPs are individualized by definition, they contain several common aspects such as specifying the delivery model by which the student is to receive certain instruction (self-contained classroom, pull-out services, co-taught general education classroom). Future work might explore how the size and persistence of learning gains is affected by service delivery model, as well as the quality and intensity of services provided.

Data and Replication

The analyses use restricted administrative data from the Massachusetts Department of Elementary and Secondary Education (MADESE), the Indiana Department of Education (IDOE), and the Connecticut State Department of Education (CTDOE). Access to these data requires a data-use agreement with each agency. Upon publication, replication code and detailed documentation will be made available through the Harvard Dataverse.

References

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

Bhatt, M. P., Guryan, J., Khan, S. A., LaForest-Tucker, M., and Mishra, B. (2024). Can technology facilitate scale? evidence from a randomized evaluation of high dosage tutoring. Technical report, National Bureau of Economic Research.

Callaway, B. and Sant'Anna, P. H. C. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2):200–230.

Carlana, M. and La Ferrara, E. (2025). Apart but connected: Online tutoring, cognitive outcomes, and soft skills. *American Economic Review*, 115(10):3487–3513.

Curran, F. C., Bal, A., Goff, P., and Mitchell, N. (2021). Estimating the relationship between special education de-identification for emotional disturbance and academic and school discipline outcomes: Evidence from Wisconsin's longitudinal data. *Education and Urban Society*, 53(1):83–112.

Escueta, M., Nickow, A. J., Oreopoulos, P., and Quan, V. (2020). Upgrading education with technology: Insights from experimental research. *Journal of Economic Literature*, 58(4):897–996.

Fryer Jr, R. G. (2017). The production of human capital in developed countries: Evidence from 196 randomized field experiments. In *Handbook of economic field experiments*, volume 2, pages 95–322. Elsevier.

Guryan, J., Ludwig, J., Bhatt, M. P., Cook, P. J., Davis, J. M., Dodge, K., Farkas, G., Fryer Jr, R. G., Mayer, S., Pollack, H., et al. (2023). Not too late: Improving academic outcomes among adolescents. *American Economic Review*, 113(3):738–765.

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

Hurwitz, S., Perry, B., Cohen, E. D., and Skiba, R. (2020). Special education and individualized academic growth: A longitudinal assessment of outcomes for students with disabilities. *American Educational Research Journal*, 57(2):576–611.

Imberman, S. A. and Johnson, A. S. (2025). The effect of charter schools on identification, service provision, and achievement of students with disabilities. *Education Finance and Policy*, pages 1–65.

Indiana Department of Education (2023). Article 7: Rules for Special Education. Accessed: 2025-02-04.

Indiana Department of Education (2024). Special Education Funding in Indiana. Accessed: 2025-02-04.

Kahn-Lang, A. and Lang, K. (2020). The promise and pitfalls of differences-in-differences: Reflections on 16 and pregnant and other applications. *Journal of Business and Economic Statistics*, 38(3):613–620.

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

Kraft, M. A., List, J. A., Livingston, J. A., and Sadoff, S. (2022). Online tutoring by college volunteers: Experimental evidence from a pilot program. In *AEA Papers and Proceedings*, volume 112, pages 614–618. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

Lee, J. Y. and Solon, G. (2011). The fragility of estimated effects of unilateral divorce laws on divorce rates. *The B.E. Journal of Economic Analysis and Policy*, 11(1).

Massachusetts Department of Elementary and Secondary Education (2023a). 603 CMR 28.02: Special Education Eligibility Criteria. Accessed: 2025-02-04.

Massachusetts Department of Elementary and Secondary Education (2023b). Massachusetts Special Education Funding and Circuit Breaker Program. Accessed: 2025-02-04.

Müller, C. M. (2010). Negative peer influence in special needs classes: A risk for students with problem behaviour? *European Journal of Special Needs Education*, 25(4):431–444.

Muralidharan, K., Singh, A., and Ganimian, A. J. (2019). Disrupting education? experimental evidence on technology-aided instruction in india. *American Economic Review*, 109(4):1426–1460.

National Center for Education Statistics (2023a). Digest of education statistics, table 204.30: Children 3 to 21 years old served under Individuals with Disabilities Education Act (idea), Part B, by type of disability: Selected years, 1976-77 through 2021-22. Accessed: 2025-02-04.

National Center for Education Statistics (2023b). Table 204.70. Percentage distribution of students enrolled in grades 1 through 12, by public school type: Selected years, Fall 1999 through Fall 2021. Accessed: 2025-02-05.

Nickow, A., Oreopoulos, P., and Quan, V. (2024). The promise of tutoring for prek–12 learning: A systematic review and meta-analysis of the experimental evidence. *American Educational Research Journal*, 61(1):74–107.

O'Hagan, K. G. and Stiefel, L. (2024). Does special education work? A systematic literature review of evidence from administrative data. *Remedial and Special Education*.

Oreopoulos, P., Gibbs, C., Jensen, M., and Price, J. (2024). Teaching teachers to use computer assisted learning effectively: Experimental and quasi-experimental evidence. Technical report, National Bureau of Economic Research.

Rambachan, A. and Roth, J. (2023). A more credible approach to parallel trends. *Review of Economic Studies*, 90(5):2555–2591.

School and State Finance Project (2024). Excess cost grant: Connecticut's method for sharing extraordinary special education costs. Report.

Schwartz, A. E., Hopkins, B. G., and Stiefel, L. (2021). The effects of special education on the academic performance of students with learning disabilities. *Journal of Policy Analysis and Management*, 40(2):480–520.

Setren, E. (2015). Special education and english language learner students in boston charter schools: Impact and classification. *Cambridge, MA: Massachusetts Institute of Technology*, pages 1073–1112.

Shifrer, D. (2013). Stigma of a label: Educational expectations for high school students labeled with learning disabilities. *Journal of Health and Social Behavior*, 54(4):462–480.

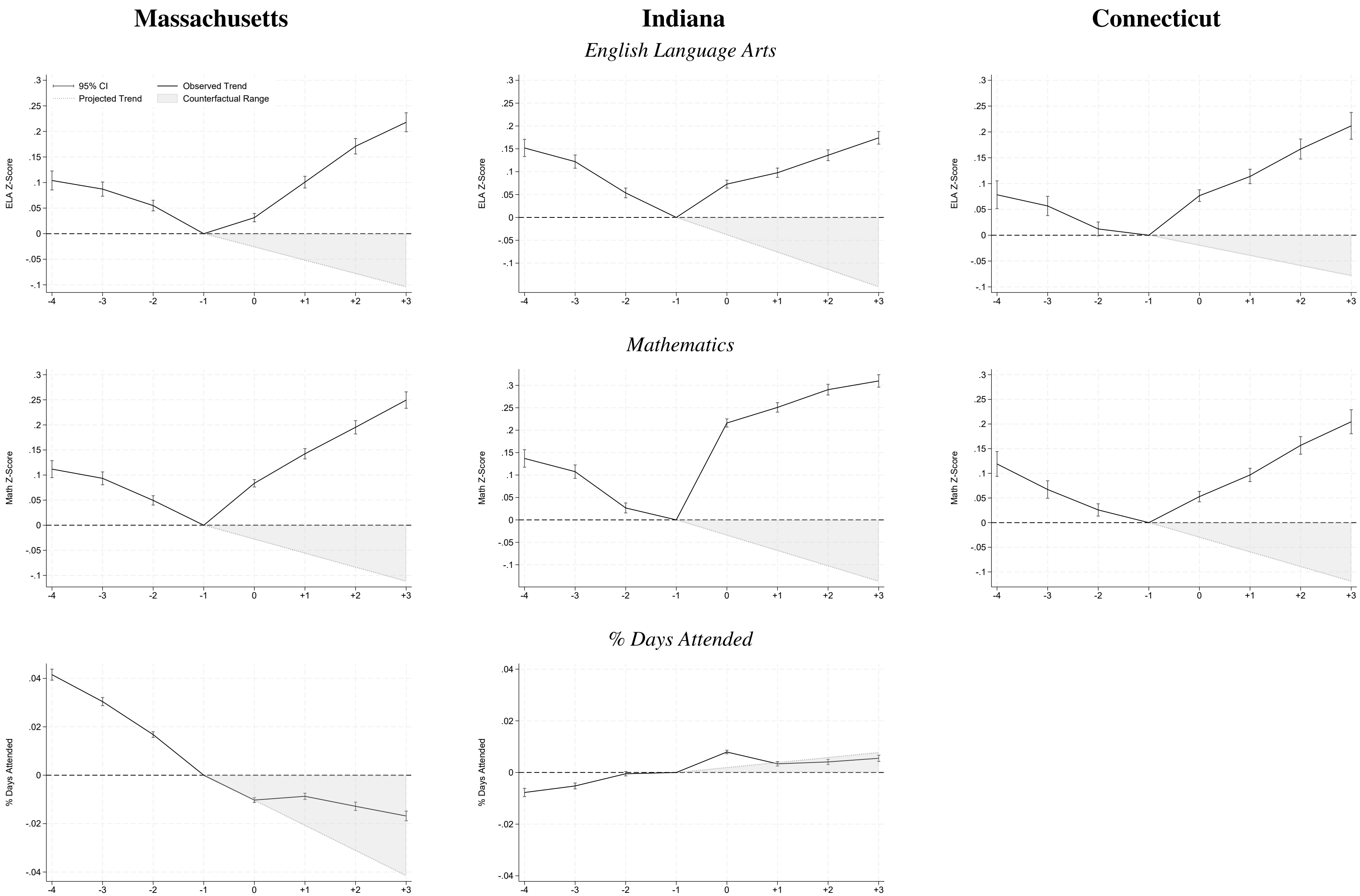
Shifrer, D., Callahan, R. M., and Muller, C. (2013). Equity or marginalization? The high school course-taking of students labeled with a learning disability. *American Educational Research Journal*, 50(4):656–682.

U.S. Department of Education (2021). U.S. Department of Education Budget History Table: Fiscal Years 1980–2021.

Wolfers, J. (2006). Did unilateral divorce laws raise divorce rates? A reconciliation and new results. *American Economic Review*, 96(5):1802–1820.

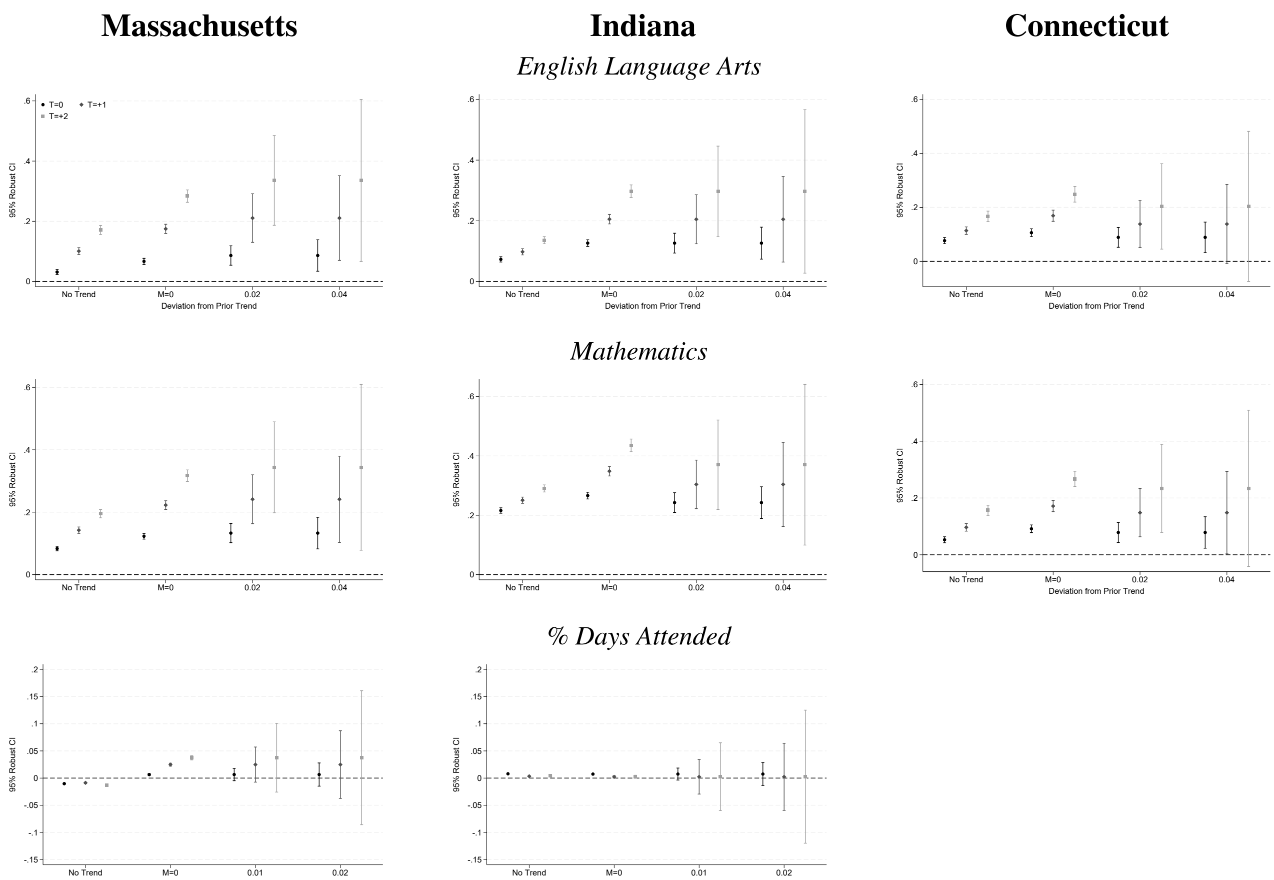
Woods, A. D., Gloski, C., Wang, Y., and Morgan, P. (2023). A multivariate meta-analysis of the effects of business-as-usual special education services in U.S. schools. *PsyArXiv Working Paper*.

Figure 1: Event Study Results: Illustrating Pattern of Outcomes Relative to Initial Classification



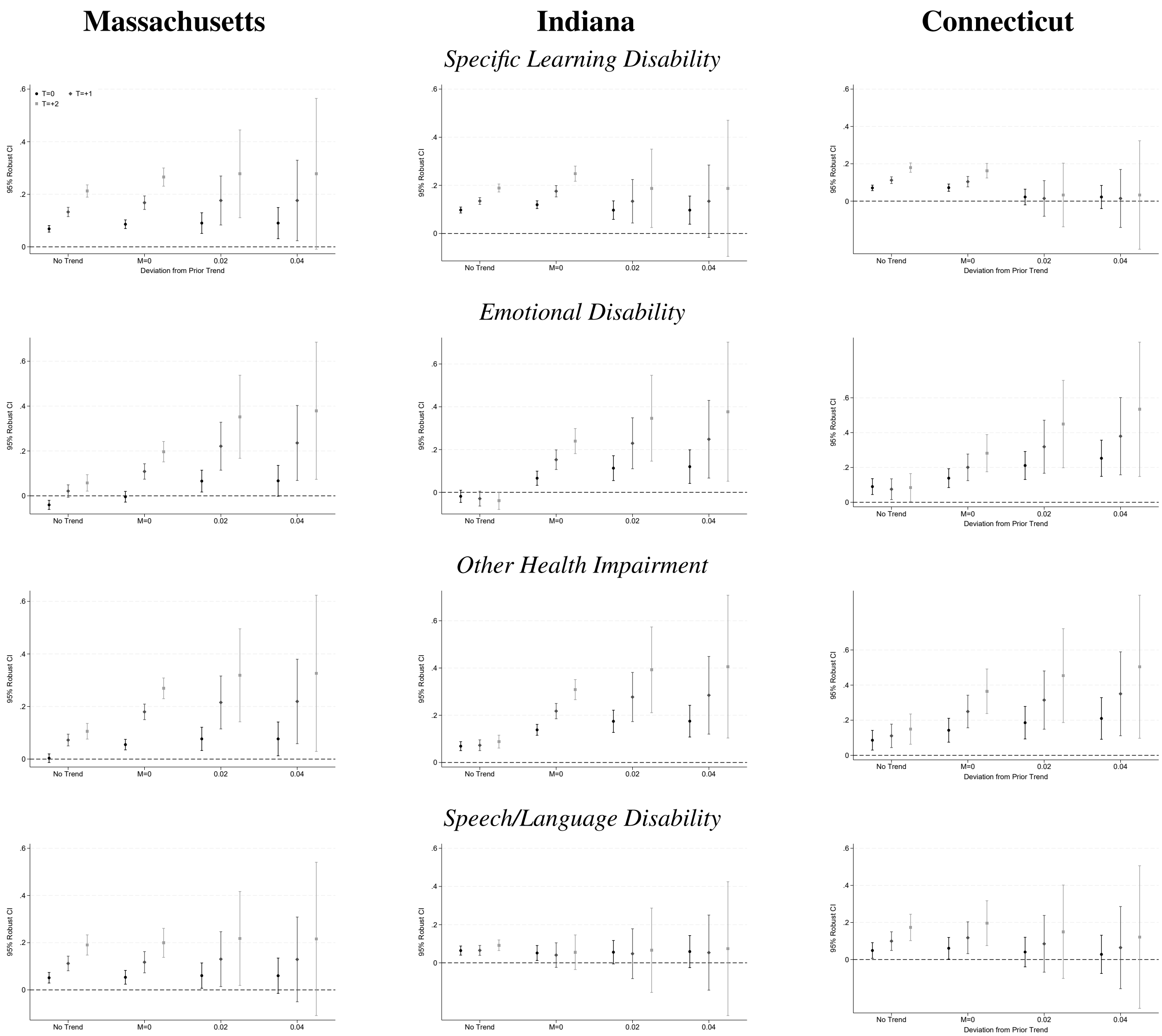
Notes: The figure above illustrates coefficients and 95% confidence intervals from regressions with student and school fixed effects. The X-axis describes years relative to initial classification, with T=0 representing the student's first year with an IEP. The sample includes observations within 4 years of initial disability classification for students observed with and without an IEP during the sample period. The dotted line illustrates continuation of the linear pre-treatment trend. Robust standard errors are clustered by student.

Figure 2: Estimated Effect of Disability Classification Allowing for Deviation from Prior Trends



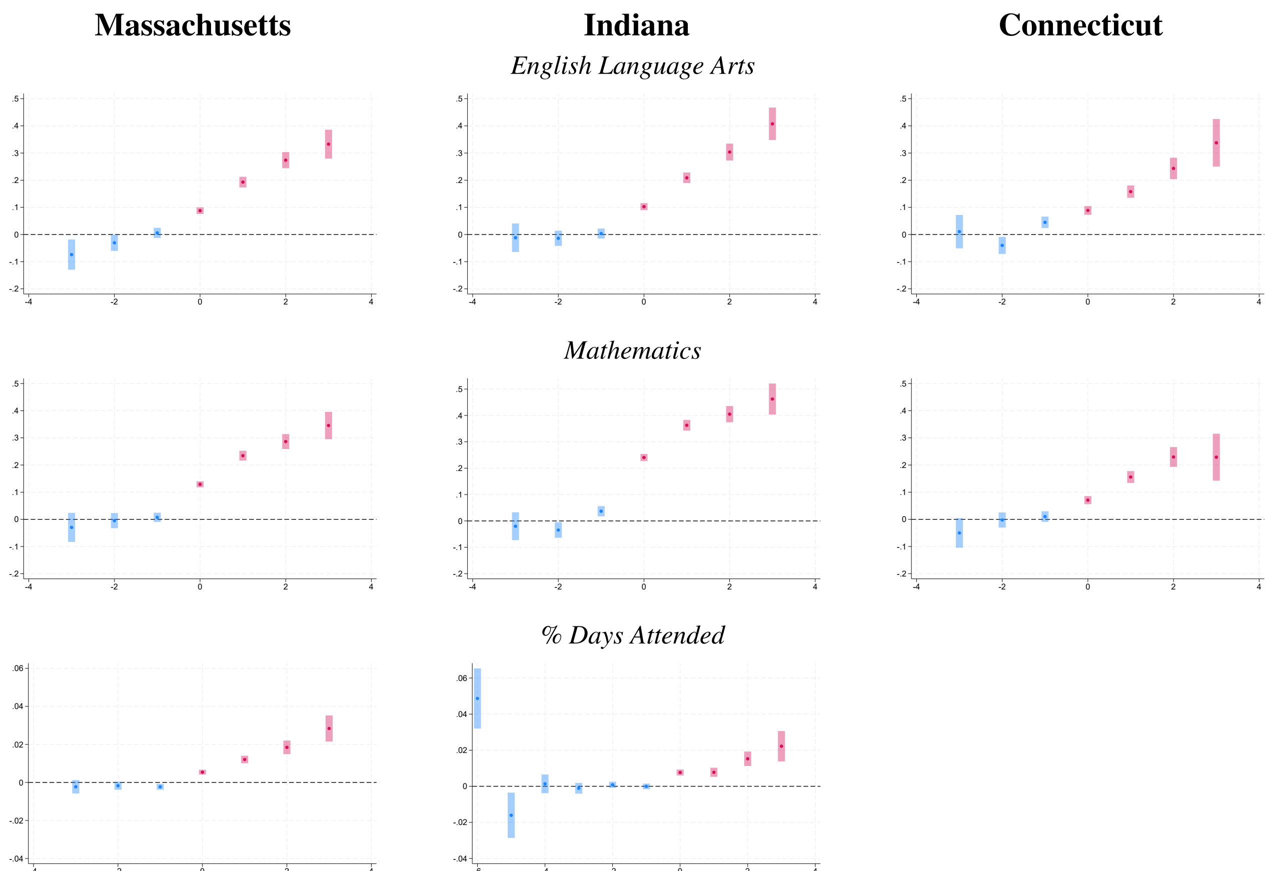
Notes: The figure above illustrates 95% confidence intervals resulting from the HonestDiD approach to the full sample. Each panel illustrates estimates for effects during the initial classified year (T=0) and two subsequent years. The X-axis groups estimates according to the specified amount the model allows the post-treatment counterfactual trend to annually deviate from the linear pre-treatment trend. The first set of results assumes a flat counterfactual trend. The second set of estimates, M = 0, represents results where the counterfactual is the linear projection of the pre-treatment trend. Estimation samples include observations within 4 years of initial disability classification for students observed with and without an IEP during the sample period. The allowed deviation amounts differ by outcome to align with the appropriate scale.

Figure 3: Effect of Classification by Select Classifications: ELA



Notes: The figure above illustrates 95% confidence intervals resulting from the HonestDiD approach within samples restricted to students initially classified with a given disability category. Each panel illustrates estimates for effects during the initial classified year (T=0) and two subsequent years. The X-axis groups estimates according to the specified amount the model allows the post-treatment counterfactual trend to annually deviate from the linear pre-treatment trend. The first set of results assumes a flat counterfactual trend. The second set of estimates, M = 0, represents results where the counterfactual is the linear projection of the pre-treatment trend. Estimation samples include observations within 4 years of initial disability classification for students observed with and without an IEP during the sample period.

Figure 4: Difference-in-Differences Estimates



Notes: The figure above illustrates coefficients and 95% confidence intervals from difference-in-differences regressions comparing treated students to not-yet treated cohorts. The X-axis describes the number of grade levels relative to initial classification, with 0 representing the student's first grade with an IEP. Robust standard errors are clustered by student.

Table 1: Impact of Accounting for Accommodations on Event Study Estimates from Massachusetts: ELA & Math

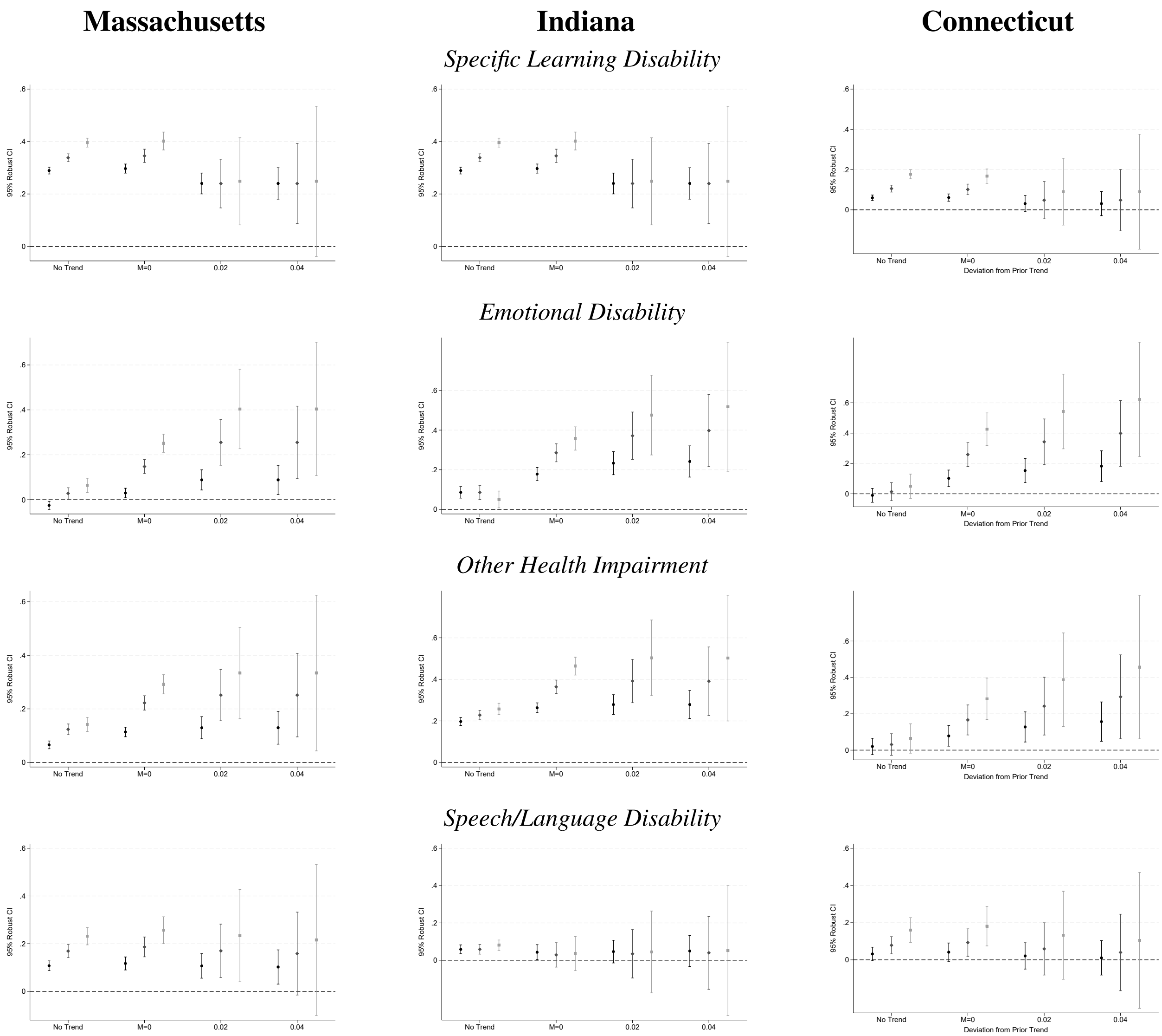
	(1)	(2)	(3)	(4)	(5)
	ELA		Math		
T-4	0.104***	0.102***	0.112***	0.112***	0.105***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.010)
T-3	0.087***	0.086***	0.093***	0.093***	0.086***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.008)
T-2	0.055***	0.054***	0.050***	0.050***	0.042***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
T-1	0	0	0	0	0
T=0	0.031***	0.024***	0.084***	0.080***	0.073***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)
T+1	0.101***	0.093***	0.142***	0.137***	0.130***
	(0.006)	(0.006)	(0.005)	(0.005)	(0.006)
T+2	0.171***	0.162***	0.195***	0.190***	0.183***
	(0.008)	(0.008)	(0.007)	(0.007)	(0.008)
T+3	0.218***	0.208***	0.249***	0.244***	0.238***
	(0.009)	(0.009)	(0.008)	(0.008)	(0.009)
Read Aloud		0.288***			
		(0.016)			
Scribe		0.183***			
		(0.026)			
Calculator				0.083***	
				(0.009)	
Accom Math					0.020***
					(0.005)
N	127,743	127,743	128,159	128,159	113,010
R^2	0.791	0.792	0.820	0.820	0.828

Notes: The table above shows estimates from event study regressions with student and school fixed effects, with T=0 representing the student's first year with an IEP. Columns 1 and 3 represent our main specification, while columsn 2, 4 and 5 add controls for various accommodations. The Massachusetts sample includes observations within 4 years of initial disability classification for students observed with and without an IEP during the sample period. Robust standard errors are clustered by student (* p<0.10 ** p<0.05 *** p<0.01.

Online Appendix

A Exhibits

Figure A.1: Effect of Classification by Select Classifications: Math

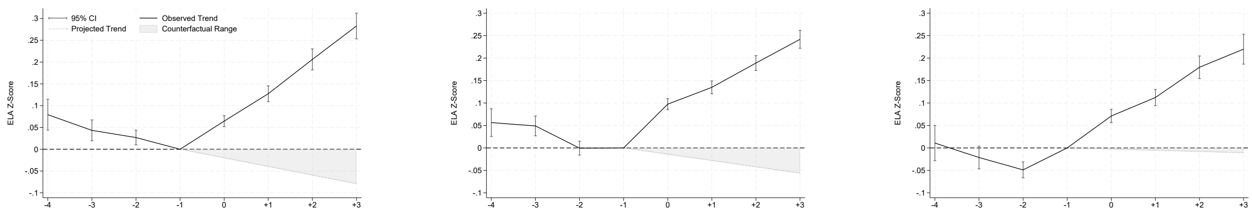


Notes: The figure above illustrates 95% confidence intervals resulting from the HonestDiD approach within samples restricted to students initially classified with a given disability category. Each panel illustrates estimates for effects during the initial classified year (T=0) and two subsequent years. The X-axis groups estimates according to the specified amount the model allows the post-treatment counterfactual trend to annually deviate from the linear pre-treatment trend. The first set of results assumes a flat counterfactual trend. The second set of estimates, M = 0, represents results where the counterfactual is the linear projection of the pre-treatment trend. Estimation samples include observations within 4 years of initial disability classification for students observed with and without an IEP during the sample period.

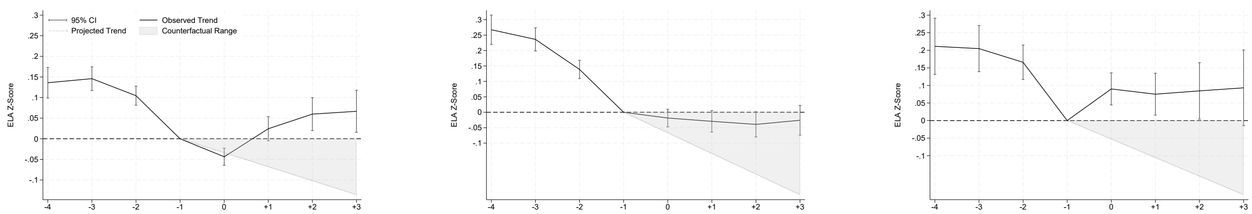
Figure A.2: Event Study by Select Classifications: ELA

Massachusetts **Indiana** **Connecticut**

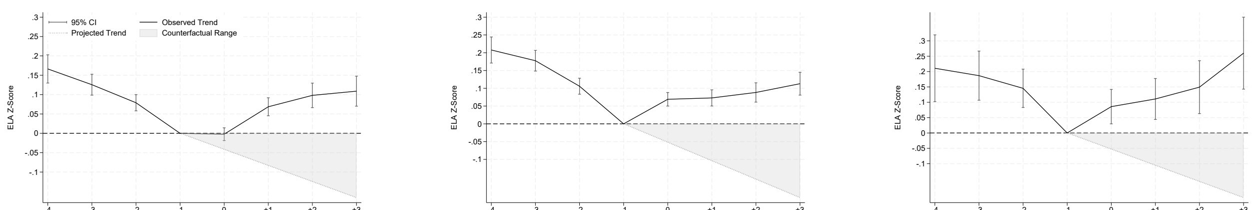
Specific Learning Disability



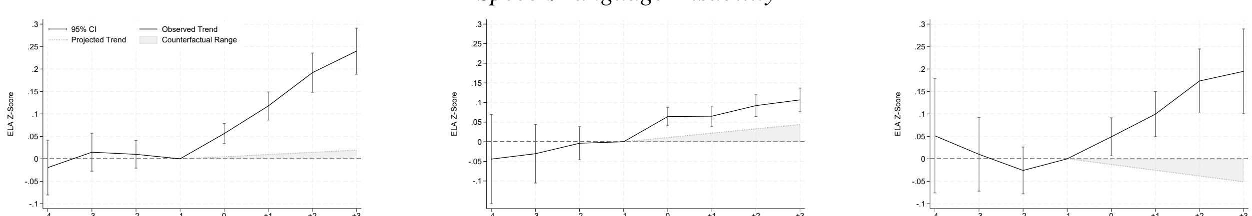
Emotional Disability



Other Health Impairment

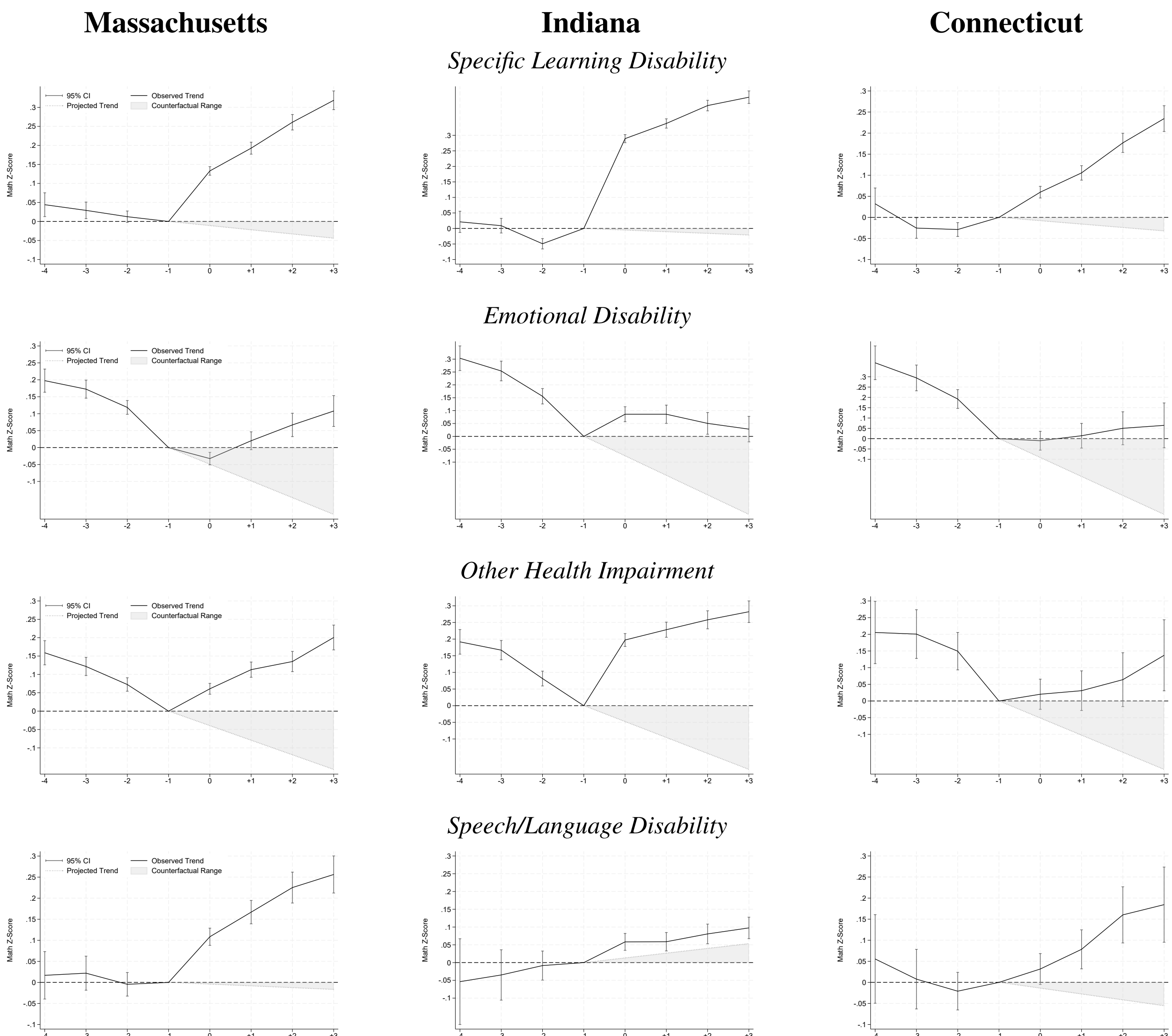


Speech/Language Disability



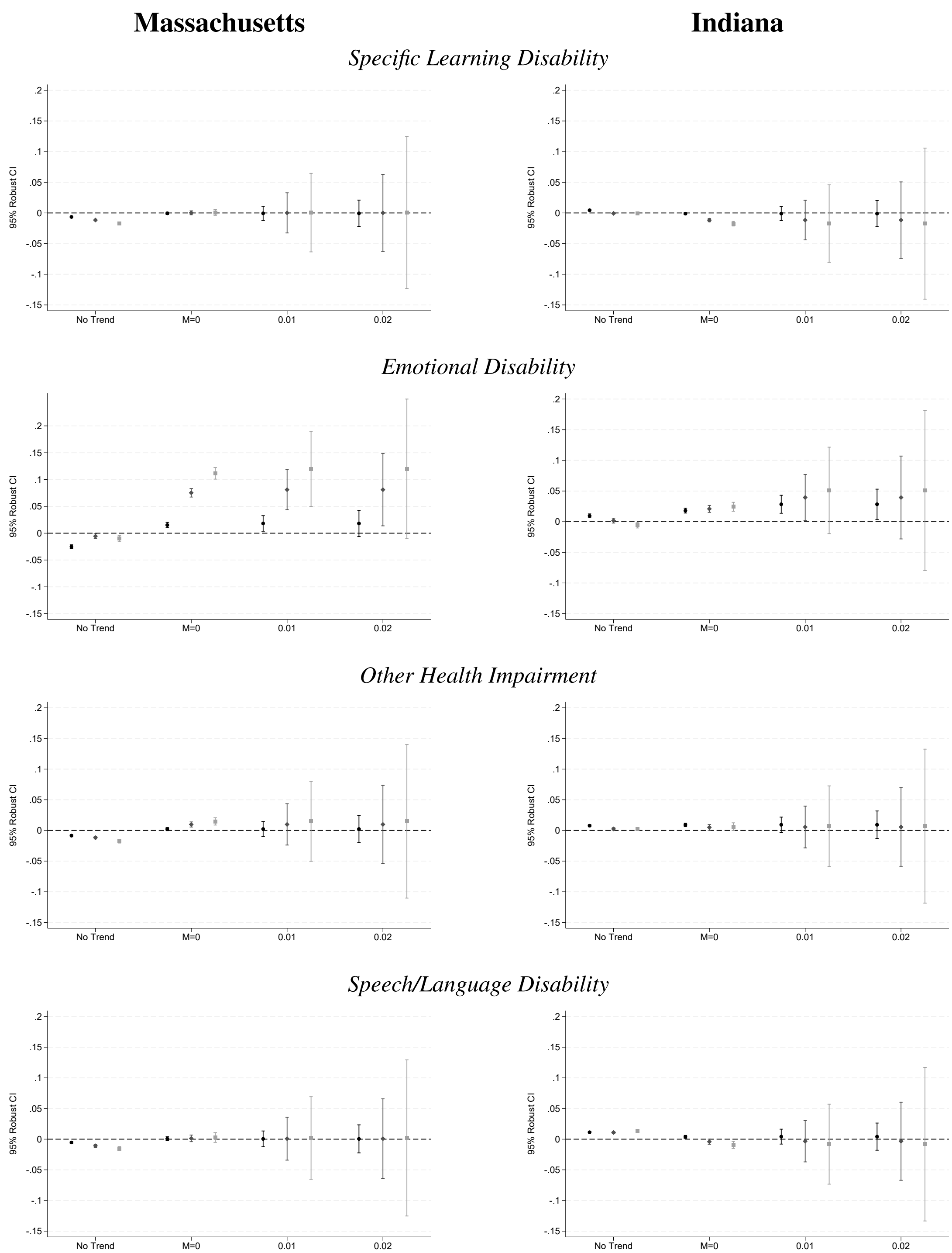
Notes: The figure above illustrates coefficients and 95% confidence intervals from regressions with student and school fixed effects. The X-axis describes years relative to initial classification, with T=0 representing the student's first year with an IEP. The sample includes observations within 4 years of initial disability classification for students observed with and without an IEP during the sample period. The dotted line illustrates continuation of the linear pre-treatment trend. Robust standard errors are clustered by student.

Figure A.3: Event Study by Select Classifications: Math



Notes: The figure above illustrates coefficients and 95% confidence intervals from regressions with student and school fixed effects. The X-axis describes years relative to initial classification, with T=0 representing the student's first year with an IEP. The sample includes observations within 4 years of initial disability classification for students observed with and without an IEP during the sample period. The dotted line illustrates continuation of the linear pre-treatment trend. Robust standard errors are clustered by student.

Figure A.4: Effect of Classification by Select Classifications: Attendance Rate

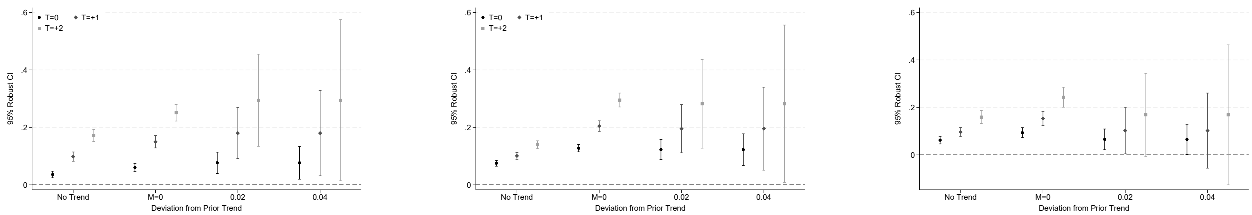


Notes: The figure above illustrates 95% confidence intervals resulting from the HonestDiD approach within samples restricted to students initially classified with a given disability category. Each panel illustrates estimates for effects during the initial classified year (T=0) and two subsequent years. The X-axis groups estimates according to the specified amount the model allows the post-treatment counterfactual trend to annually deviate from the linear pre-treatment trend. The first set of results assumes a flat counterfactual trend. The second set of estimates, M = 0, represents results where the counterfactual is the linear projection of the pre-treatment trend. Estimation samples include observations within 4 years of initial disability classification for students observed with and without an IEP during the sample period.

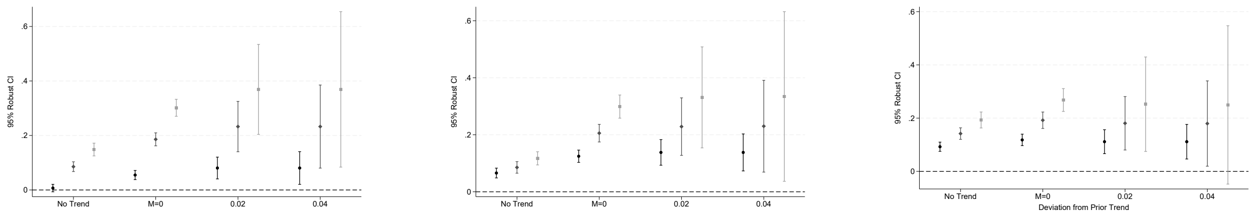
Figure A.5: Effect of Classification by SES & Race/Ethnicity: ELA

Massachusetts **Indiana** **Connecticut**

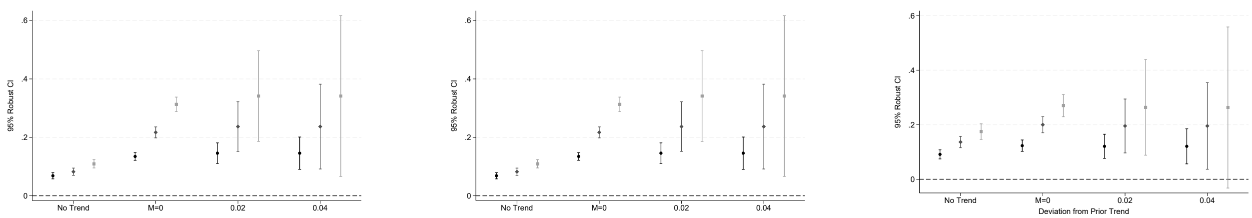
Eligible for Free or Reduced Priced Lunch



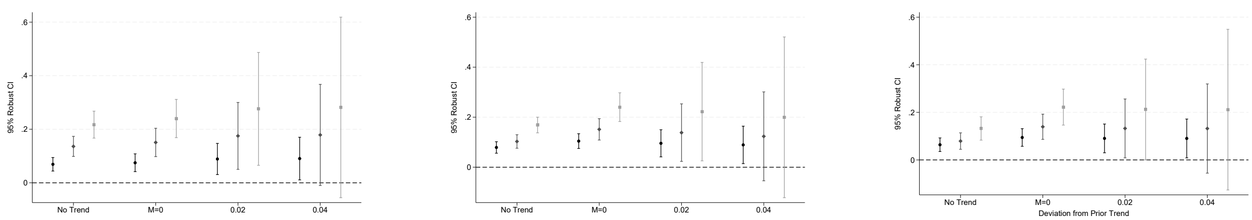
Not Eligible for Free or Reduced Priced Lunch



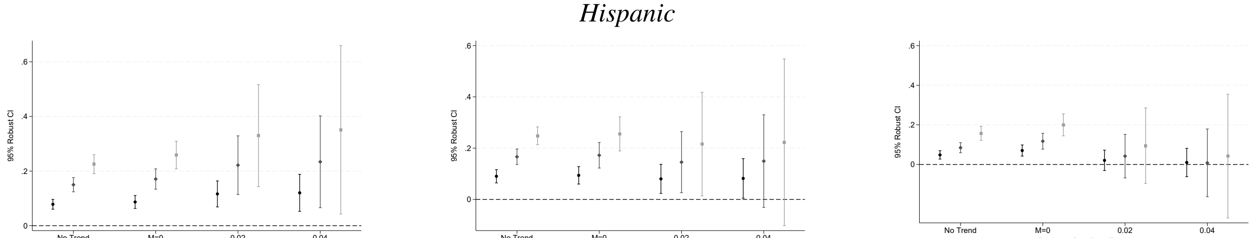
White



Black

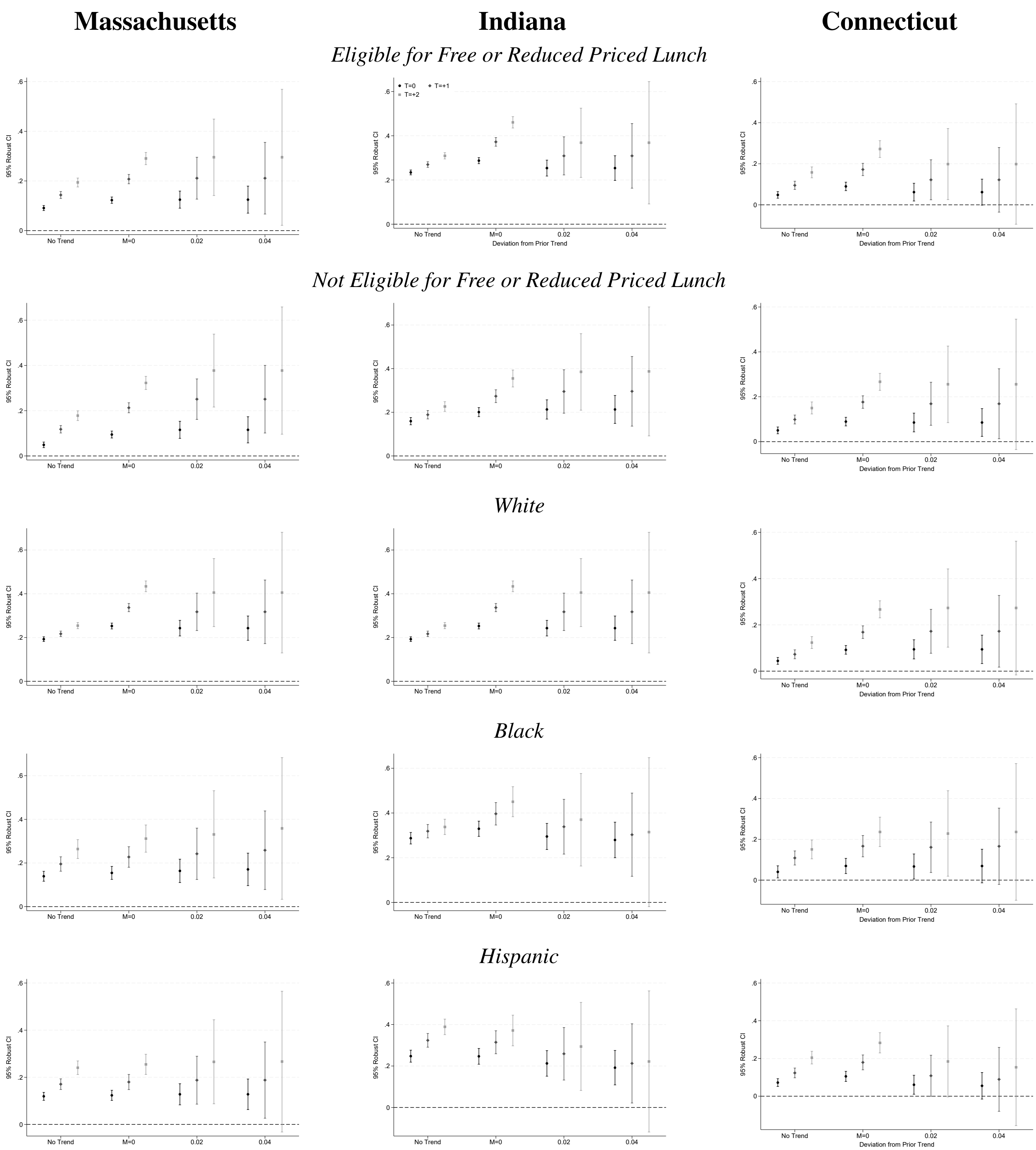


Hispanic



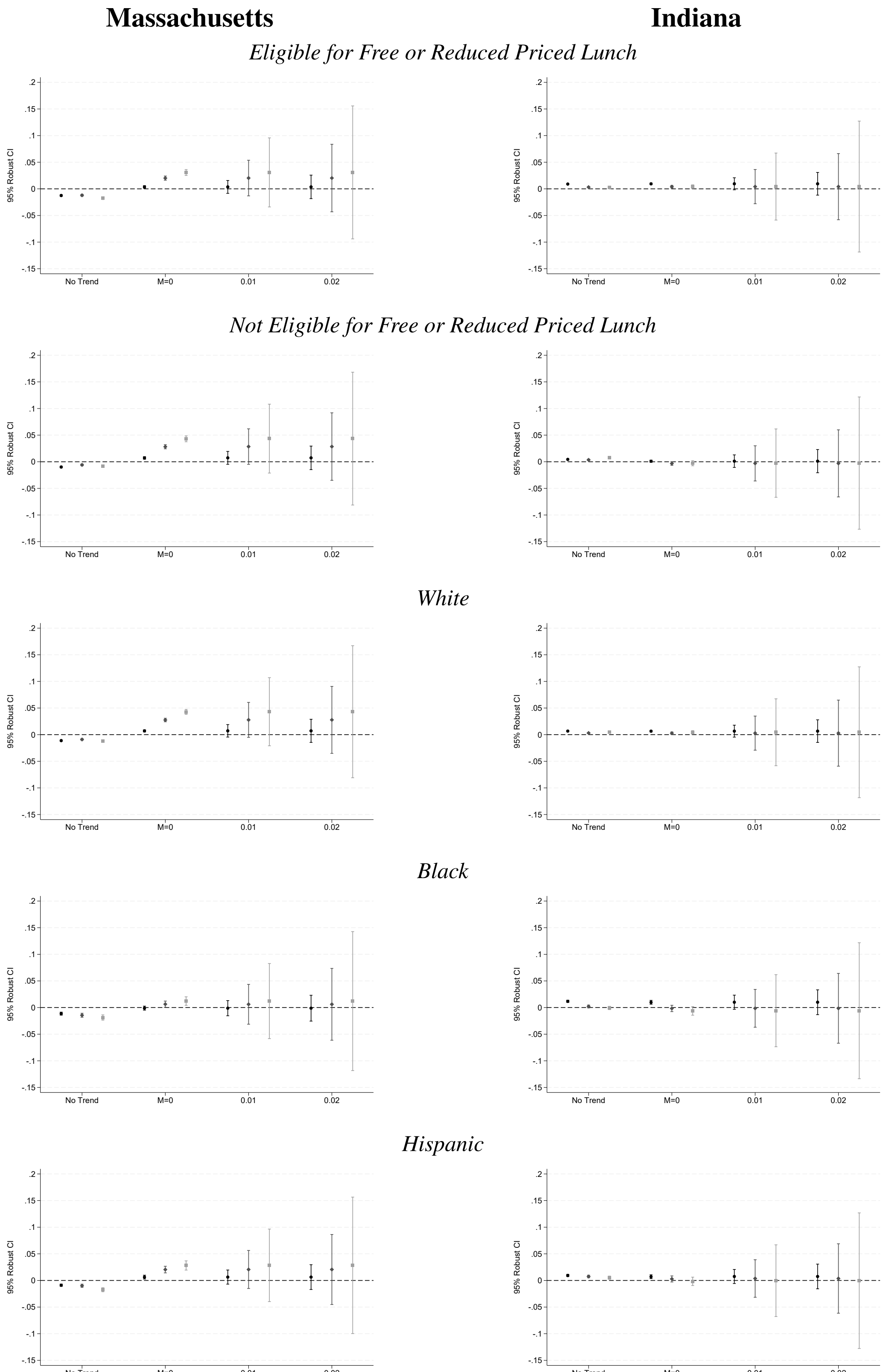
Notes: The figure above illustrates 95% confidence intervals resulting from the HonestDiD approach within samples defined by socioeconomic status and race/ethnicity. Each panel illustrates estimates for effects during the initial classified year (T=0) and two subsequent years. The X-axis groups estimates according to the specified amount the model allows the post-treatment counterfactual trend to annually deviate from the linear pre-treatment trend. The first set of results assumes a flat counterfactual trend. The second set of estimates, M = 0, represents results where the counterfactual is the linear projection of the pre-treatment trend. Estimation samples include observations within 4 years of initial disability classification for students observed with and without an IEP during the sample period.

Figure A.6: Effect of Classification by SES & Race/Ethnicity: Math



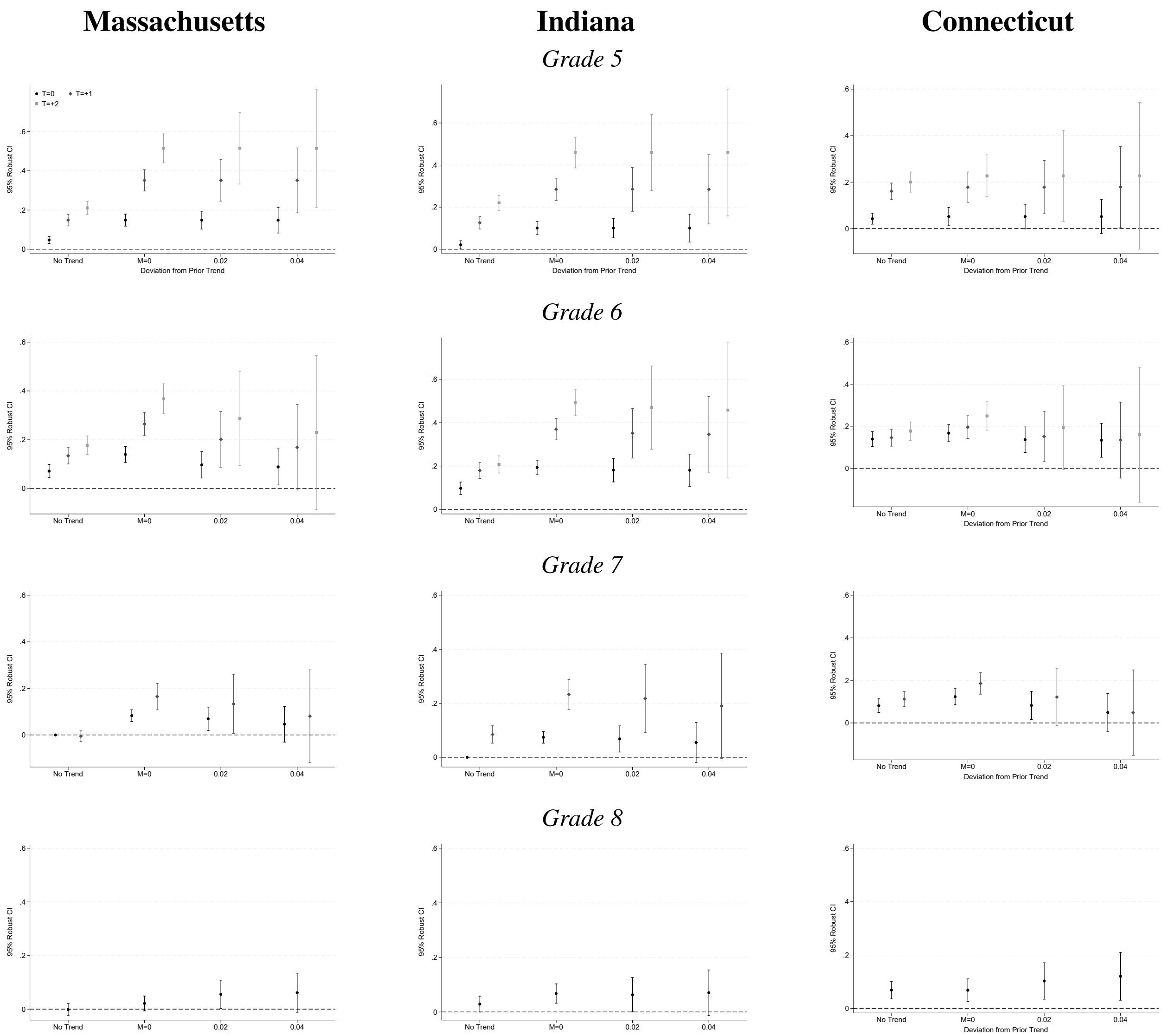
Notes: The figure above illustrates 95% confidence intervals resulting from the HonestDiD approach within samples defined by socioeconomic status and race/ethnicity. Each panel illustrates estimates for effects during the initial classified year (T=0) and two subsequent years. The X-axis groups estimates according to the specified amount the model allows the post-treatment counterfactual trend to annually deviate from the linear pre-treatment trend. The first set of results assumes a flat counterfactual trend. The second set of estimates, M = 0, represents results where the counterfactual is the linear projection of the pre-treatment trend. Estimation samples include observations within 4 years of initial disability classification for students observed with and without an IEP during the sample period.

Figure A.7: Effect of Classification by SES & Race/Ethnicity: Attendance Rate



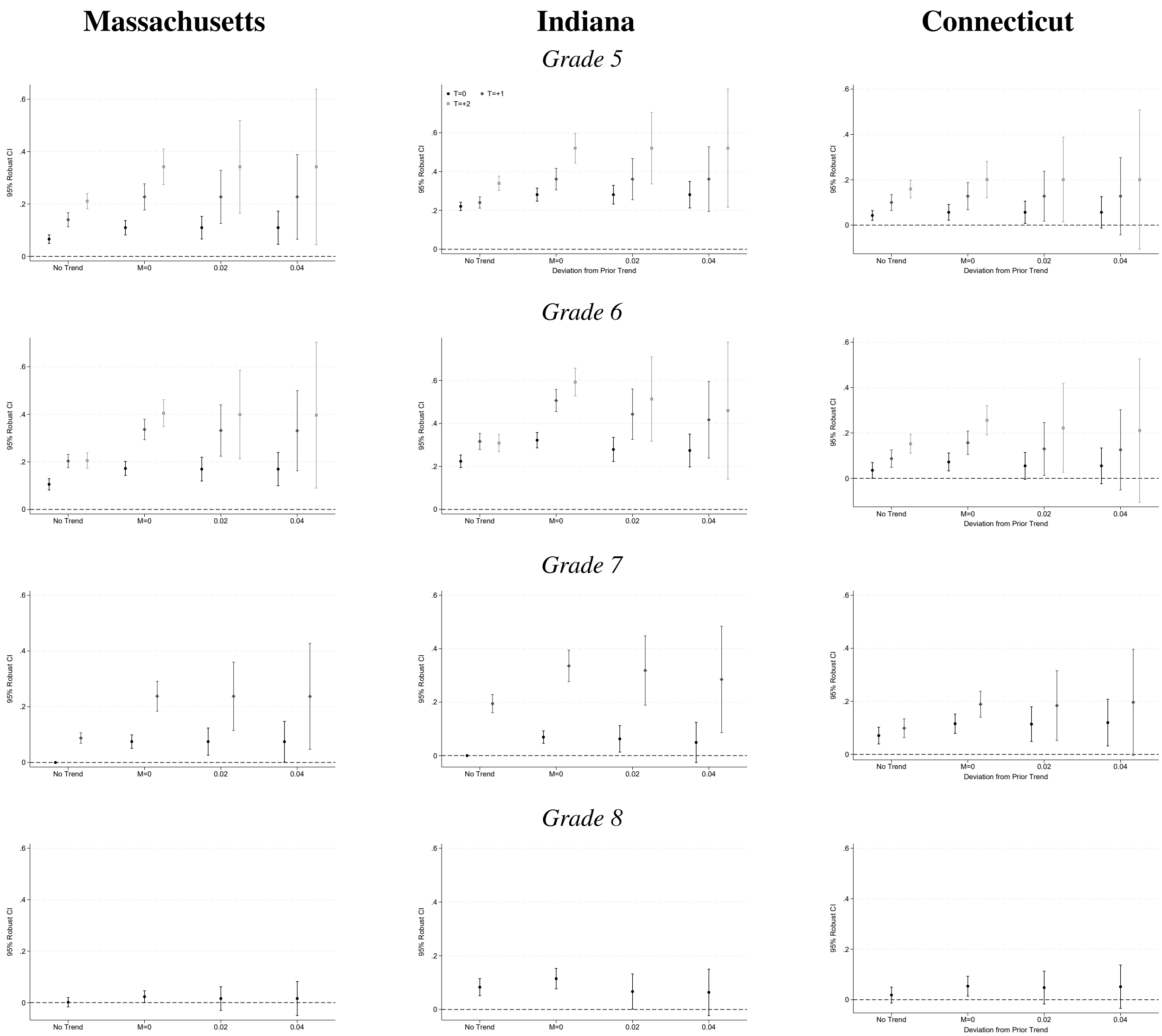
Notes: The figure above illustrates 95% confidence intervals resulting from the HonestDiD approach within samples defined by socioeconomic status and race/ethnicity. Each panel illustrates estimates for effects during the initial classified year (T=0) and two subsequent years. The X-axis groups estimates according to the specified amount the model allows the post-treatment counterfactual trend to annually deviate from the linear pre-treatment trend. The first set of results assumes a flat counterfactual trend. The second set of estimates, M = 0, represents results where the counterfactual is the linear projection of the pre-treatment trend. Estimation samples include observations within 4 years of initial disability

Figure A.8: Effect of Classification by Grade Classified: ELA



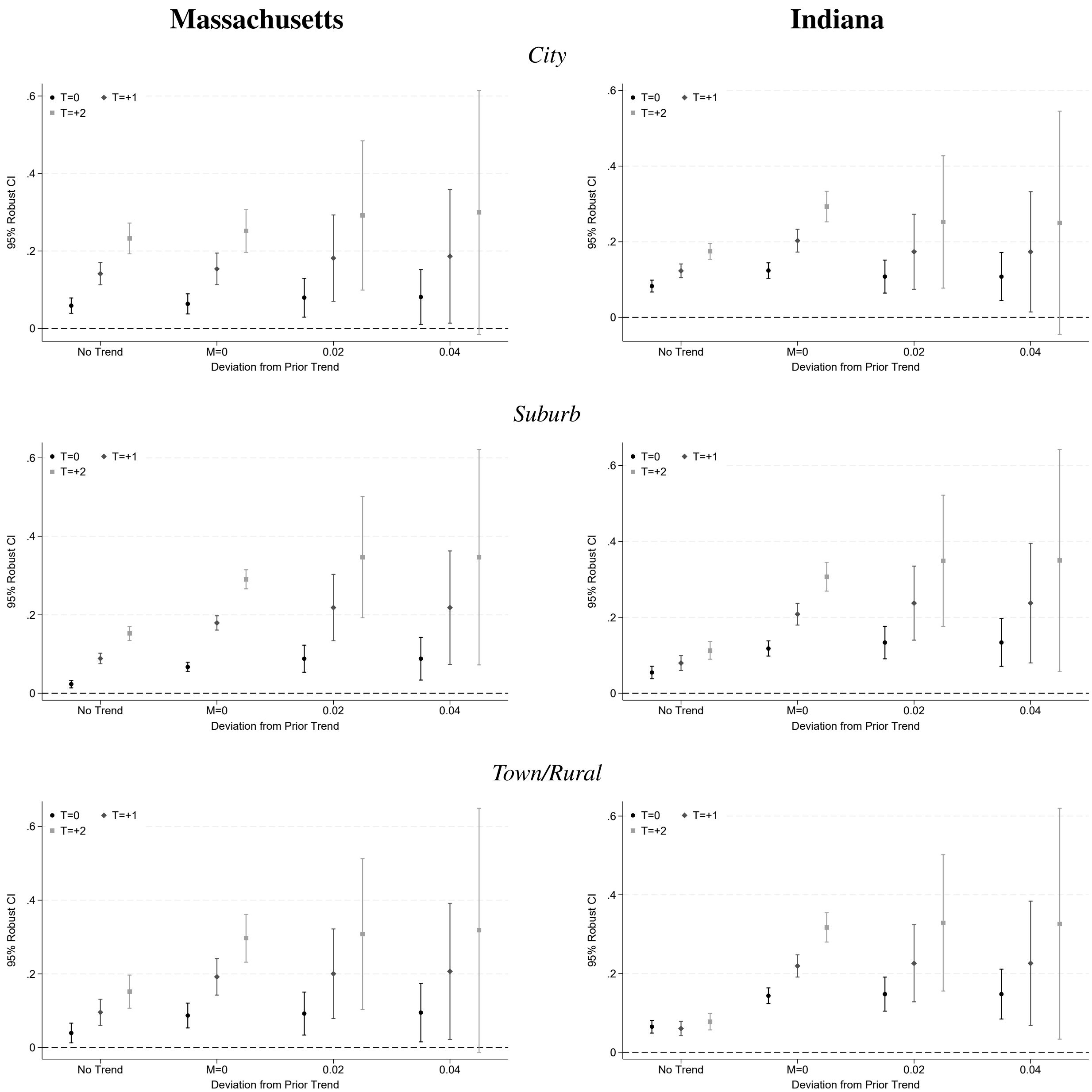
Notes: The figure above illustrates 95% confidence intervals resulting from the HonestDiD approach within samples restricted to students initially classified into special education in a given grade. Each panel illustrates estimates for effects during the initial classified year (T=0) and two subsequent years. The X-axis groups estimates according to the specified amount the model allows the post-treatment counterfactual trend to annually deviate from the linear pre-treatment trend. The first set of results assumes a flat counterfactual trend. The second set of estimates, M = 0, represents results where the counterfactual is the linear projection of the pre-treatment trend. Estimation samples include observations within 4 years of initial disability classification for students observed with and without an IEP during the sample period.

Figure A.9: Effect of Classification by Grade Classified: Math



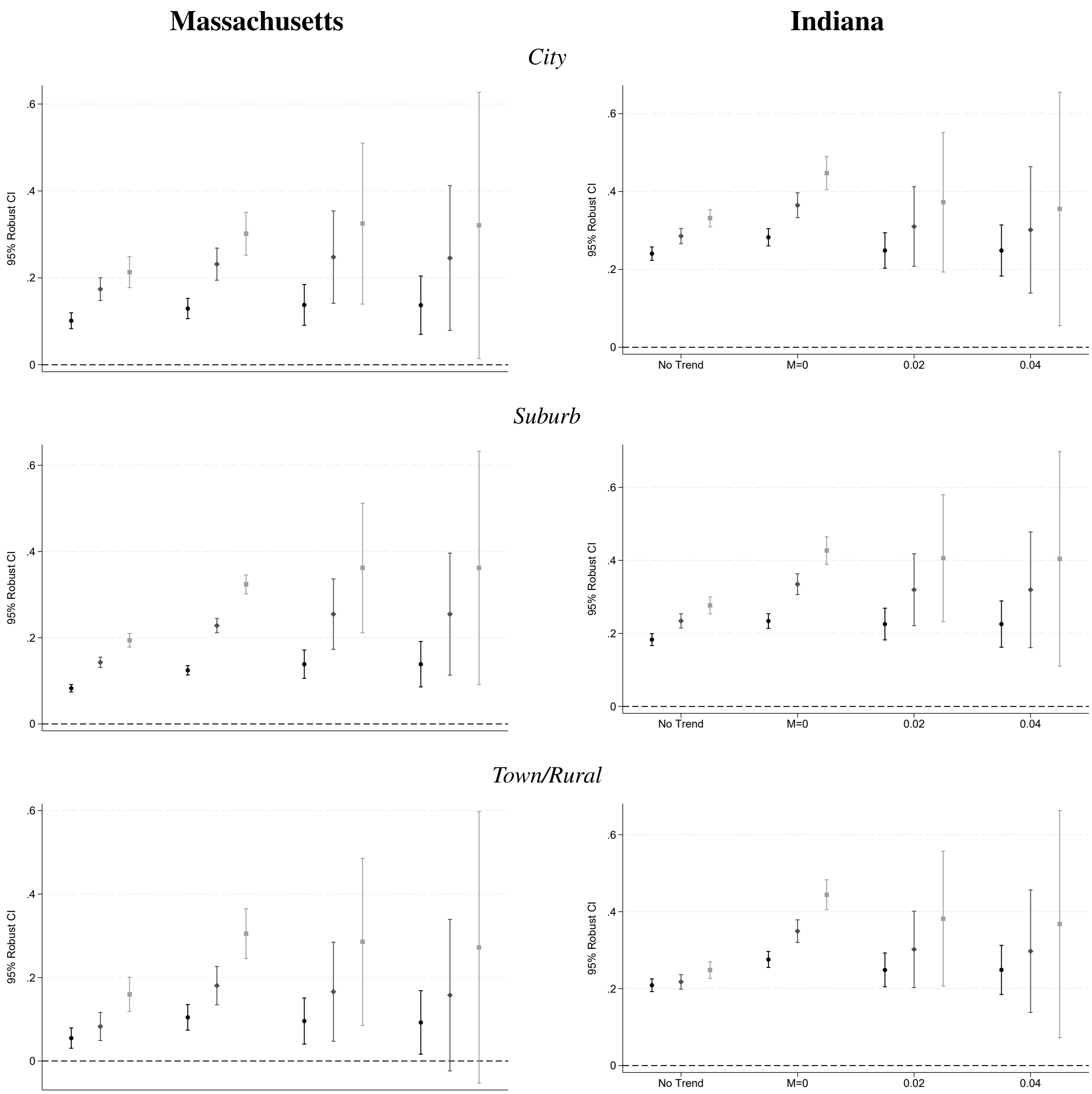
Notes: The figure above illustrates 95% confidence intervals resulting from the HonestDiD approach within samples restricted to students initially classified into special education in a given grade. Each panel illustrates estimates for effects during the initial classified year (T=0) and two subsequent years. The X-axis groups estimates according to the specified amount the model allows the post-treatment counterfactual trend to annually deviate from the linear pre-treatment trend. The first set of results assumes a flat counterfactual trend. The second set of estimates, M = 0, represents results where the counterfactual is the linear projection of the pre-treatment trend. Estimation samples include observations within 4 years of initial disability classification for students observed with and without an IEP during the sample period.

Figure A.10: Effect of Classification by Urbanicity: ELA



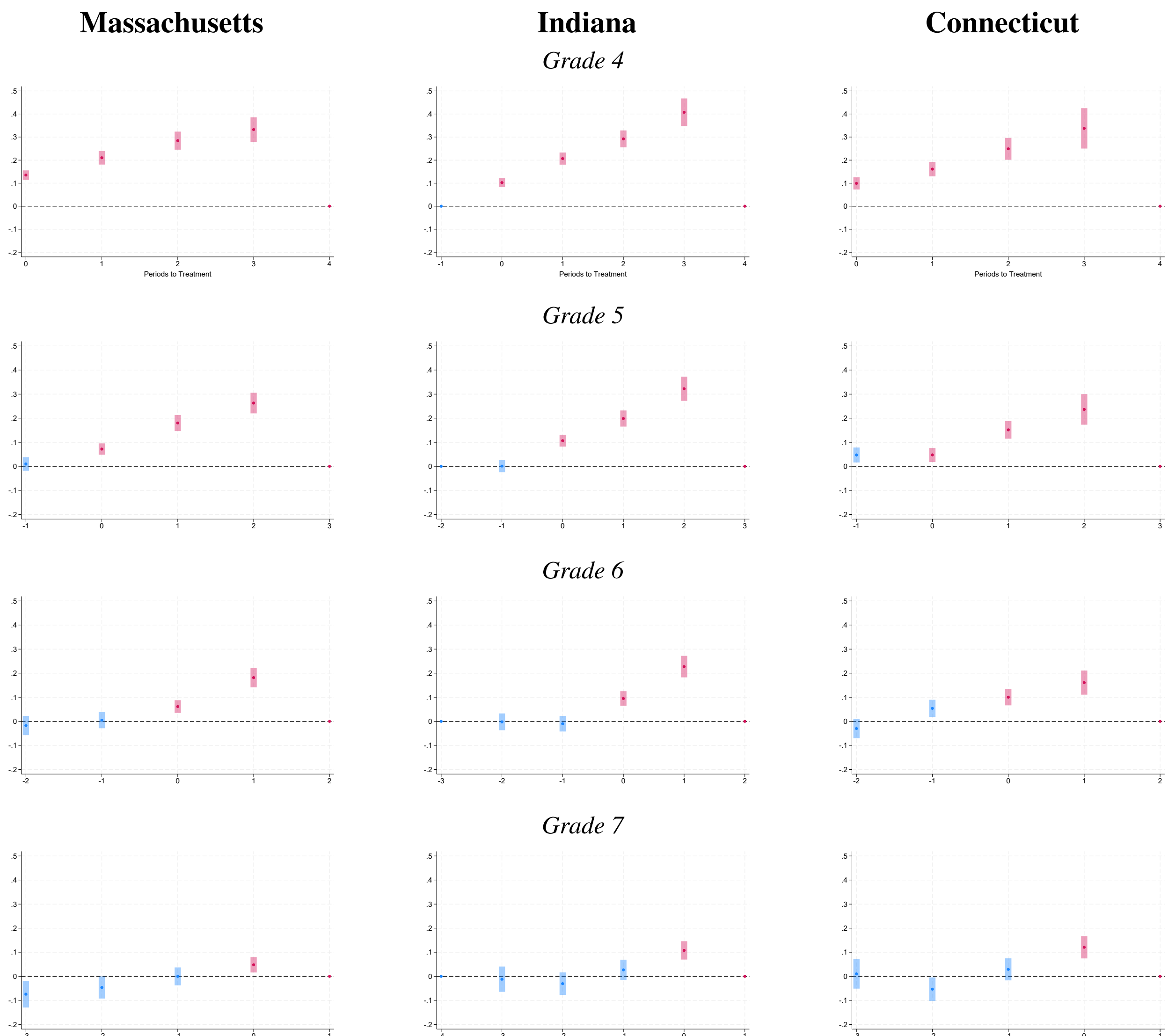
Notes: The figure above illustrates 95% confidence intervals resulting from the HonestDiD approach within samples defined by urbanicity. Each panel illustrates estimates for effects during the initial classified year (T=0) and two subsequent years. The X-axis groups estimates according to the specified amount the model allows the post-treatment counterfactual trend to annually deviate from the linear pre-treatment trend. The first set of results assumes a flat counterfactual trend. The second set of estimates, M = 0, represents results where the counterfactual is the linear projection of the pre-treatment trend. Estimation samples include observations within 4 years of initial disability classification for students observed with and without an IEP during the sample period.

Figure A.11: Effect of Classification by Urbanicity: Math



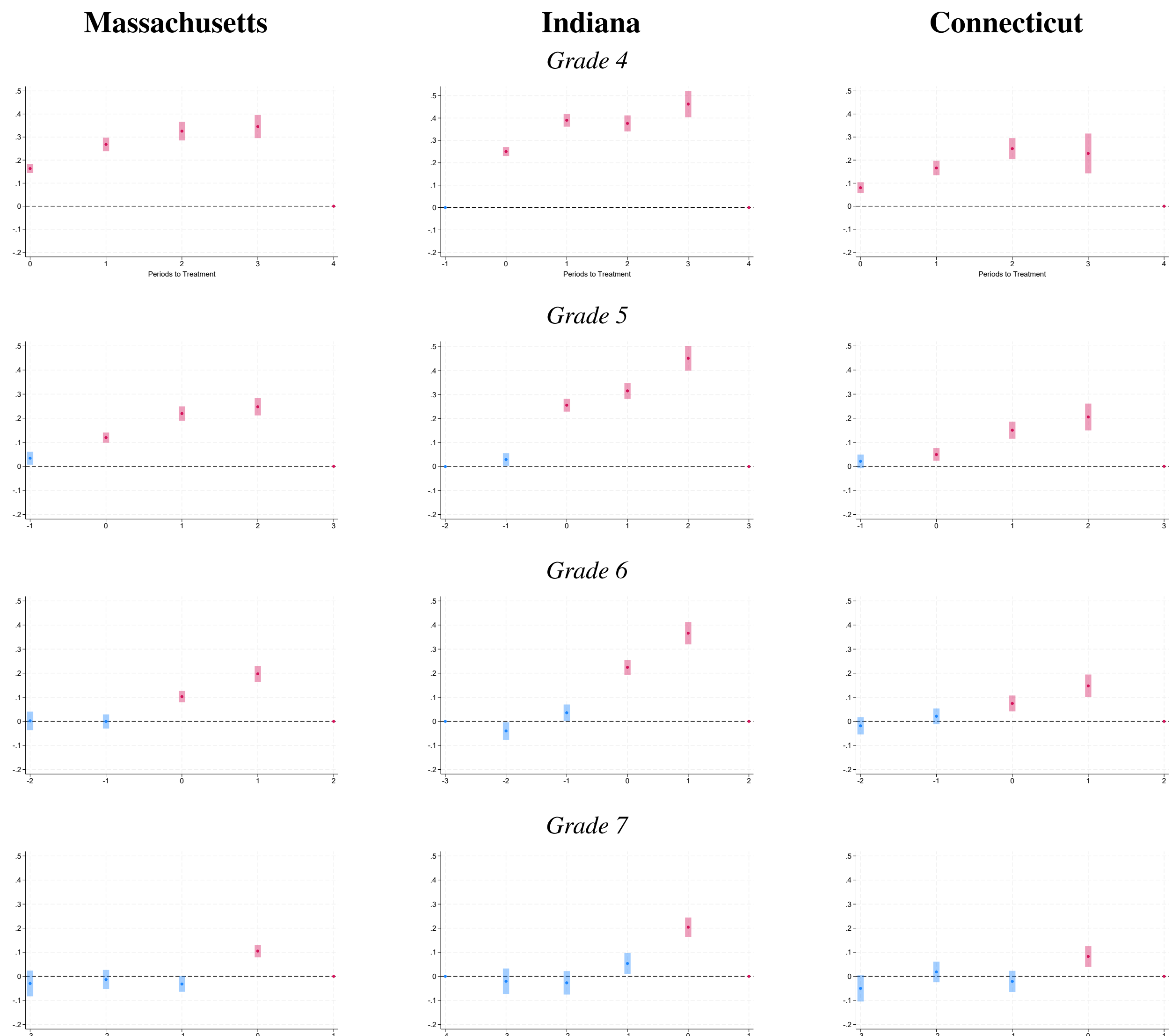
Notes: The figure above illustrates 95% confidence intervals resulting from the HonestDiD approach within samples defined by urbanicity. Each panel illustrates estimates for effects during the initial classified year (T=0) and two subsequent years. The X-axis groups estimates according to the specified amount the model allows the post-treatment counterfactual trend to annually deviate from the linear pre-treatment trend. The first set of results assumes a flat counterfactual trend. The second set of estimates, M = 0, represents results where the counterfactual is the linear projection of the pre-treatment trend. Estimation samples include observations within 4 years of initial disability classification for students observed with and without an IEP during the sample period.

Figure A.12: Difference-in-Differences Estimates by Grade Classified: ELA



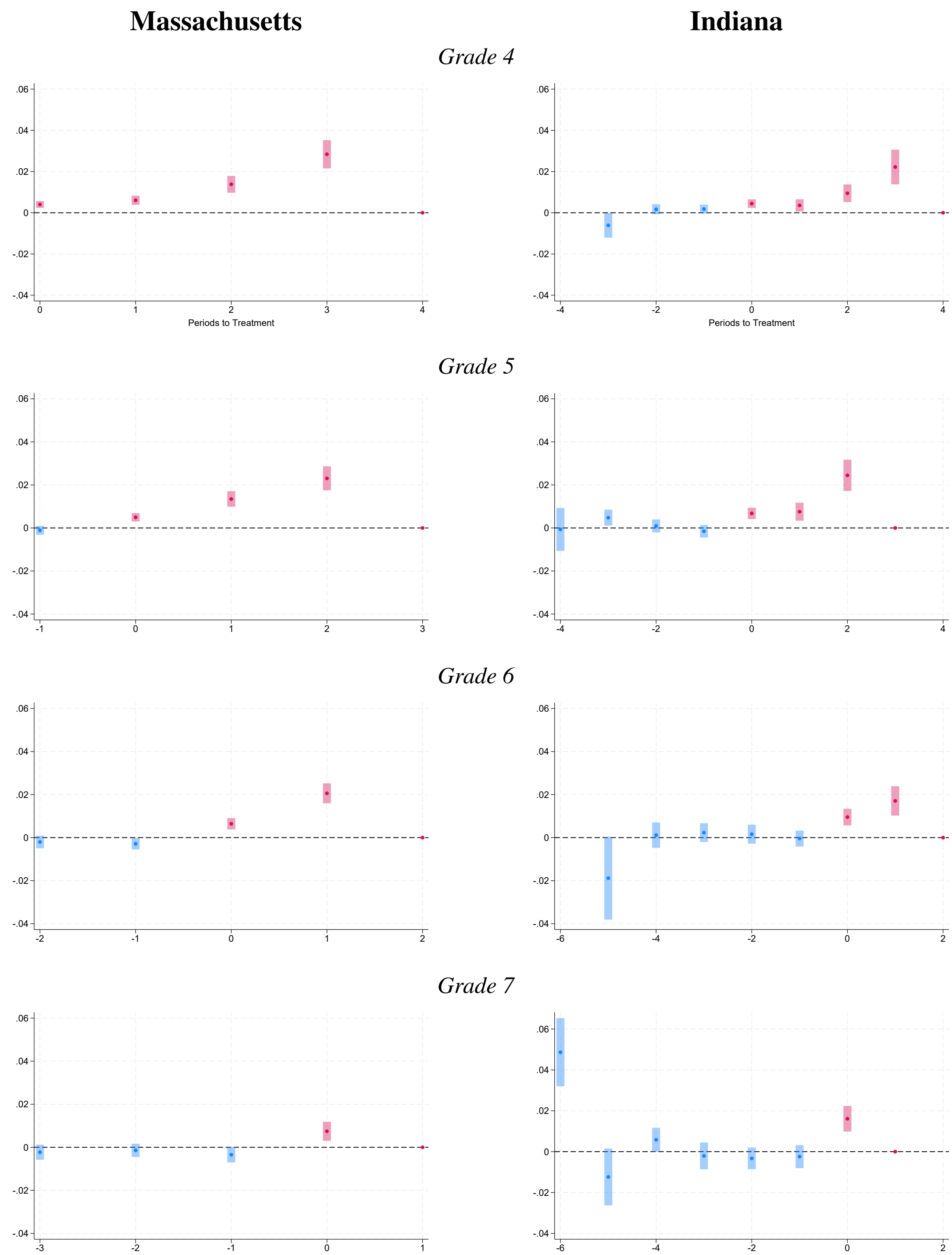
Notes: The figure above illustrates coefficients and 95% confidence intervals from difference-in-differences regressions comparing treated students to not-yet treated cohorts, within samples restricted to students initially classified into special education in a given grade.. The X-axis describes the number of grade levels relative to initial classification, with 0 representing the student's first grade with an IEP. Robust standard errors are clustered by student.

Figure A.13: Difference-in-Differences Estimates by Grade Classified: Math



Notes: The figure above illustrates coefficients and 95% confidence intervals from difference-in-differences regressions comparing treated students to not-yet treated cohorts, within samples restricted to students initially classified into special education in a given grade.. The X-axis describes the number of grade levels relative to initial classification, with 0 representing the student's first grade with an IEP. Robust standard errors are clustered by student.

Figure A.14: Difference-in-Differences Estimates by Grade Classified: Attendance Rate



Notes: The figure above illustrates coefficients and 95% confidence intervals from difference-in-differences regressions comparing treated students to not-yet treated cohorts, within samples restricted to students initially classified into special education in a given grade.. The X-axis describes the number of grade levels relative to initial classification, with 0 representing the student's first grade with an IEP. Robust standard errors are clustered by student.

Table A.1: Descriptive Statistics: Indiana

	(1) All	(2) Never	(3) Always	(4) Change	(5) Grade 4	(6) Grade 5	(7) Grade 6	(8) Grade 7	(9) Grade 8	(10) Learning	(11) Speech	(12) Emo	(13) Health	(14) Other IEP
Avg ELA	-0.0312	0.138	-1.056	-0.477	-0.665	-0.676	-0.747	-0.807	-0.855	-1.080	-0.0150	-0.750	-0.868	-0.777
	(0.963)	(0.871)	(0.924)	(0.920)	(0.891)	(0.866)	(0.850)	(0.828)	(0.797)	(0.639)	(0.824)	(0.857)	(0.734)	(0.882)
Avg Math	-0.0305	0.112	-0.872	-0.428	-0.576	-0.649	-0.752	-0.832	-0.912	-1.014	0.0399	-0.736	-0.817	-0.749
	(0.966)	(0.899)	(0.962)	(0.961)	(0.920)	(0.905)	(0.922)	(0.901)	(0.850)	(0.705)	(0.860)	(0.879)	(0.784)	(0.935)
Female	0.488	0.517	0.343	0.387	0.409	0.421	0.393	0.374	0.371	0.490	0.365	0.290	0.320	0.394
	(0.500)	(0.500)	(0.475)	(0.487)	(0.492)	(0.494)	(0.488)	(0.484)	(0.483)	(0.500)	(0.481)	(0.454)	(0.467)	(0.489)
Black	0.125	0.122	0.144	0.135	0.147	0.154	0.168	0.169	0.179	0.199	0.0793	0.190	0.162	0.177
	(0.331)	(0.327)	(0.351)	(0.341)	(0.354)	(0.361)	(0.374)	(0.375)	(0.383)	(0.399)	(0.270)	(0.392)	(0.368)	(0.382)
Hispanic	0.115	0.120	0.0941	0.102	0.111	0.116	0.103	0.0883	0.0843	0.143	0.0952	0.0575	0.0652	0.101
	(0.319)	(0.325)	(0.292)	(0.302)	(0.314)	(0.320)	(0.303)	(0.284)	(0.278)	(0.350)	(0.294)	(0.233)	(0.247)	(0.301)
White	0.686	0.683	0.699	0.700	0.672	0.666	0.663	0.687	0.676	0.596	0.765	0.673	0.704	0.654
	(0.464)	(0.465)	(0.459)	(0.458)	(0.469)	(0.472)	(0.473)	(0.464)	(0.468)	(0.491)	(0.424)	(0.469)	(0.457)	(0.476)
Ever Lunch	0.606	0.577	0.747	0.703	0.725	0.739	0.741	0.741	0.756	0.809	0.616	0.855	0.734	0.760
	(0.489)	(0.494)	(0.435)	(0.457)	(0.447)	(0.439)	(0.438)	(0.438)	(0.430)	(0.393)	(0.486)	(0.352)	(0.442)	(0.427)
Ever EL	0.0714	0.0744	0.0562	0.0623	0.0741	0.0797	0.0773	0.0553	0.0525	0.107	0.0511	0.0201	0.0317	0.0695
	(0.257)	(0.262)	(0.230)	(0.242)	(0.262)	(0.271)	(0.267)	(0.229)	(0.223)	(0.309)	(0.220)	(0.140)	(0.175)	(0.254)
Yrs Observed	3.437	3.387	3.233	4.666	4.570	4.557	4.414	4.216	4.135	4.707	4.775	4.628	4.695	4.688
	(1.775)	(1.771)	(1.763)	(1.342)	(1.349)	(1.103)	(1.210)	(1.490)	(1.594)	(1.352)	(1.272)	(1.337)	(1.338)	(1.356)
SLD	0.0670		0.408	0.257	0.373	0.373	0.393	0.368	0.260					
	(0.250)		(0.492)	(0.437)	(0.484)	(0.484)	(0.488)	(0.482)	(0.439)					
Communication	0.0464		0.151	0.492	0.354	0.309	0.203	0.115	0.0254					
	(0.210)		(0.358)	(0.500)	(0.478)	(0.462)	(0.402)	(0.319)	(0.157)					
Other Health	0.0267		0.155	0.122	0.152	0.161	0.198	0.237	0.200					
	(0.161)		(0.362)	(0.327)	(0.359)	(0.367)	(0.399)	(0.426)	(0.400)					
Emotional	0.0157		0.0932	0.0661	0.0664	0.0869	0.113	0.151	0.147					
	(0.124)		(0.291)	(0.248)	(0.249)	(0.282)	(0.317)	(0.358)	(0.354)					
Other IEP Category	0.0536 (0.225)		0.322 (0.467)	0.216 (0.412)	0.253 (0.435)	0.287 (0.452)	0.294 (0.456)	0.329 (0.470)	0.371 (0.483)					
N	1,171,390	955,456	152,022	63,912	14,180	9,778	6,336	4,249	2,955	16,414	31,418	4,226	7,793	13,835

Notes: Shown above for Indiana are sample means (with standard deviations in parentheses). Column 1 is the full sample, columns 2 and 3 represent students never or always observed in special education, and column 4 represents the main analytic sample of students first classified in grades 4-8. Columns 5-8 separate the sample in column 4 by the grade first classified. Columns 10-14 separate the sample in column 4 by initial disability classification.

Table A.2: Descriptive Statistics: Massachusetts

	(1) All	(2) Never	(3) Always	(4) Change	(5) Grade 4	(6) Grade 5	(7) Grade 6	(8) Grade 7	(9) Grade 8	(10) Learning	(11) Speech	(12) Emo	(13) Health	(14) Other IEP
Avg ELA	-0.0433	0.168	-0.875	-0.719	-0.807	-0.785	-0.766	-0.703	-0.646	-0.925	-0.872	-0.486	-0.633	-0.725
	(0.972)	(0.877)	(0.881)	(0.835)	(0.737)	(0.761)	(0.768)	(0.817)	(0.881)	(0.745)	(0.789)	(0.894)	(0.739)	(0.940)
Avg Math	-0.0359	0.170	-0.849	-0.756	-0.797	-0.819	-0.811	-0.766	-0.712	-0.959	-0.798	-0.587	-0.684	-0.732
	(0.969)	(0.884)	(0.856)	(0.804)	(0.770)	(0.767)	(0.758)	(0.767)	(0.780)	(0.682)	(0.792)	(0.869)	(0.747)	(0.906)
Female	0.487	0.520	0.365	0.418	0.440	0.415	0.392	0.367	0.412	0.475	0.430	0.459	0.295	0.361
	(0.500)	(0.500)	(0.481)	(0.493)	(0.496)	(0.493)	(0.488)	(0.482)	(0.492)	(0.499)	(0.495)	(0.498)	(0.456)	(0.480)
Black	0.0917	0.0883	0.105	0.111	0.107	0.113	0.114	0.118	0.0973	0.132	0.127	0.102	0.0756	0.126
	(0.289)	(0.284)	(0.306)	(0.314)	(0.309)	(0.317)	(0.318)	(0.323)	(0.296)	(0.338)	(0.333)	(0.302)	(0.264)	(0.332)
Hispanic	0.185	0.176	0.218	0.211	0.222	0.235	0.207	0.196	0.197	0.251	0.246	0.192	0.158	0.203
	(0.388)	(0.381)	(0.413)	(0.408)	(0.415)	(0.424)	(0.405)	(0.397)	(0.398)	(0.434)	(0.431)	(0.394)	(0.365)	(0.402)
White	0.626	0.629	0.616	0.613	0.597	0.585	0.613	0.626	0.646	0.559	0.534	0.641	0.714	0.606
	(0.484)	(0.483)	(0.486)	(0.487)	(0.490)	(0.493)	(0.487)	(0.484)	(0.478)	(0.497)	(0.499)	(0.480)	(0.452)	(0.489)
Ever FRPL	0.430	0.397	0.548	0.568	0.554	0.566	0.572	0.587	0.583	0.610	0.585	0.629	0.490	0.539
	(0.495)	(0.489)	(0.498)	(0.495)	(0.497)	(0.496)	(0.495)	(0.492)	(0.493)	(0.488)	(0.493)	(0.483)	(0.500)	(0.499)
Ever EL	0.124	0.126	0.118	0.154	0.176	0.184	0.158	0.132	0.126	0.203	0.230	0.0873	0.0773	0.189
	(0.330)	(0.332)	(0.322)	(0.361)	(0.381)	(0.388)	(0.365)	(0.339)	(0.331)	(0.402)	(0.421)	(0.282)	(0.267)	(0.392)
Ever Retain	0.0349	0.0229	0.0765	0.0949	0.0349	0.0538	0.0686	0.113	0.111	0.0915	0.0740	0.139	0.0770	0.102
	(0.183)	(0.149)	(0.266)	(0.293)	(0.184)	(0.226)	(0.253)	(0.316)	(0.314)	(0.288)	(0.262)	(0.346)	(0.267)	(0.303)
Yrs Observed	3.426	3.345	3.729	4.931	4.517	4.953	5.129	5.374	5.147	4.947	5.166	5.086	5.101	5.084
	(2.000)	(1.989)	(2.014)	(1.715)	(1.826)	(1.763)	(1.681)	(1.337)	(1.468)	(1.714)	(1.672)	(1.679)	(1.666)	(1.669)
Learning	0.0862		0.398	0.379	0.490	0.446	0.376	0.307	0.302					
	(0.281)		(0.489)	(0.485)	(0.500)	(0.497)	(0.484)	(0.461)	(0.459)					
Speech/Language	0.0412		0.178	0.119	0.169	0.151	0.111	0.102	0.0807					
	(0.199)		(0.383)	(0.324)	(0.375)	(0.358)	(0.315)	(0.302)	(0.272)					
Other Health	0.0387		0.181	0.260	0.224	0.257	0.291	0.309	0.271					
	(0.193)		(0.385)	(0.439)	(0.417)	(0.437)	(0.454)	(0.462)	(0.444)					
Emotional	0.0296		0.139	0.234	0.133	0.166	0.208	0.274	0.316					
	(0.170)		(0.346)	(0.423)	(0.339)	(0.372)	(0.406)	(0.446)	(0.465)					
Other IEP	0.0645		0.296	0.170	0.157	0.172	0.183	0.192	0.171					
Category	(0.246)		(0.457)	(0.376)	(0.364)	(0.377)	(0.387)	(0.394)	(0.377)					
Accom	0.186		0.729	0.669	0.807	0.702	0.660	0.690	0.626	0.715	0.703	0.612	0.731	0.714
Comprehend	(0.389)		(0.444)	(0.471)	(0.395)	(0.458)	(0.474)	(0.463)	(0.484)	(0.452)	(0.457)	(0.487)	(0.443)	(0.452)
Accom Math	0.193		0.782	0.811	0.870	0.868	0.851	0.853	0.783	0.871	0.811	0.732	0.859	0.832
	(0.395)		(0.413)	(0.391)	(0.336)	(0.338)	(0.356)	(0.355)	(0.412)	(0.335)	(0.392)	(0.443)	(0.348)	(0.374)
Read Aloud	0.0255		0.118	0.0608	0.0741	0.0776	0.0592	0.0598	0.0503	0.0934	0.0792	0.0253	0.0359	0.0920
	(0.158)		(0.322)	(0.239)	(0.262)	(0.268)	(0.236)	(0.237)	(0.219)	(0.291)	(0.270)	(0.157)	(0.186)	(0.289)
Scribe	0.0119		0.0525	0.0222	0.0386	0.0251	0.0207	0.0188	0.0116	0.0231	0.0187	0.0165	0.0226	0.0386
	(0.108)		(0.223)	(0.147)	(0.193)	(0.156)	(0.143)	(0.136)	(0.107)	(0.150)	(0.135)	(0.128)	(0.149)	(0.193)
Calculator	0.0372		0.173	0.112	0.0977	0.104	0.121	0.128	0.121	0.158	0.108	0.0775	0.0884	0.159
	(0.189)		(0.378)	(0.315)	(0.297)	(0.306)	(0.326)	(0.334)	(0.327)	(0.364)	(0.311)	(0.267)	(0.284)	(0.366)
N	1,210,622	954,391	255,894	35,952	7,935	6,986	5,353	4,245	4,748	13,609	4,281	8,409	9,342	6,118

Notes: Shown above for Massachusetts are sample means (with standard deviations in parentheses). Column 1 is the full sample, columns 2 and 3 represent students never or always observed in special education, and column 4 represents the main analytic sample of students first classified in grades 4-8. Columns 5-8 separate the sample in column 4 by the grade first classified. Columns 10-14 separate the sample in column 4 by initial disability classification.

Table A.3: Descriptive Statistics: Connecticut

	(1) All	(2) Never	(3) Always	(4) Change	(5) Grade 4	(6) Grade 5	(7) Grade 6	(8) Grade 7	(9) Grade 8	(10) SLD	(11) Speech	(12) Emo	(13) Health	(14) Other IEP
Avg ELA	-0.00987	0.155	-1.001	-0.640	-0.823	-0.848	-0.814	-0.772	-0.699	-0.864	-0.295	-0.624	-0.393	-0.535
	(0.961)	(0.892)	(0.806)	(0.836)	(0.789)	(0.753)	(0.775)	(0.794)	(0.830)	(0.713)	(0.851)	(0.908)	(0.852)	(0.879)
Avg Math	-0.00950	0.157	-1.019	-0.629	-0.818	-0.829	-0.833	-0.795	-0.704	-0.861	-0.212	-0.667	-0.454	-0.524
	(0.969)	(0.891)	(0.862)	(0.865)	(0.837)	(0.788)	(0.764)	(0.807)	(0.835)	(0.745)	(0.877)	(0.891)	(0.850)	(0.908)
Female	0.487	0.516	0.328	0.389	0.420	0.428	0.433	0.373	0.356	0.484	0.348	0.321	0.383	0.261
	(0.500)	(0.500)	(0.469)	(0.487)	(0.494)	(0.495)	(0.496)	(0.484)	(0.479)	(0.500)	(0.476)	(0.467)	(0.486)	(0.439)
Black	0.129	0.123	0.162	0.146	0.149	0.168	0.164	0.165	0.164	0.163	0.108	0.186	0.101	0.135
	(0.335)	(0.328)	(0.368)	(0.353)	(0.356)	(0.374)	(0.370)	(0.371)	(0.370)	(0.369)	(0.310)	(0.389)	(0.301)	(0.342)
Hispanic	0.253	0.243	0.313	0.279	0.313	0.293	0.295	0.287	0.244	0.314	0.264	0.278	0.193	0.242
	(0.435)	(0.429)	(0.464)	(0.449)	(0.464)	(0.455)	(0.456)	(0.452)	(0.430)	(0.464)	(0.441)	(0.448)	(0.395)	(0.428)
White	0.529	0.538	0.466	0.518	0.480	0.478	0.489	0.500	0.542	0.477	0.546	0.474	0.653	0.563
	(0.499)	(0.499)	(0.499)	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)	(0.498)	(0.499)	(0.498)	(0.499)	(0.476)	(0.496)
Ever Lunch	0.464	0.437	0.603	0.582	0.614	0.616	0.625	0.616	0.573	0.618	0.511	0.712	0.502	0.542
	(0.499)	(0.496)	(0.489)	(0.493)	(0.487)	(0.487)	(0.484)	(0.487)	(0.495)	(0.486)	(0.500)	(0.453)	(0.500)	(0.498)
Ever Retain	0.00792	0.00579	0.0158	0.0215	0.0203	0.0189	0.0195	0.0344	0.0388	0.0204	0.00795	0.0545	0.0216	0.0213
	(0.0886)	(0.0759)	(0.125)	(0.145)	(0.141)	(0.136)	(0.138)	(0.182)	(0.193)	(0.142)	(0.0888)	(0.227)	(0.145)	(0.144)
Yrs Observed	3.261	3.217	3.078	4.497	4.307	4.832	4.910	4.663	4.016	4.560	4.485	4.546	4.513	4.534
	(1.702)	(1.700)	(1.665)	(1.300)	(1.438)	(1.080)	(0.938)	(1.183)	(1.509)	(1.284)	(1.311)	(1.324)	(1.308)	(1.283)
SLD	0.444		0.433	0.473	0.604	0.603	0.555	0.436	0.350					
	(0.497)		(0.496)	(0.499)	(0.489)	(0.489)	(0.497)	(0.496)	(0.477)					
Communication	0.160		0.158	0.166	0.0952	0.0849	0.0743	0.0541	0.0411					
	(0.367)		(0.365)	(0.372)	(0.294)	(0.279)	(0.262)	(0.226)	(0.199)					
Other Health	0.0668		0.0684	0.0629	0.0541	0.0582	0.0645	0.0829	0.0964					
	(0.250)		(0.252)	(0.243)	(0.226)	(0.234)	(0.246)	(0.276)	(0.295)					
Emotional	0.0722		0.0659	0.0884	0.0695	0.0769	0.0927	0.141	0.199					
	(0.259)		(0.248)	(0.284)	(0.254)	(0.266)	(0.290)	(0.348)	(0.400)					
Other IEP Category	0.0622		0.408	0.263	0.250	0.260	0.257	0.318	0.334					
	(0.241)		(0.492)	(0.440)	(0.433)	(0.438)	(0.437)	(0.466)	(0.472)					
N	508,611	422,514	61,800	24,297	5,325	4,230	3,486	2,701	2,553	11,492	4,025	2,148	1,528	6,396

Notes: Shown above for Connecticut are sample means (with standard deviations in parentheses). Column 1 is the full sample, columns 2 and 3 represent students never or always observed in special education, and column 4 represents the main analytic sample of students first classified in grades 4-8. Columns 5-8 separate the sample in column 4 by the grade first classified. Columns 10-14 separate the sample in column 4 by initial disability classification.

B Comparing Special Education Across States

All three states studied here - Indiana, Massachusetts, and Connecticut — adhere to the federal framework under the Individuals with Disabilities Education Act (IDEA), which requires both a comprehensive evaluation and the exclusion of non-disability factors (such as limited instruction or linguistic background) when determining eligibility. The states differ somewhat in how those components are operationalized. In Indiana, Article 7 regulations place strong procedural emphasis on documenting and ruling out exclusionary factors before eligibility can be established (Indiana Department of Education, 2023). In Massachusetts, regulations similarly reference exclusionary criteria but situate them within a broader "comprehensive evaluation" process that relies heavily on multidisciplinary team judgment about educational impact (Massachusetts Department of Elementary and Secondary Education, 2023a). Connecticut's Planning and Placement Team model blends these approaches: the team must document that external factors do not primarily explain a student's difficulties while also exercising professional discretion in interpreting educational performance.

The three states also apply different funding models for special education. Indiana uses a categorical approach, distributing funding to districts on a per-student basis, with allocations that vary depending on the type and intensity of each student's disability (Indiana Department of Education, 2024). Massachusetts employs a census-based model, which assumes a fixed percentage of students in each district require special education services and provides a standardized amount per student, regardless of individual need (Massachusetts Department of Elementary and Secondary Education, 2023b). In contrast, Connecticut relies on a hybrid reimbursement model: districts initially cover all special education costs locally, and the state reimburses a portion of expenses that exceed 4.5 times the district's average per-pupil expenditure. However, this Excess Cost Grant is subject to available appropriations and does not guarantee full reimbursement (School and State Finance Project, 2024).

C Details of Difference-in-Differences Approach

We estimate cohort- and time-specific average treatment effects, $\text{ATT}_{g,t}$, following Callaway and Sant'Anna (2021). Students are grouped by the grade in which they are first placed into special education ($G_i = g$), and in each grade we compare that cohort with students who have *not yet* been classified. Because the effect appears to grow as students progress through later grades we report the full dynamic path, $\widehat{\text{ATT}}_{g,t}$. Formally, for each entry grade $g \in \{4, 5, 6, 7\}$ we estimate:

$$\widehat{\text{ATT}}_{h,g} = \underbrace{\frac{1}{N^T_{h,g}} \sum_{i\,:\,H_i=h} y_{ig}}_{\substack{\text{treated:}\\ \text{first classified}\\ \text{in grade } h}} - \underbrace{\frac{1}{N^C_{h,g}} \sum_{i\,:\,H_i>h} y_{ig}}_{\substack{\text{controls:}\\ \text{not yet classified}\\ \text{by grade } h}}, \tag{8}$$

Where y_{ig} is the outcome of student i in grade g; H_i denotes the first grade in which student i enters special education; $H_i = h$ defines the treatment cohort; $N^T_{h,g}$ is the number of students first classified in grade h who are observed in grade g (treated group); $N^C_{h,g}$ is the number of students who have not yet been classified by grade h (comparison group); T_h is the count of post-classification grades observed for cohort h. For each outcome we aggregate across grade-classified cohorts to present an overall estimate for the dynamic path of student outcomes relative to placement into special education overall.

Our analysis departs slightly from the standard application of this approach by tracking student progress across grade levels rather than calendar time. This reflects our view that, for a given student, the key temporal dimension shaping outcomes is the grade in which they first enter special education, not the school year. The identifying assumption underlying this strategy is that absent classification students who are first placed into special education in a given grade would have followed similar outcome trajectories in subsequent grades as students who had not yet been classified with a disability.