EdWorkingPaper No. 24-1010

The Effects of Response to Intervention on Disability Identification and Achievement

Nicholas Ainsworth
University of California, Irvine

Christopher Cleveland
Brown University

Andrew Penner
University of California, Irvine

Currently 15 percent of U.S. students receive special education services, a widespread intensive intervention with variable effects on students. Spurred by changes in federal policy, many states and districts have begun adopting the Response to Intervention (RTI) approach to identifying students to receive special education services. RTI seeks to provide a system for targeting interventions to children facing early academic challenges and identifying children with specific learning disabilities (SLD). This paper uses a difference-in-differences design to examine the effects of RTI adoption across Oregon on elementary students' disability identification and state-standardized achievement test scores. RTI adoption reduced special education identification by 1.4 percentage points (11%) and SLD identification by 0.5 percentage points (15%). RTI also caused moderately large reading test score gains for Black students (0.15 SD) and did not reduce other students' achievement. These findings suggest RTI is a promising approach to supporting struggling students.

VERSION: August 2024

Suggested citation: Ainsworth, Nicholas, Christopher Cleveland, and Andrew Penner. (2024). The Effects of Response to Intervention on Disability Identification and Achievement. (EdWorkingPaper: 24-1010). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/72jm-2m09

The Effects of Response to Intervention on Disability Identification and Achievement*

Nicholas Ainsworth† Christopher Cleveland‡ Andrew Penner§

Abstract

Currently 15 percent of U.S. students receive special education services, a widespread intensive intervention with variable effects on students. Spurred by changes in federal policy, many states and districts have begun adopting the Response to Intervention (RTI) approach to identifying students to receive special education services. RTI seeks to provide a system for targeting interventions to children facing early academic challenges and identifying children with specific learning disabilities (SLD). This paper uses a difference-in-differences design to examine the effects of RTI adoption across Oregon on elementary students' disability identification and state-standardized achievement test scores. RTI adoption reduced special education identification by 1.4 percentage points (11%) and SLD identification by 0.5 percentage points (15%). RTI also caused moderately large reading test score gains for Black students (0.15 SD) and did not reduce other students' achievement. These findings suggest RTI is a promising approach to supporting struggling students.

---

*We thank Aaron Ainsworth, Andy de Barros, and Emily Penner for their detailed feedback on our manuscript as well as Alejandro Granados Vargas, Young-Suk Kim, and Julie Washington for their engagement with early drafts of this paper. We also thank members of the Early Childhood Policy Lab at UC Irvine and the Center for Administrative Data Analysis at UC Irvine for their helpful feedback at various stages of this project. We are grateful to the Oregon Department of Education for sharing the data for this research. Work on this project was supported by the Integrated Research Training-Language & Literacy Disabilities fellowship granted to UC Irvine from the Office of Special Education Programs, U.S. Department of Education (H325D190031) as well as the Eunice Kennedy Shriver National Institute of Child Health and Human Development of the National Institutes of Health (R01HD094007). All views expressed are those of the authors and do not represent the views of the Oregon Department of Education, U.S. Department of Education, or National Institutes of Health.

†Nicholas Ainsworth is a PhD student in the School of Education at the University of California, Irvine (Corresponding Author: ainsworn@uci.edu).

‡Christopher Cleveland is an assistant professor in the Department of Education at Brown University.

§Andrew Penner is a professor in the Department of Sociology at the University of California, Irvine.

1 Introduction

Despite the progress made in the past several decades in advancing the rights of students with disabilities, determining how to best provide special education services presents persistent challenges. In the years since the passage of the Individuals with Disabilities Education Act (IDEA) in 1975, researchers have noted that sharp increases in disability identification have largely been unaccompanied by improvements in student achievement compared to general education students (National Center for Education Statistics, 2017; Gilmour et al., 2019; Ladner, 2021), raising questions about the effectiveness and costs of special education services delivered at scale (Morgan et al., 2010; Aron and Loprest, 2012; Ballis and Heath, 2021a; O'Hagan and Stiefel, 2024). Relatedly, scholars have worried that students, particularly low-income and students of color, are being disproportionately identified for special education due to a lack of access to effective general education instruction or broader cultural biases, leading to the provision of services that are, at best, unnecessary, and at worst, harmful to student outcomes (National Research Council, 1982; Fuchs et al., 2002; Ballis and Heath, 2023; Hart and Lindsay, 2024).[1] As such, ensuring that students are appropriately identified for special education services and provided access to high-quality instruction remains a significant concern for policymakers.

In direct response to these concerns, the 2004 reauthorization of IDEA made substantial changes to the identification practices for a specific learning disability (SLD), the most commonly identified disability in U.S. schools (National Center for Education Statistics, 2023). In particular, the reauthorization allowed states to use Response to Intervention (RTI) for SLD identification, which reconceptualizes many learning disabilities as an inadequate re-

[1] While, historically, researchers have worried that students of color were disproportionately overidentified for special education services (e.g., Dunn, 1968; Artiles and Trent, 1994; Donovan and Cross, 2002; Skiba et al., 2005) more recent research questions whether these students might be underidentified based on underlying levels of need (e.g., Hibel et al., 2010; Morgan et al., 2015, 2017). Other researchers suggest that disproportionate identification is context-dependent, with students of color being overrepresented in some school settings and underrepresented in others (e.g., Fish, 2019; Elder et al., 2021; Stiefel et al., 2023). Nevertheless, regulations issued by the U.S. Department of Education require that state educational agencies intervene when districts evidence significant disproportionality above certain thresholds, suggesting that federal policy is still primarily concerned with overrepresentation (U.S. Department of Education, 2017).

sponse to evidence-based instruction (Vaughn and Fuchs, 2003; Fuchs and Fuchs, 2006). In doing so, policymakers aimed to address outstanding problems with the prevailing method of identifying SLD, which relied upon discrepancies between a student's IQ and achievement relative to their peers. At the same time, policymakers also sought to encourage a fundamental shift in instruction and intervention within schools (Fuchs et al., 2003). Because full implementation of RTI requires the use of evidence-based core curricula, targeted small-group interventions for students making insufficient academic progress, and shifts in disability identification practices, policymakers hoped that changes brought about by RTI would universally benefit both students with and without disabilities across the school system (Fletcher and Vaughn, 2009).

Given the attractiveness of the proposition that RTI can support positive outcomes for all students, its use has become widespread over the past twenty years. Indeed, RTI is currently used in some form across all 50 states and upwards of half of U.S. elementary schools (Williams, 2022; Balu et al., 2015; Pendharkar, 2023). RTI's growing prevalence, therefore, raises questions about the impacts of RTI adoption on student outcomes and the degree to which RTI has met its intended policy objectives.

While certain aspects of the RTI model have been studied extensively (e.g., Wanzek and Vaughn, 2007; Wanzek et al., 2016), research on the impacts of system-wide RTI adoption is limited. A significant challenge for researchers in this area has been identifying exogenous variation that can be used to understand program impacts. Because RTI is a multi-faceted school or district-level reform, assigning participation at random is often infeasible (Burns, 2010). Thus, the outcomes observed in practice for schools that selectively adopt RTI may be endogenous to student and school characteristics or pre-existing trends in student outcomes, making it difficult to disentangle the causal effects of the reform. To resolve this tension, researchers have typically elected to study only components of RTI, such as the impacts of tiered interventions, using robust research designs (e.g., Balu et al., 2015; Coyne et al., 2018) or to analyze outcomes following system-wide adoptions using designs with lower internal validity, recognizing that such estimates may not represent causal effects (e.g., Wanzek and Vaughn, 2011).

In this paper, we provide one of the first causal analyses of system-wide adoption of RTI on student outcomes by leveraging the staggered rollout of the program and a longitudinal administrative data set on the population of students in the state of Oregon. Starting in the 2005-2006 school year, the Oregon Department of Education elected to gradually expand RTI across the state by selecting cohorts of districts through a program called Oregon Response to Instruction and Intervention (ORTIi). Districts received technical assistance, training, and limited funding for adopting RTI with fidelity to its core components. By 2018, almost half of all districts had participated in the program, reaching more than 40% of the total student population (ORTIi, 2023).

Taking advantage of the staggered roll-out of the program across districts, we leverage a quasi-experimental difference-in-differences design to estimate credibly causal impacts of RTI adoption on student outcomes. We find that RTI reduced special education identification rates by 1.4 percentage points (11%; $p < .01$) and SLD identification rates by 0.5 percentage points (15%, $p < .01$). Interestingly, we identify spillover effects on disability categories not directly targeted by the policy, with adoption leading to statistically significant decreases in the identification of speech or language impairment. Overall, these changes to disability identification did not accompany increases in average reading test scores, the academic area targeted by the policy, nor were there durable spillovers on math achievement or student discipline. Nevertheless, we can rule out average decreases in test scores greater than .03 standard deviations (SD), suggesting that test scores did not decline due to changes in disability identification. Finally, while there were null overall effects on achievement, we find evidence that Black students in RTI schools experienced reading test score gains of .15 SD ($p < .001$), suggesting that RTI adoption was equity-enhancing for some student populations. These results provide novel evidence on the impacts of a longstanding and understudied education policy reform as it was implemented at scale and contribute to conversations on how to structure school systems to effectively meet the needs of students with and without disabilities.

2 Background

2.1 Origins of RTI

Response to Intervention's use in schools came about from a decades-long effort to determine how to best identify and support children who face academic challenges. RTI initially arose in the field of special education as an alternative to the prevailing method used to identify specific learning disabilities known as the IQ-achievement discrepancy model, which required a significant difference between a child's achievement levels and intellectual ability to recognize them for services (Fuchs et al., 2002, 2003). The IQ-discrepancy model was widely criticized on the grounds of validity due to its inability to distinguish between the academic profiles of discrepant and non-discrepant children (Stuebing et al., 2002) and reliability due to the instability of classification decisions arising from imposing arbitrary eligibility cut points on a continuous distribution (Francis et al., 2005). Further critiques included the fact that this model typically resulted in children experiencing years of academic failure before receiving any intervention (often called the "wait-to-fail" approach) and the limited instructional relevance of the eligibility process (Vaughn and Fuchs, 2003).

RTI departs substantially from the IQ-achievement discrepancy model by attempting to intervene early for children demonstrating academic difficulties and by providing special education supports only to those children unresponsive to less intensive interventions. Notably, because of RTI's intervention and prevention focus, full implementation is conceptualized as a school-wide reform and reaches far beyond children receiving special education services (Baker et al., 2010). Central to RTI's theory of change is a multi-tiered approach to instruction and intervention. In this model, all children are supposed to be provided generally effective classroom teaching using a strong core instructional program, which is called Tier

1. Children not making satisfactory progress are identified using universal screeners and progress monitoring data and are provided Tier 2 interventions in a small group setting. Students not responding adequately to Tier 2 interventions are identified using progress monitoring data and provided increasingly intensive and individualized interventions at Tier 3 (Fletcher and Vaughn, 2009; Fuchs and Vaughn, 2012). Depending on the state, children not progressing at Tier 2 or Tier 3 are referred for evaluation for SLD (Berkeley et al., 2009; Zirkel and Thomas, 2010).

The catalyst for RTI's adoption in schools came about in 2004 with the reauthorization of IDEA. In it, the federal government shifted from encouraging the use of the discrepancy model to allowing for RTI as an alternative method for SLD identification (Fuchs and Fuchs, 2006). This reauthorization led to a proliferation of states and districts adopting RTI, with all states now allowing for the use of RTI and thirteen states requiring its use as the only method for SLD identification (Berkeley et al., 2009; Hauerwas et al., 2013; Williams, 2022). Despite the seemingly widespread use of RTI, obtaining estimates of the number of schools using RTI for prevention, intervention, and identification has proven difficult. Estimates place the use of multi-tiered systems of support (MTSS, the successor to RTI and a more general name for tiered approaches) at anywhere from fifty to seventy-five percent of the nation's schools. Many use RTI specifically for reading intervention and disability identification (Balu et al., 2015; Pendharkar, 2023). Nevertheless, large-scale surveys indicate that only one-third of school psychologists use RTI as the primary method for identifying SLD whereas twenty percent of special education administrators report using RTI as the only method for SLD identification (Maki and Adams, 2019; Lockwood et al., 2022). Thus, while Response to Intervention is increasingly popular across the U.S., its adoption in schools is not yet universal.

2.2 RTI's Impacts on Students

Because RTI incorporates several interrelated components, researchers often try to isolate how each component contributes to student outcomes. Some scholars study how enhancing Tier 1 core instruction impacts children's academic performance, typically finding positive results (e.g., Fien et al., 2015, 2021; Smith et al., 2016). More attention has been paid to the impacts of targeted interventions for students identified as needing additional supports, particularly in the area of reading. In general, there is a consensus that well-designed interventions for struggling readers improve their reading performance. Multiple meta-analyses and systematic reviews find that interventions similar to those provided at Tiers 2 and 3 yield positive effects for at-risk or struggling readers, especially those in younger grades (Wanzek and Vaughn, 2007; Wanzek et al., 2016, 2018; Gersten et al., 2017b, 2009, 2020). Many of these studies, however, rely upon small-scale researcher-designed interventions and do not reflect what happens when schools attempt to provide interventions themselves at scale.

Notably, the impacts of school-led intervention efforts reveal a more complicated picture. The most widely cited analysis of RTI interventions delivered at scale is that of Balu et al. (2015), commissioned by the Institute of Education Sciences in 2010. The team of researchers utilized a sample of 146 schools across 13 states that assigned students to Tier 2 and Tier 3 interventions using a cut score to examine impacts on reading outcomes using a regression discontinuity design (RDD). For students in 2nd and 3rd grade, they found null results. However, for 1st grade students, assignment to intervention led to significantly negative impacts on reading test scores that were moderately large in magnitude (-0.17 SD). In reconciling these findings with the positive results often found for early literacy interventions, researchers cited problems with implementation fidelity and the difficulties of scaling RTI in some school contexts. These researchers also noted that the scope of the evaluation was narrow, given that the RDD only allowed for the estimation of intervention impacts for students local to the cut score, and could not speak to the overall impacts of RTI adoption on a broader set of students and outcomes (Fuchs and Fuchs, 2017; Arden et al., 2017;

Gersten et al., 2017a).

While the question of RTI's impacts on an entire student population when adopted as a comprehensive school or district-wide reform is arguably the most relevant for judging RTI's overall success, the literature on this topic is limited. Because identifying exogenous variation in RTI adoption is difficult, researchers have frequently turned to describing pre-post trends in outcomes or making cross-sectional cohort comparisons within a few schools or a single district. Most of these smaller-scale studies focus on disability identification outcomes, typically finding evidence of lower identification or referral rates for cohorts exposed to RTI (VanDerHeyden et al., 2007; Wanzek and Vaughn, 2011; O'Connor et al., 2013). Nevertheless, given their small sample sizes, they often fail to detect statistically significant differences and, thus, portray changes in identification as largely descriptive. Some of these studies also look at achievement outcomes, with some finding higher scores for cohorts exposed to RTI but others not (e.g., O'Connor et al., 2013; O'Connor et al., 2014; Grapin et al., 2019). Overall, an early meta-analysis of the effects of different RTI models implemented prior to changes in IDEA policy in smaller-scale contexts found positive effects on student achievement and reductions in disability identification (Burns et al., 2005).

Only a handful of studies have examined the impacts of RTI adoption across a large number of students. In one such study, Torgesen (2009) examined changes in reading achievement and disability identification among students enrolled in 318 Reading First schools in the state of Florida. He found large decreases in SLD identification (40%) for 3rd graders in year 3 of implementation compared to year 1 and a 30% decline in 3rd graders evincing reading difficulties over the same time period. Using statewide administrative data for the population of students in Tennessee, Gilmour et al. (2023) found reductions in overall special education and SLD identification in grades K-5 after statewide adoption, with declines that were larger for Black and low-income students. While these studies both make important contributions to understanding the relationship between RTI adoption and student outcomes for large populations of students, given their pre-post designs they cannot rule out the possibility that the differences they found were driven by unobserved confounding factors beyond RTI adoption. Thus, questions remain regarding the causal effect of full-scale adoption of RTI on disability identification and student achievement.[2]

In summary, RTI's emergence as an alternative method for SLD identification has led to the development of a school-wide reform model for improving the performance of all students. In response to federal policy change, RTI's use across states and districts in the past few decades became widespread, although adoption of the model is not yet universal. Extant research suggests that interventions in an RTI framework can improve student reading outcomes and may reduce disability identification. Nevertheless, there is limited causal evidence about how adoption at scale influences student outcomes and the small set of existing studies point to effects in different directions. Thus, examining how RTI adoption and implementation impacts student outcomes as practiced in schools at scale is essential for fully understanding the effects of this reform. We turn now to describing the context of RTI adoption in the state of Oregon to motivate carrying out such an analysis.

2.3 The Policy Context in the State of Oregon

The state of Oregon presents a unique opportunity to study the impact of RTI adoption. Starting in the fall of 2005 the Oregon Department of Education (ODE) decided to roll out RTI by providing training and technical assistance to cadres of districts selected by the state. ODE contracted with the Tigard-Tualatin School District, a district that had successfully implemented RTI for several years, to provide assistance to the selected districts. This effort came to be known as the Oregon Response to Instruction and Intervention (ORTIi) project and is still in operation as of 2024. As a part of this project, school districts apply to receive technical assistance and professional development from ORTIi to adopt RTI. In

[2]One recent study uses a difference-in-differences design to examine the effects of state-level RTI policy changes on disability identification (Shea and Jenkins, 2024) Notably, the state-level policies examined in this study differed in their correspondence to RTI implementation (e.g., in some states policies mandated the use of certain components of RTI, while in other states policies simply allowed or encouraged districts to adopt these components). As such, we view this study as focused on the effects of state-level policy changes rather than full-scale implementation of the RTI model.

doing so, the districts commit to implementing the RTI model district wide. There are several requirements of districts participating, including the use of RTI to identify SLD, the adoption of a research-based core instructional program, the use of universal screeners and progress monitoring tools, the development of a data management system as well as the collection of student outcomes data and implementation fidelity data (ORTIi, 2023). Although there was funding attached to the technical assistance and professional development, districts needed to be able to support the continued use of RTI without the state's financial support (Stepanek and Peixotto, 2009).

As conceived by the state, ORTIi is primarily concerned with early reading outcomes and SLD identification. Indeed, the explicit goals of the program reference increasing the percentage of students reading proficiently by 3rd grade and beyond, reducing the percentage of students receiving the most intensive level of support through SLD identification, and ameliorating inequities in these outcomes by race/ethnicity and socioeconomic status (ORTIi, 2023). Our three research questions are guided by these aims, asking:

1. How does RTI adoption impact student identification for special education?
   (a) What are the effects on specific learning disability identification?
   (b) Are there spillover effects on identification for other disabilities?
2. How does RTI adoption impact student reading performance as measured by state standardized test scores?
   (a) Are there spillover effects on math achievement or student discipline?
3. Do the effects of RTI vary by student race/ethnicity and socioeconomic status?

3 Data

The data for this project come from the Oregon Department of Education's administrative data for the universe of students enrolled in Oregon public schools from the 2004-2005 to

2021-2022 school years. These data are at the individual student level and include information on each student's school and district enrollments, demographics, state standardized test scores, attendance, and disciplinary incidents. In addition, ODE provided a detailed special education file starting in the 2007-2008 school year that, for each student with an Individualized Education Program (IEP), contains their specific disability eligibility category. This file is constructed using information collected during the fall of each school year and allows us to observe each student's primary disability classification. For analysis, we restrict the data to elementary school students (grades K to 5) with non-missing school enrollment and demographic information from the 2008 to 2022 school years.[3] This gives us over 3.3 million student-year observations with an average annual N of around 220,000.

We use these administrative data to construct several outcomes of interest. First, we use a student's primary disability code to construct indicators for the identification of each disability category in Oregon as well as an indicator for any special education service receipt in a given school year. For achievement outcomes, we rely upon state standardized test scores from summative assessments administered by the Oregon Department of Education in the spring of each school year that measure mastery of grade-level content standards. Test score outcomes are only measured for students beginning in grade 3, so we use test scores from students in grades 3 through 5, which we standardize within subject, grade, and year using the entire population of student test takers. For disciplinary outcomes, we use incident-level data on suspensions to construct an indicator for whether a student received one or more suspensions within a given school year. Due to disruptions caused by the COVID-19 pandemic, test score data are unavailable for the 2020 and 2021 school years while discipline outcomes are unavailable for 2021. Those years are dropped from the achievement and suspension analyses, respectively.

[3] In addition to excluding students with missing demographic and enrollment information, we also exclude students enrolled in schools that switched district IDs during the panel. This restriction drops 12 schools and 0.17% of the student sample. Overall, these combined restrictions exclude 0.75% of the total sample. Our main estimates are not sensitive to retaining these observations nor are there effects of RTI on exclusion rates based on this sample restriction (results available upon request).

To evaluate the impacts of RTI on these student outcomes, we construct a novel panel data set on program adoption and participation. Notably, on the ORTIi website the organization publishes a list of adopting districts as well as the cohort and school year of adoption. Utilizing this list, we constructed variables at the district-year level indicating whether the district had been "treated" by RTI. We define treatment as any year during and after participation in the ORTIi program. While formal programming only lasts for a few school years, developing the infrastructure to implement RTI represents an organizational change that is likely to persist. Moreover, the program's intent was to help districts make the initial switch to using RTI and for official support from program staff to fade over time so that districts could operate these models independently. We merge these adoption data with the individual-level student data using district of enrollment to create a novel 15-year panel data set to use for the analysis of RTI's impacts.

3.1 Descriptive RTI Adoption Trends

To better understand how RTI was rolled out across the state of Oregon, we present information on RTI adoption trends in Figure 1. In Panel A of this figure, we plot the number of districts and schools newly adopting RTI from 2008 to 2022. Because we do not have access to detailed special education information until 2008, we restrict our analyses to RTI adoptions that occurred in 2008 or later, though we note that the first year of formal program adoption was the 2005-2006 school year.[4] As shown in Panel A, the mid to late 2010s saw a large expansion of the program, rising from 6 districts newly adopting RTI in 2010 to 20 districts in 2015. Notably, districts that adopted RTI were not of equal size, with the patterns of adoption at the school level differing from the counts at the district level. For example, while 12 districts newly adopted RTI in 2018, this was the largest year of new schools adopting the program at 75. Similarly, while only 6 districts adopted RTI in 2010, this represented an increase of 54 schools. After 2018, no new schools or districts adopted

[4] In 2006, 4 districts (23 schools) adopted RTI while 9 districts (51 schools) adopted RTI in 2007.

RTI through participation in the ORTIi program.

While school and district-level adoption provides important insights into the spread of RTI across the state, it may misrepresent the true reach of the program. This is because schools and districts do not serve equal-sized student populations. Thus, if RTI tended to be adopted in small districts, a large proportion of districts may have participated in the program, but the proportion of the state's students impacted may be much smaller. The converse is true if larger districts disproportionately adopted the program. In Panel B of Figure 1, we plot the cumulative number of kindergarten through 5th grade students in districts that had adopted RTI versus the number of K-5 students in districts that had not adopted RTI by academic year. We also present the cumulative percentage of Oregon students in participating districts. Panel B demonstrates that the program has expanded dramatically in its reach over time. In 2008, only 12% of Oregon's K-5 students were in RTI districts, but by 2018 that figure hit 47%. This also meant that RTI was impacting over 100,000 students annually by 2018. Thus, RTI has made substantial inroads in Oregon's public schools and now impacts a meaningfully large segment of the state's student population.

3.2 Sample Description

In Table 1, we report the average grade 3-5 student achievement, K-5 student demographics, and school characteristics for students in districts that never adopt RTI (i.e., never adopters) and districts that ever (or eventually) adopt RTI (i.e., ever adopters) in the 2007-2008 school year. We also calculate the difference in means between never-adopters and ever-adopters and test the statistical significance of these differences using OLS regression models that regress each characteristic on an indicator for ever adopting RTI, with point estimates and district-clustered standard errors presented in column 3. We restrict sample characteristics to the 2007-2008 school year because, once we exclude districts that adopted RTI in 2008, we are able to make comparisons prior to the potential influence of RTI. This is important because several characteristics, such as disability identification and achievement, are hypothesized to be affected by RTI adoption.

Ever RTI adopters and never RTI adopters are similar on many dimensions, and of the 25 differences in characteristics examined, only 4 are statistically distinguishable from 0 at the .10 level. Most notably, students in districts that eventually adopt RTI have lower average achievement in both reading (-.17 SD; $p < .001$) and math (-.21 SD; $p < .001$) compared to students in never adopting districts. In terms of student characteristics, students in RTI-adopting districts are 3.7 percentage points less likely to identify as Asian/Pacific Islander ($p < .01$) and 8 percentage points more likely to identify as low income ($p < .01$). There were no statistically significant differences across other characteristics, including the likelihood of receiving special education, school size, and whether the district was rural, urban, or suburban.

4 Methods

To estimate the impacts of RTI adoption on elementary student outcomes, we utilize a quasi-experimental difference-in-differences (DiD) design. This design is meant to address two challenges to inference in non-experimental settings. First, comparing within unit changes pre- and post-intervention may confound the impact of the intervention with secular changes in an outcome over time. For example, if one observes higher test scores for students in RTI schools after adoption, this may reflect the program's impacts or a more general trend in test scores rising over time due to some other cause. Second, making cross-sectional comparisons between RTI adopters and non-adopters may confound differences in outcomes with other characteristics that are related both to adoption and outcomes. If, for example, schools with lower test scores select into RTI (as is the case in Oregon), comparing the outcomes of students in those schools to non-adopters will confound program effects with pre-existing differences in test scores, leading to biased inferences about RTI's impact.

DiD overcomes these two challenges to inference by leveraging trends, rather than levels, in outcomes for estimation. Specifically, DiD uses the trend in outcomes among the untreated units to project what the trend would have been for treated units had they not received treatment. Treated units are then compared to this projected value as the counterfactual to estimate treatment impacts. Because differences in levels are irrelevant for making this projection, treated and non-treated units can differ in their levels of the outcome prior to treatment (i.e., there can be treatment selection on levels of the outcome variable). However, treatment should be exogenous to *trends* in the outcomes for inference to be unbiased (Roth et al., 2023). As such, the critical assumption underlying DiD is parallel trends. That is, the design assumes that treated units would have followed the same trend in outcomes as the untreated units had they also not received treatment. This assumption is untestable because one does not observe the counterfactual trend for treated units in the post-treatment period. Nevertheless, evidence for the plausibility of this assumption can be generated by examining the trends in outcomes prior to treatment to determine whether treated and untreated units appeared to be following parallel trends (de Chaisemartin and D'Haultfœuille, 2023).

The traditional approach for implementing a difference-in-differences design is to estimate a two-way fixed effects (TWFE) model that takes the following form:

$$Y_{st} = \theta_s + \delta_t + \beta RTI_{st} + \epsilon_{st} \tag{1}$$

where Y_{st} is the outcome of interest indexed for school s in year t, θ_s is a school fixed effect, δ_t is a year fixed effect, and ϵ_{st} is an error term. RTI_{st} is an indicator variable that takes on a value of 1 for treated schools in all years after school-wide participation in the RTI program and 0 otherwise. Thus, β represents the average effect of RTI on student outcomes in the post-adoption years and can be thought of as an estimate of the average treatment effect on the treated (ATT). Model 1 can be extended by replacing RTI_{st} with indicators for relative time to treatment to produce event study estimates that allow for the examination of the plausibility of the parallel trends assumption and the possibility of dynamic treatment effects (i.e., those that change over time post-adoption).[5]

There is a growing recognition in the econometrics literature, however, of the potential problems with utilizing TWFE estimators in situations involving staggered treatment timing and heterogeneous treatment effects (see Roth et al., 2023, for a recent overview). A major problem in this context is the assignment of negative weights. As shown by de Chaisemartin and D'Haultfœuille (2020) and Goodman-Bacon (2021), among others, TWFE estimators with staggered treatment timing may assign negative weights for some group-time comparisons in the presence of heterogeneous treatment effects based on timing of adoption, with negative weights typically applied to longer-run treatment effect estimates. In some extreme cases, this may lead coefficients to reverse signs entirely. The reason this negative weighting problem occurs is that static TWFE estimators can make "forbidden comparisons," or comparisons between newly treated units and earlier treated units (Goodman-Bacon, 2021). Notably, these problems may not be solved with dynamic TWFE estimators, which can also assign negative weights to some treated units and can cause cross-lag contamination, with some lagged treatment effect estimates being influenced by estimates for earlier treated years (Sun and Abraham, 2021). Together, these problems can lead to substantial bias in the estimation of treatment effects and a set of weighted comparisons that are not intuitive or practically relevant in many policy contexts (Roth et al., 2023).

To address these issues, econometricians have developed several new DiD estimators, including the Callaway and Sant'Anna (2021) estimator we implement here. Callaway and Sant'Anna (CS) take as the foundation of their estimator what they call the "group-time

[5] The event study model is specified as follows:

$$Y_{st} = \theta_s + \delta_t + \sum_{\tau=2}^{n} \beta_{-\tau} RTI_{s,t-\tau} + \sum_{\tau=0}^{n} \beta_{\tau} RTI_{s,t+\tau} + \epsilon_{st}$$

where all terms are defined the same as equation 1 with the exceptions of RTI_{st} and β. Here, $RTI_{s,t\pm\tau}$ are relative time to treatment indicators that take on a value of 1 for schools that are τ years from adopting RTI. As such, $\beta_{-\tau}$ represents the difference of being τ years from RTI adoption compared to never adopters or being 1 year prior to adoption for RTI adopters. If $\beta_{-\tau}$ are statistically indistinguishable from 0 for the pre-treatment years, this can provide evidence in support of the plausibility of the parallel trends assumptions. $\beta_{+\tau}$ are the coefficients of interest and represent the effect of RTI τ years after adoption.

average treatment effect" which are separately estimated Average Treatment Effects on the Treated ($ATT(g, t)$) for each group g (defined by year of initial treatment) and time t (defined by calendar time).[6] Each of these differences is calculated with respect to an explicit comparison group, which can be specified as either never treated or not-yet treated units. Under similar assumptions regarding parallel trends, these $ATT(g, t)$ represent clearly defined causal parameters that are akin to the canonical DiD set up with two periods and two groups. Indeed, when no covariates are incorporated, estimates of the $ATT(g, t)$ may be obtained as β from estimating equation 1 with data subset to include only observations from treated cohort (g) and the comparison group at times t and $g - 1$.[7] An attractive feature of the CS estimator is that these group time average effects can then be aggregated and weighted to provide an estimate of the ATT for the entire post-treatment period as well as for years defined as time relative to treatment (i.e., to produce event study estimates). By specifying the reference group and applying an explicit weighting scheme to aggregate treatment effect estimates, the CS estimator avoids the negative weighting problem and produces an interpretable causal estimand.

To implement the CS estimator, we collapse the data to the school-by-year level and weight by the school's student population size. We select the never treated schools as the comparison group and drop always treated schools from the analysis sample.[8] Standard errors are clustered at the district level, as that is the level at which treatment was assigned (e.g., MacKinnon et al., 2023). In the main text, we provide estimates of the ATT by

[6]The $ATT(g, t)$ estimand is defined as follows:

$$ATT(g, t) = E[Y_{i,t} - Y_{i,g-1}|G_i = g] - E[Y_{i,t} - Y_{i,g-1}|G_i = G_{comp}]$$

where G_{comp} represents the comparison group (either never treated or not-yet treated units).

[7]One notable feature of the CS estimator is that it relies upon a different parallel trends assumption compared to other estimators. Specifically, the CS estimator assumes parallel trends hold only in the post-treatment period and, consequently, makes all post-treatment comparisons to the last pre-treatment year ($g - 1$). This feature is different from other recent estimators (see, for example, Sun and Abraham, 2021; Gardner, 2022) and should be taken into consideration when interpreting estimates produced by the CS estimator (see Roth et al., 2023, for a discussion of the advantages and disadvantages of this parallel trends assumption).

[8]We implement the Callaway and Sant'Anna (2021) estimator with the `csdid` package written by Rios-Avila et al. (2021) using Stata 17.

aggregating the group time average effects for the entire post-adoption period, though we also present event study estimates for some outcomes (with others presented in the appendix). Although we treat the CS estimator as our preferred specification, in the appendix we present sensitivity tests of our main estimates to utilizing the TWFE model specified in equation 1 as well as the two-stage DiD estimator developed by Gardner (2022), showing that these approaches provide similar estimates. We turn now to presenting the results of the effects of RTI adoption on student outcomes in the state of Oregon.

5 Results

5.1 Special Education Identification

To understand the impacts of RTI adoption on disability outcomes, we first examine the effects on overall special education (SpEd) identification. In Figure 2, we present an event study plot with relative time to RTI adoption on the x-axis and the difference in SpEd identification rates for students in grades K through 5 on the y-axis. A dashed line is plotted at time 0 to represent the first year a school experienced treatment. Ninety-five percent confidence intervals are displayed for each pre- and post-treatment point estimate.

As shown in Figure 2, RTI adoption had a substantial effect on special education identification. In the first 3 years after adoption, RTI schools saw around a 1 percentage point decline in SpEd identification rates. In all post-treatment years, with the exception of the first year of treatment (year 0), the ninety-five percent confidence intervals do not overlap with 0, indicating that the effects are statistically significant at the .05 level. Notably, the effects of RTI adoption grow over time, reaching a 2 percentage point decline in SpEd identification 5 to 8 years after adoption and almost 3 percentage points by 10 years out. Given that the average SpEd identification rate for treated schools prior to RTI adoption was 13.7%, these changes represent 7% to 22% declines compared to baseline identification rates. Looking at the pre-treatment period, all point estimates are close in magnitude to 0 while none are statistically distinguishable from 0 at the .05 level. This suggests that the observed declines in SpEd identification after RTI adoption reflect changes brought about by participation in the program rather than the continuation of a preexisting trend in disability identification. If anything, SpEd identification rates were trending slightly upwards prior to RTI adoption, with this trend reversing entirely once schools began participating in the program.

Although these event study estimates make clear that special education identification declined in RTI schools after program adoption, they raise questions regarding which disability categories were impacted. Because RTI was expressly designed as an alternative approach for identifying specific learning disabilities, one might expect SLD to be the only (or primary) disability impacted by its adoption. Nevertheless, researchers have identified RTI impacts on other disability categories (e.g., Gilmour et al., 2023), suggesting that the examination of other categories is warranted. In Table 2, we present the effects of RTI adoption on the identification rates for the six most prevalent school-based disabilities in Oregon. Combined, these disability categories capture almost 96% of all elementary school students receiving SpEd services in the state. Columns represent separate models with the aggregated average treatment effect on the treated (ATT) post-RTI adoption recorded in each cell. We also display mean identification rates for each disability in RTI schools prior to adoption at the bottom of the table.

Turning first to overall SpEd identification impacts, we find that the declines observed in the event study plot in Figure 2 represent an average decline of 1.44 percentage points ($p < .01$) post-RTI adoption, or 11% of the pre-treatment mean. Next, we examine impacts on SLD identification, the primary disability targeted by the policy. As shown in Table 2, RTI adoption led to an average decline in SLD identification rates of 0.54 percentage points ($p < .01$). This represents a 15% decrease compared to the average pre-treatment SLD identification rate in RTI schools, which was 3.5%. In Appendix Figure A1 we also plot event study estimates for SLD identification, which reveal a near constant treatment effect post-RTI adoption of around 0.5 percentage points. Thus, RTI appears to have reduced SLD identification as intended, though effects on SLD only account for a little over a third of the overall effect on special education receipt.

Table 2 also reveals evidence of spillover effects on other disability categories. In particular, we find spillover effects on the identification of speech or language impairment (SLI). Post-RTI adoption, SLI identification rates for K-5 students in treated schools declined by 0.71 percentage points ($p < .05$). This represents a 12% drop relative to the average SLI rate of 5.7% in RTI schools. Event study estimates presented in Figure A1 also show that the RTI impacts on SLI identification grew over time, with decreases of around 0.5 percentage points in the first several years after adoption that reach 2 percentage points by 10 years post-adoption. Thus, it appears that approximately half of the overall decline in SpEd identification rates is driven by reductions in services for speech or language impairment and that the dynamic decreases in identification observed in the event study plot are explained by these drops in SLI.

Beyond SLD and SLI, RTI did not have consistent impacts on other disability categories. For emotional behavior disability (EBD), autism (ASD), and intellectual disability (ID), we find that both the overall effect estimates reported in Table 2 and the event study estimates in Figure A1 were consistently insignificant. By contrast, the event study estimates of RTI's effect on receiving an other health impairment (OHI) classification presented in Appendix Figure A1 suggest that RTI led to statistically significant declines in OHI identification for a few years after adoption, but that these effects faded over the long term, with estimates trending towards 0 by 8 to 10 years later. As a result, the average effect reported in Table 2 is not statistically significant. RTI, therefore, appears to have reduced disability identification when adopted in Oregon schools in part through its intended reductions to SLD classification as well as through spillover effects on SLI classification.

5.2 Academic Achievement and Student Discipline

We turn now to investigating whether the impacts of RTI adoption on disability identification extended to other academic or disciplinary outcomes. Of particular interest are the impacts on reading achievement, as reading was the primary academic area targeted by the policy. We also examine whether there were spillover effects on math achievement and suspensions. These results are presented together in Table 3. Test score outcomes are restricted to students in grades 3 to 5 (i.e., the elementary school grades with test score data available) while suspension outcomes include all students from grades K to 5. Similar to Table 2, we also report the mean for treated schools prior to RTI adoption.

Overall, RTI had limited effects on these other outcomes. As shown in Table 3, RTI had null effects on reading and math achievement as well as suspension rates. Each of the point estimates are close to 0 and are fairly precisely estimated. This precision enables us to rule out modest changes in outcomes, particularly for achievement. For example, with a point estimate of .0033 SD and a standard error of .0164, a ninety-five percent confidence interval can rule out reading test score declines larger than .029 SD and increases larger than .035 SD. We can also rule out math test score declines larger than .039 SD. Effects of that magnitude are typically considered small for most educational interventions (e.g., Kraft, 2020). In Appendix Figure A2, we explore whether the null average post-treatment impacts mask dynamic changes in these outcomes over time. No noteworthy patterns emerge. As such, RTI adoption appears not to have resulted in significant changes in the overall academic achievement or disciplinary outcomes of students who experienced this policy change.

5.3 Heterogeneity by Race/Ethnicity and Family Income

Beyond impacting disability identification rates and student achievement, a central goal of RTI as a policy and its implementation in the context of Oregon schools was to ameliorate disparities in these outcomes between advantaged and disadvantaged students. In particular, RTI sought to reduce disproportionality in special education identification and to improve academic outcomes for students of color and students from low-income families. Figure 3 presents a coefficient plot with the impacts of RTI on overall SpEd identification, reading and math achievement, and suspensions for six different racial/ethnic student groups. In the same plot, we also present estimates for students who are and are not from low-income families, which we define as qualifying for free or reduced price lunch as indicated in the Oregon administrative data. Point estimates are from separately estimated models subset to each student subgroup using the CS estimator with ninety-five percent confidence intervals plotted around each point estimate.

Examining special education identification, we observe that most student subgroups experienced changes post-RTI adoption. For all students, the points estimates for disability identification are negative while they are statistically different from 0 ($p < .05$) for six out of the eight subgroups. We find that many of the marginalized student groups in the state experienced relatively large declines in disability identification. American Indian/Alaska Native (AIAN) students, the population with the highest overall disability identification rates in the state, saw the largest decline at 3.5 percentage points (though the estimate for AIAN students is only marginally significant; see Appendix Table A2). Black, Multiracial, and low-income students all saw declines of around two percentage points, while Hispanic, White, and higher-income students had approximately one percentage point declines. However, due to the size of the standard errors, none of the declines for students of color or for low-income students are statistically distinguishable from the identification changes for White and higher-income students, respectively.

Figure 3 also reports subgroup-specific effects of RTI on academic achievement and suspensions. We find clear evidence that Black students' reading test scores increased under RTI. Where most student groups have relatively small, statistically insignificant positive point estimates, Black students in RTI schools experienced moderately large and statistically significant average reading test score increases of .146 SD ($p < .001$). This increase occurred immediately after RTI adoption and does not appear to be the continuation of a preexisting reading test score trend (see Appendix Figure A3). Beyond the positive effects of RTI on Black students, we see limited evidence of subgroup heterogeneity. For suspension rates, point estimates across all subgroups are close to 0, suggesting that no student subgroup was particularly impacted. Point estimates are likewise small in magnitude for math achievement. Taken together, the results on academic and suspension outcomes highlight that there were no groups that appear to have been harmed by the implementation of RTI, and that Black students benefited substantially.

6 Discussion

Response to Intervention's inclusion in the reauthorization of IDEA in 2004 marked a new chapter in the identification of specific learning disabilities as well as the approach schools take to intervention and instruction more broadly. In the two and a half decades preceding RTI, SLD rates were climbing rapidly, leading observers to worry about the possible inappropriate identification of learning disabilities and the potential costs associated with student mis-identification (Fuchs et al., 2002; Preston et al., 2016). Since the 2004 reauthorization of IDEA, however, national SLD rates have declined in every subsequent year (see Gilmour et al., 2023), leading observers to wonder whether RTI has caused these declines (Fuchs and Vaughn, 2012). At the same time, researchers have viewed the legacy of RTI as entwined with the broader educational policy changes of its time, most notably the shifts in attention to raising the achievement of lower-performing students brought about by No Child Left Behind. Indeed, some argue that RTI was more aligned to meeting these objectives both by design (Kavale and Spaulding, 2008) and as implemented in practice (Fuchs et al., 2010). RTI also served as the foundation for the development of Multi-Tiered Systems of Support (MTSS), which have become among the most popular school-based models for providing academic and behavioral supports to students (Bailey, 2019; Pendharkar, 2023). Despite these legacies, however, determining the extent to which RTI has causally contributed to changes in disability identification rates and student achievement outcomes since its creation has proven difficult.

In this paper, we provide one of the first causal analyses of the impacts of RTI on both special education identification and academic achievement. Examining the staggered rollout of RTI in Oregon, we find that RTI adoption led to significant decreases in overall special education identification rates (1.4 percentage points, 11% on average). This finding is consistent with theoretical expectations and extant empirical work on small and large-scale RTI adoption (Torgesen, 2009; Wanzek and Vaughn, 2011; Gilmour et al., 2023). Interestingly, however, only a third of this decline was driven by reductions in SLD identification. This stands in contrast with prior research that found that declines in SLD accounted for the vast majority of the reductions in special education service receipt in other contexts (Gilmour et al., 2023). Instead, we see that half of the disability identification declines in Oregon are driven by reductions in services for speech or language impairment. This result is surprising, particularly given that RTI only expressly targeted the identification procedures for SLD and not other disability categories.

One reason this pattern may occur is the high co-occurrence of SLD with SLI. Estimates of the comorbidity of these disabilities range anywhere from 10% to upwards of 50% (Catts et al., 2005; Snowling et al., 2019, 2020). Given the overlap of these disability categories and the subjectivity inherent in the special education referral process (Lloyd et al., 1991; Fish, 2017), it is possible that the referral decisions of school staff for special education more generally were responsive to the shifting policy context under RTI. For example, in the face of uncertainty regarding how the challenges students presented in the classroom mapped onto disability categories, educators may have elected to refer students for the RTI process instead of making a direct referral for a special education evaluation. To the extent that lower referral rates ultimately result in lower identification rates, this could be one reason why we observe declines in SLI designations. However, this is among many potential explanations, and our study cannot speak to the mechanisms that drive these spillover effects. Future research might identify these potential mechanisms and determine whether these patterns hold across the U.S.

Although prior research has found that declines in disability identification rates were harmful to student outcomes in some contexts (Ballis and Heath, 2021a), we find no evidence of declines in student achievement following RTI adoption in Oregon, despite large declines in disability identification rates (11% overall and 15% for SLD). For both reading and math achievement, we find null effects and are able to rule out meaningful declines in test scores (.03 SD in reading and .04 SD in math). Given that special education services are generally seen as effective for the marginal student (O'Hagan and Stiefel, 2024), this might suggest that the instructional supports put into place by RTI were sufficient to meet the academic needs of the marginal students who were not identified for services due to RTI adoption. This is notable, especially because policy-induced decreases in SpEd identification in other states resulted in worse academic outcomes, likely because students who lost out on services were not provided any additional supports (Ballis and Heath, 2021a). Therefore, to the extent that the state is able to reduce SpEd identification without any trade-offs for student achievement, this is likely a positive outcome. This is particularly true for state budgets given the additional costs of educating students with disabilities (Banks, 2020) but also potentially for the long-term outcomes of the students who avoided unnecessary special education services (e.g., Ballis and Heath, 2021b, 2023). However, whether the decline in SpEd services in elementary school due to RTI adoption has implications for long-term student outcomes is a question for future research.

Finally, our results reveal an interesting pattern of effects for student subgroups in the state. In general, more marginalized students experienced larger declines in disability identification rates. Reductions were the largest for American Indian/Alaska Native, Multiracial, and Black students as well as for students from low-income families. Yet, these larger declines in special education designations related to improved reading achievement outcomes only for Black students. Black students in RTI schools experienced reading test score increases of .15 SD, which are medium-sized for a policy operating in schools at scale (e.g., Kraft, 2020) and comparable in magnitude to other effective literacy reform policies (e.g., Novicoff and Dee, 2023). The fact that achievement impacts are not directly related to the size of the disability declines across subgroups suggests that different components of the RTI model may be operating to impact different segments of the student population.

7 Conclusion

Among the changes made to federal special education policy in the past two decades, perhaps none has been as significant as the shift to Response to Intervention for specific learning disability identification. Beyond dramatically changing identification practices to address outstanding concerns regarding the over-identification of SLD (Preston et al., 2016), the instructional changes implied by the RTI model encouraged fundamental shifts in the organization of instruction and intervention within schools with the aim of improving outcomes for all students (Fletcher and Vaughn, 2009). Even though RTI as a policy attempted to address both of these goals, little research to date has examined how well RTI met these objectives as it was brought to scale in schools nationwide (but see descriptive evidence from Torgesen, 2009; Gilmour et al., 2023). Our paper contributes to this conversation by providing the first causal analyses of the disability identification and achievement impacts of large-scale RTI adoption in the state of Oregon. We find that RTI reduced special education identification, including SLD, without decreasing overall (or subgroup specific) academic achievement. Although RTI did not yield overall improvements in academic outcomes as intended, it increased reading achievement for Black students, thus meeting its aim of boosting equity. On balance, our findings indicate that RTI in Oregon achieved its intended disability identification outcomes and was at least partially successful in ameliorating inequities, highlighting the promise of the RTI model for effectively supporting students at scale.

References

Arden, S. V., Gandhi, A. G., Zumeta Edmonds, R., and Danielson, L. (2017). Toward More Effective Tiered Systems: Lessons From National Implementation Efforts. *Exceptional Children*, 83(3):269–280.

Aron, L. and Loprest, P. (2012). Disability and the Education System. *The Future of Children*, 22(1):97–122.

Artiles, A. J. and Trent, S. C. (1994). Overrepresentation of Minority Students in Special Education: A Continuing Debate. *The Journal of Special Education*, 27(4):410–437.

Bailey, T. R. (2019). Is MTSS the new RTI? Depends on Where You Live.

Baker, S. K., Fien, H., and Baker, D. L. (2010). Robust reading instruction in the early grades: Conceptual and practical issues in the integration and evaluation of Tier 1 and Tier 2 instructional supports. *Focus on Exceptional Children*, 42(9):1.

Ballis, B. and Heath, K. (2021a). The Long-Run Impacts of Special Education. *American Economic Journal: Economic Policy*, 13(4):72–111.

Ballis, B. and Heath, K. (2021b). Special education: Beneficial to many, harmful to others. *Brookings*.

Ballis, B. and Heath, K. (2023). The Long-Run Impacts of Reducing Racial Gaps in Special Education.

Balu, R., Zhu, P., Doolittle, F., Schiller, E., Jenkins, J., and Gersten, R. (2015). Evaluation of Response to Intervention Practices for Elementary School Reading. Technical Report NCEE 2016-4000., National Center for Education Evaluation and Regional Assistance.

Banks, J. (2020). Examining the Cost of Special Education. In *Oxford Research Encyclopedia of Education*.

Berkeley, S., Bender, W. N., Gregg Peaster, L., and Saunders, L. (2009). Implementation of Response to Intervention: A Snapshot of Progress. *Journal of Learning Disabilities*, 42(1):85–95.

Burns, M. K. (2010). Response-to-intervention research: Is the sum of the parts as great as the whole. *Perspectives on language and literacy*, 36(2):13–15.

Burns, M. K., Appleton, J. J., and Stehouwer, J. D. (2005). Meta-Analytic Review of Responsiveness-To-Intervention Research: Examining Field-Based and Research-Implemented Models. *Journal of Psychoeducational Assessment*, 23(4):381–394.

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

Catts, H. W., Adlof, S. M., Hogan, T. P., and Weismer, S. E. (2005). Are Specific Language Impairment and Dyslexia Distinct Disorders? *Journal of Speech, Language, and Hearing Research*, 48(6):1378–1396. Publisher: American Speech-Language-Hearing Association.

Coyne, M. D., Oldham, A., Dougherty, S. M., Leonard, K., Koriakin, T., Gage, N. A., Burns, D., and Gillis, M. (2018). Evaluating the Effects of Supplemental Reading Intervention Within an MTSS or RTI Reading Reform Initiative Using a Regression Discontinuity Design. *Exceptional Children*, 84(4):350–367.

de Chaisemartin, C. and D'Haultfœuille, X. (2020). Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects. *American Economic Review*, 110(9):2964–2996.

de Chaisemartin, C. and D'Haultfœuille, X. (2023). Credible Answers to Hard Questions: Differences-in-Differences for Natural Experiments.

Donovan, M. S. and Cross, C. T. (2002). *Minority Students in Special and Gifted Education.* National Academies Press, Washington, D.C.

Dunn, L. M. (1968). Special Education for the Mildly Retarded—Is Much of it Justifiable? *Exceptional Children*, 35(1):5–22.

Elder, T. E., Figlio, D. N., Imberman, S. A., and Persico, C. L. (2021). School Segregation and Racial Gaps in Special Education Identification. *Journal of Labor Economics*, 39(S1):S151–S197.

Fien, H., Nelson, N. J., Smolkowski, K., Kosty, D., Pilger, M., Baker, S. K., and Smith, J. L. M. (2021). A Conceptual Replication Study of the Enhanced Core Reading Instruction MTSS-Reading Model. *Exceptional Children*, 87(3):265–288.

Fien, H., Smith, J. L. M., Smolkowski, K., Baker, S. K., Nelson, N. J., and Chaparro, E. (2015). An Examination of the Efficacy of a Multitiered Intervention on Early Reading Outcomes for First Grade Students at Risk for Reading Difficulties. *Journal of Learning Disabilities*, 48(6):602–621.

Fish, R. E. (2017). The racialized construction of exceptionality: Experimental evidence of race/ethnicity effects on teachers' interventions. *Social Science Research*, 62:317–334.

Fish, R. E. (2019). Standing Out and Sorting In: Exploring the Role of Racial Composition in Racial Disparities in Special Education. *American Educational Research Journal*, 56(6):2573–2608.

Fletcher, J. M. and Vaughn, S. (2009). Response to Intervention: Preventing and Remediating Academic Difficulties. *Child Development Perspectives*, 3(1):30–37.

Francis, D. J., Fletcher, J. M., Stuebing, K. K., Lyon, G. R., Shaywitz, B. A., and Shaywitz, S. E. (2005). Psychometric Approaches to the Identification of LD: IQ and Achievement Scores Are Not Sufficient. *Journal of Learning Disabilities*, 38(2):98–108.

Fuchs, D. and Fuchs, L. S. (2006). Introduction to Response to Intervention: What, Why, and How Valid Is It? *Reading Research Quarterly*, 41(1):93–99.

Fuchs, D. and Fuchs, L. S. (2017). Critique of the National Evaluation of Response to Intervention: A Case for Simpler Frameworks. *Exceptional Children*, 83(3):255–268.

Fuchs, D., Fuchs, L. S., and Stecker, P. M. (2010). The "Blurring" of Special Education in a New Continuum of General Education Placements and Services. *Exceptional Children*, 76(3):301–323.

Fuchs, D., Mock, D., Morgan, P. L., and Young, C. L. (2003). Responsiveness-to-Intervention: Definitions, Evidence, and Implications for the Learning Disabilities Construct. *Learning Disabilities Research & Practice*, 18(3):157–171.

Fuchs, L. S., Fuchs, D., and Speece, D. L. (2002). Treatment Validity as a Unifying Construct for Identifying Learning Disabilities. *Learning Disability Quarterly*, 25(1):33–45.

Fuchs, L. S. and Vaughn, S. (2012). Responsiveness-to-Intervention: A Decade Later. *Journal* *of Learning Disabilities*, 45(3):195–203.

Gardner, J. (2022). Two-stage differences in differences. arXiv:2207.05943.

Gersten, R., Compton, D., Connor, C. M., Dimino, J., Santoro, L., Linan-Thompson, S., and Tilly, W. D. (2009). Assisting Students Struggling with Reading: Response to Intervention (RtI) and Multi-Tier Intervention in the Primary Grades. Technical Report NCEE 2009-4045, National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Gersten, R., Haymond, K., Newman-Gonchar, R., Dimino, J., and Jayanthi, M. (2020). Meta-Analysis of the Impact of Reading Interventions for Students in the Primary Grades. *Journal of Research on Educational Effectiveness*, 13(2):401–427.

Gersten, R., Jayanthi, M., and Dimino, J. (2017a). Too Much, Too Soon? Unanswered Questions From National Response to Intervention Evaluation. *Exceptional Children*, 83(3):244–254.

Gersten, R., Newman-Gonchar, R., Haymond, K. S., and Dimino, J. (2017b). What Is the Evidence Base to Support Reading Interventions for Improving Student Outcomes in Grades 1-3? REL 2017-271. Technical report, Regional Educational Laboratory Southeast.

Gilmour, A. F., Fuchs, D., and Wehby, J. H. (2019). Are Students With Disabilities Accessing the Curriculum? A Meta-Analysis of the Reading Achievement Gap Between Students With and Without Disabilities. *Exceptional Children*, 85(3):329–346.

Gilmour, A. F., Harper, J., Lloyd, B., and Van Camp, A. (2023). Response to Intervention and Specific Learning Disability Identification: Evidence From Tennessee. *Journal of Learning Disabilities*, 0(0).

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2):254–277.

Grapin, S. L., Waldron, N., and Joyce-Beaulieu, D. (2019). Longitudinal effects of RtI implementation on reading achievement outcomes. *Psychology in the Schools*, 56(2):242–254.

Hart, C. M. and Lindsay, C. A. (2024). Teacher-Student Race Match and Identification for Discretionary Educational Services. *American Educational Research Journal*.

Hauerwas, L. B., Brown, R., and Scott, A. N. (2013). Specific Learning Disability and Response to Intervention: State-Level Guidance. *Exceptional Children*, 80(1):101–120.

Hibel, J., Farkas, G., and Morgan, P. L. (2010). Who Is Placed into Special Education? *Sociology of Education*, 83(4):312–332.

Kavale, K. A. and Spaulding, L. S. (2008). Is Response to Intervention Good Policy for Specific Learning Disability? *Learning Disabilities Research & Practice*, 23(4):169–179.

Kraft, M. A. (2020). Interpreting Effect Sizes of Education Interventions. *Educational Researcher*, 49(4):241–253.

Ladner, M. (2021). The NAEP Sounded Red Alert for Students with Disabilities Before Covid-19.

Lloyd, J. W., Kauffman, J. M., Landrum, T. J., and Roe, D. L. (1991). Why do teachers refer pupils for special education? An analysis of referral records. *Exceptionality*, 2(3):115–126.

Lockwood, A. B., Farmer, R. L., Winans, S., and Sealander, K. (2022). Specific Learning Disability Identification Practices in the USA: A Survey of Special Education Administrators. *Contemporary School Psychology*, 26(4):535–544.

MacKinnon, J. G., Nielsen, M. , and Webb, M. D. (2023). Cluster-robust inference: A guide to empirical practice. *Journal of Econometrics*, 232(2):272–299.

Maki, K. E. and Adams, S. R. (2019). A current landscape of specific learning disability identification: Training, practices, and implications. *Psychology in the Schools*, 56(1):18–31.

Morgan, P. L., Farkas, G., Cook, M., Strassfeld, N. M., Hillemeier, M. M., Pun, W. H., and Schussler, D. L. (2017). Are Black Children Disproportionately Overrepresented in Special Education? A Best-Evidence Synthesis. *Exceptional Children*, 83(2):181–198.

Morgan, P. L., Farkas, G., Hillemeier, M. M., Mattison, R., Maczuga, S., Li, H., and Cook, M. (2015). Minorities Are Disproportionately Underrepresented in Special Education: Longitudinal Evidence Across Five Disability Conditions. *Educational Researcher*, 44(5):278–292.

Morgan, P. L., Frisco, M. L., Farkas, G., and 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 Education Statistics (2017). Focus on NAEP - Students With Disabilities.

National Center for Education Statistics (2023). Students With Disabilities. Condition of Education. Technical report, U.S. Department of Education, Institute of Education Sciences.

National Research Council (1982). *Placing children in special education: A strategy for equity*. National Academies Press, Washington, D.C.

Novicoff, S. and Dee, T. S. (2023). The Achievement Effects of Scaling Early Literacy Reforms. (EdWorkingPaper: 23-887). Retrieved from Annenberg Institute at Brown University.

O'Connor, R. E., Bocian, K. M., Beach, K. D., Sanchez, V., and Flynn, L. J. (2013). Special Education in a 4–Year Response to Intervention (RtI) Environment: Characteristics of Students with Learning Disability and Grade of Identification. *Learning Disabilities Research & Practice*, 28(3):98–112.

ORTIi (2023). Project Overview. https://www.oregonrti.org/project-overview.

O'Connor, R. E., Bocian, K. M., Sanchez, V., and Beach, K. D. (2014). Access to a Responsiveness to Intervention Model: Does Beginning Intervention in Kindergarten Matter? *Journal of Learning Disabilities*, 47(4):307–328.

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*.

Pendharkar, E. (2023). What We Know About Multi-Tiered Systems of Supports (MTSS), in Charts. *Education Week*.

Preston, A. I., Wood, C. L., and Stecker, P. M. (2016). Response to Intervention: Where It Came From and Where It's Going. *Preventing School Failure: Alternative Education for Children and Youth*, 60(3):173–182.

Rios-Avila, F., Sant'Anna, P. H. C., and Callaway, B. (2021). CSDID: Stata module for the estimation of Difference-in-Difference models with multiple time periods. *Statistical Software Components*, S458976. Boston College Department of Economics.

Roth, J., Sant'Anna, P. H., Bilinski, A., and Poe, J. (2023). What's trending in difference-in-differences? A synthesis of the recent econometrics literature. *Journal of Econometrics*, 235(2):2218–2244.

Shea, Z. M. and Jenkins, J. M. (2024). Impact of States' Adoption of Response to Interventions (RTI) on the Identification and Placement of Students in Special Education. (EdWorkingPaper: 24-1007). Retrieved from Annenberg Institute at Brown University.

Skiba, R. J., Poloni-Staudinger, L., Simmons, A. B., Renae Feggins-Azziz, L., and Chung, C.-G. (2005). Unproven Links: Can Poverty Explain Ethnic Disproportionality in Special Education? *The Journal of Special Education*, 39(3):130–144.

Smith, J. L. M., Nelson, N. J., Smolkowski, K., Baker, S. K., Fien, H., and Kosty, D. (2016). Examining the Efficacy of a Multitiered Intervention for At-Risk Readers in Grade 1. *The Elementary School Journal*, 116(4):549–573.

Snowling, M. J., Hayiou-Thomas, M. E., Nash, H. M., and Hulme, C. (2020). Dyslexia and Developmental Language Disorder: comorbid disorders with distinct effects on reading comprehension. *Journal of Child Psychology and Psychiatry*, 61(6):672–680.

Snowling, M. J., Nash, H. M., Gooch, D. C., Hayiou-Thomas, M. E., Hulme, C., and Wellcome Language and Reading Project Team (2019). Developmental Outcomes for Children at High Risk of Dyslexia and Children With Developmental Language Disorder. *Child Development*, 90(5):e548–e564.

Stepanek, J. and Peixotto, K. (2009). Models of response to intervention in the Northwest Region states. Technical Report REL 2009–No. 079, U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northwest, Washington, D.C.

Stiefel, L., Fatima, S. S., Cimpian, J. R., and O'Hagan, K. (2023). The Role of School Context in Explaining Racial Disproportionality in Special Education. *Annenberg Institue at Brown University*. (EdWorkingPaper: 22-661).

Stuebing, K. K., Fletcher, J. M., LeDoux, J. M., Lyon, G. R., Shaywitz, S. E., and Shaywitz, B. A. (2002). Validity of IQ-Discrepancy Classifications of Reading Disabilities: A Meta-Analysis. *American Educational Research Journal*, 39(2):469–518.

Sun, L. and Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2):175–199.

Torgesen, J. K. (2009). The Response to Intervention Instructional Model: Some Outcomes From a Large-Scale Implementation in Reading First Schools. *Child Development Perspectives*, 3(1):38–40.

U.S. Department of Education (2017). Significant Disproportionality: Essential Questions and Answers.

VanDerHeyden, A. M., Witt, J. C., and Gilbertson, D. (2007). A multi-year evaluation of the effects of a Response to Intervention (RTI) model on identification of children for special education. *Journal of School Psychology*, 45(2):225–256.

Vaughn, S. and Fuchs, L. S. (2003). Redefining Learning Disabilities as Inadequate Response to Instruction: The Promise and Potential Problems. *Learning Disabilities Research & Practice*, 18(3):137–146.

Wanzek, J., Stevens, E. A., Williams, K. J., Scammacca, N., Vaughn, S., and Sargent, K. (2018). Current Evidence on the Effects of Intensive Early Reading Interventions. *Journal of Learning Disabilities*, 51(6):612–624.

Wanzek, J. and Vaughn, S. (2007). Research-Based Implications From Extensive Early Reading Interventions. *School Psychology Review*, 36(4):541–561.

Wanzek, J. and Vaughn, S. (2011). Is a Three-Tier Reading Intervention Model Associated With Reduced Placement in Special Education? *Remedial and Special Education*, 32(2):167–175.

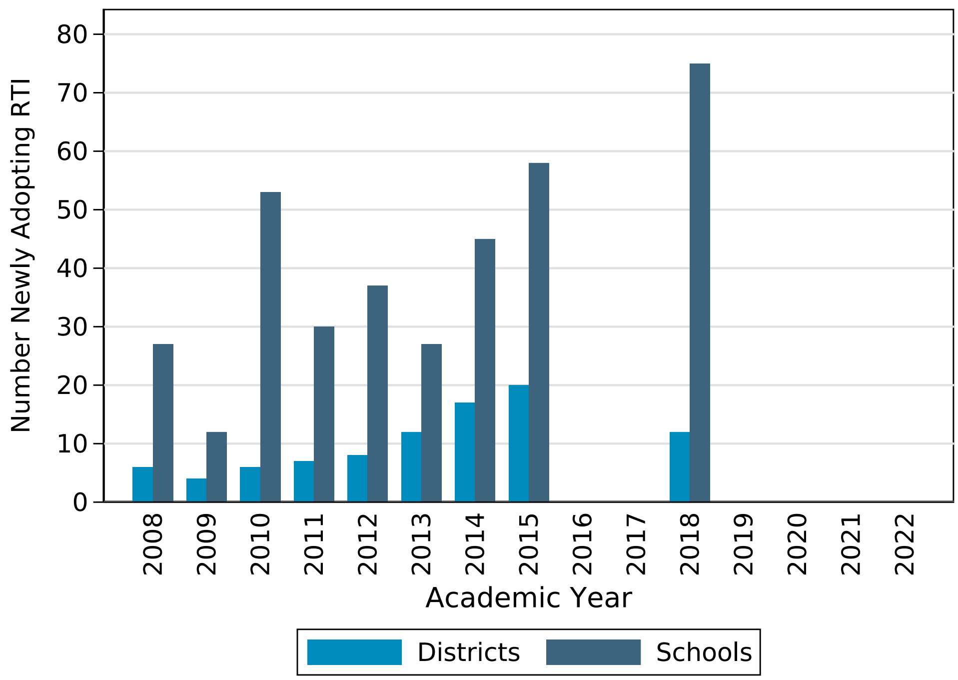
Wanzek, J., Vaughn, S., Scammacca, N., Gatlin, B., Walker, M. A., and Capin, P. (2016). Meta-Analyses of the Effects of Tier 2 Type Reading Interventions in Grades K-3. *Educational Psychology Review*, 28(3):551–576.

Williams, L. M. (2022). *The Specifics of Specific Learning Disability: An Analysis of State-Level Eligibility Criteria and Response to Intervention Practices*. PhD thesis, University of South Florida.

Zirkel, P. A. and Thomas, L. B. (2010). State Laws and Guidelines for Implementing RTI. *TEACHING Exceptional Children*, 43(1):60–73.

Figure 1: RTI Adoption by Academic Year

Panel A. Number of Adopting Districts and Schools



Panel B. Number and Percent of Students in Adopting Districts

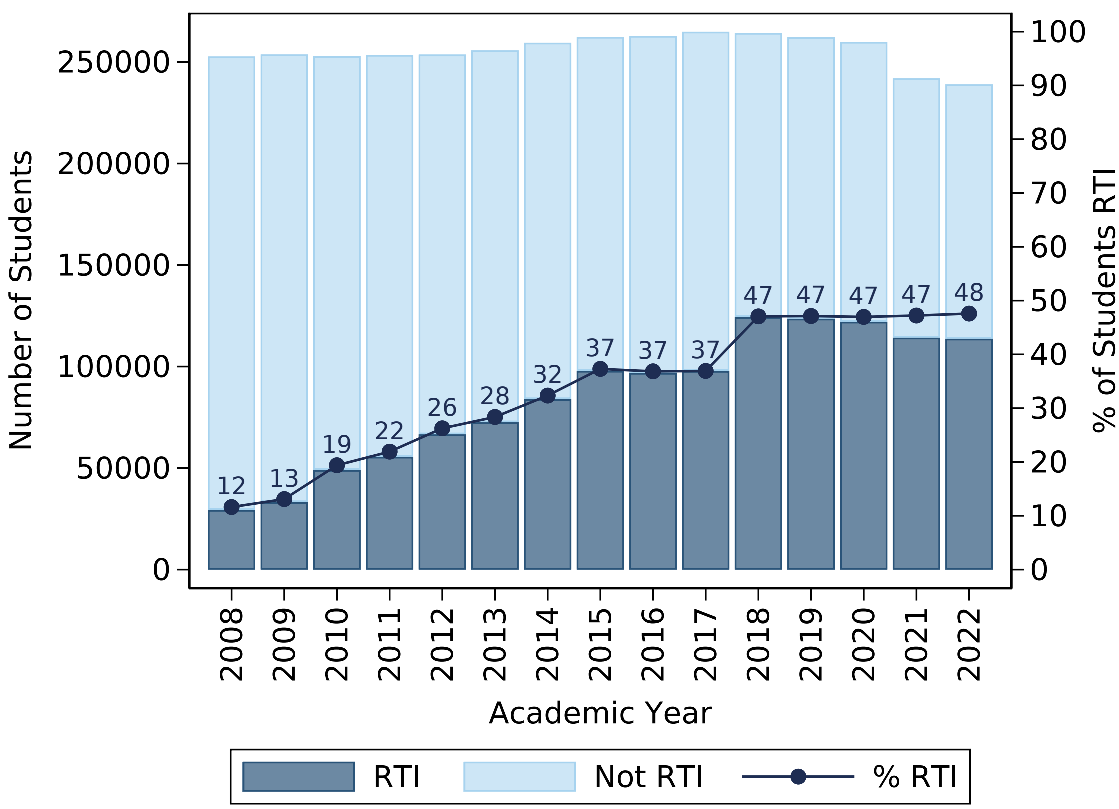
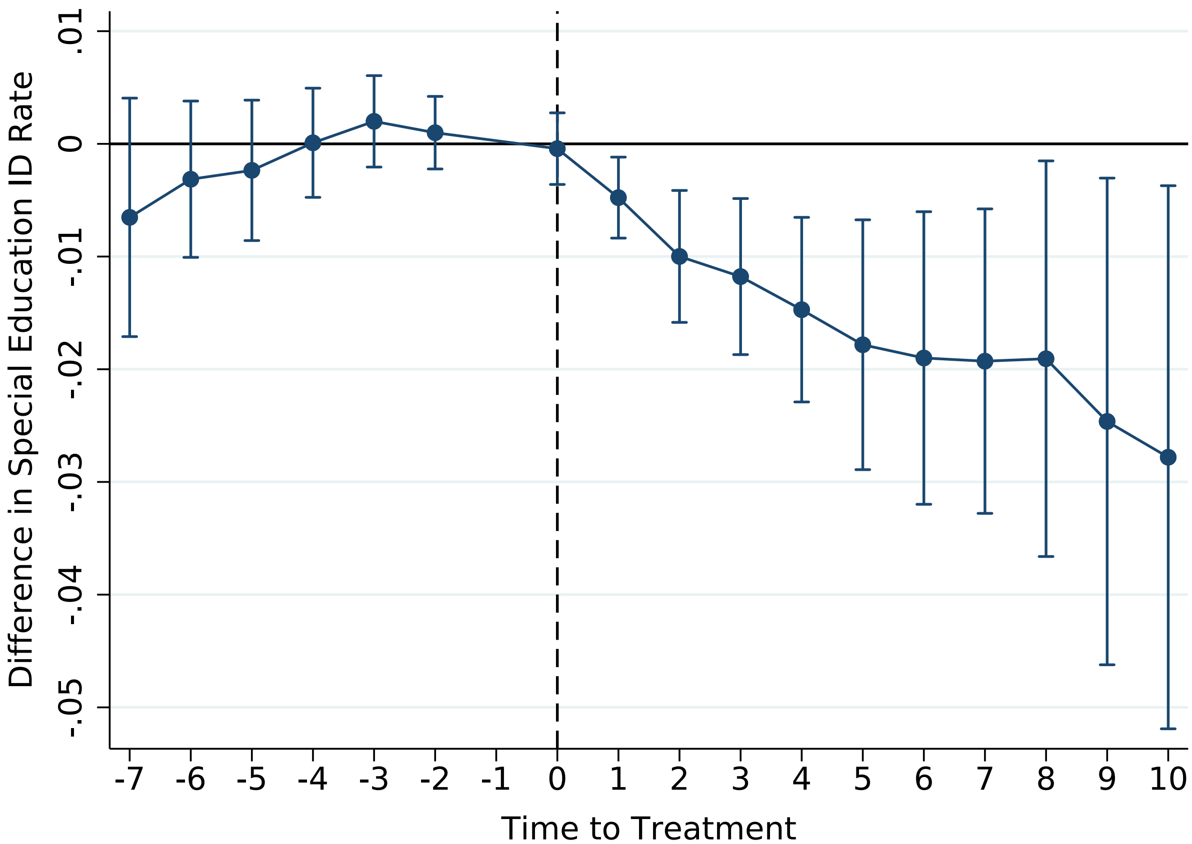


Table 1: Differences Between RTI Adopters and Never Adopters in 2008 Characteristics

	(1) Ever Adopt RTI	(2) Never Adopt RTI	(3) Difference (1)-(2)
Reading Test Score (Std.)	-0.102	0.066	-0.167***
	(0.937)	(1.043)	(0.061)
Math Test Score (Std.)	-0.116	0.094	-0.209***
	(0.927)	(1.040)	(0.066)
Special Education	0.129	0.122	0.007
			(0.006)
Specific Learning Disability	0.034	0.032	0.002
			(0.003)
Speech or Language Impairment	0.058	0.054	0.004
			(0.004)
Other Health Impairment	0.012	0.011	0.001
			(0.002)
Emotional Behavior Disability	0.005	0.005	0.001
			(0.001)
Autism	0.012	0.012	-0.001
			(0.002)
Intellectual Disability	0.005	0.004	0.000
			(0.001)
American Indian/Alaska Native	0.025	0.017	0.009
			(0.007)
Asian/Pacific Islander	0.029	0.066	-0.037**
			(0.015)
Black	0.017	0.044	-0.027
			(0.019)
Hispanic	0.221	0.206	0.015
			(0.044)
White	0.692	0.650	0.042
			(0.053)
Multiracial	0.016	0.018	-0.003
			(0.007)
Female	0.488	0.489	-0.001
			(0.003)
Low Income	0.545	0.464	0.081**
			(0.039)
Gifted	0.039	0.057	-0.018
			(0.012)
Home Language Other Than English	0.220	0.272	-0.052
			(0.055)
Suspended	0.035	0.030	0.005
			(0.004)
Truant	0.009	0.010	-0.001
			(0.004)
City	0.393	0.517	-0.123
			(0.194)
Suburb or Town	0.466	0.422	0.044
			(0.169)
Rural	0.140	0.061	0.079
			(0.057)
Total School Enrollment	407.577	423.148	-15.572
	(165.865)	(158.010)	(34.066)
N Students	89,886	122,746	212,632
N Schools	339	461	800
N Districts	85	102	187

Note. Columns (1) and (2) display the mean for each student-level characteristic with standard deviations reported in parentheses for continuous variables only. Column (3) shows the difference in means between students in districts that ever adopted RTI and those in districts that never adopted RTI estimated by regressing each characteristic on an indicator for ever adopting RTI. Standard error are robustly clustered at the district level and are reported in parentheses. Statistically significant coefficients are indicated as follows: $+p < 0.10$, $*p < 0.05$, $**p < 0.01$, $***p < 0.001$.

Figure 2: Event Study of RTI Impacts on Special Education Identification in K-5



Note. The event study model is estimated using the estimator described in Callaway and Sant'Anna (2021). To implement this estimator, data are collapsed at the school level and weighted by student enrollment. Never treated units are specified as the comparison group. Ninety-five percent confidence intervals are presented for each point estimate, which are calculated using standard errors clustered at the district level.

Table 2: Impact of RTI on Disability Identification in K-5

	(1) SpEd	(2) SLD	(3) SLI	(4) OHI	(5) EBD	(6) ASD	(7) ID
RTI	-0.0144**	-0.0054**	-0.0071*	-0.0013	-0.0005	-0.0009	0.0005
	(0.0051)	(0.0019)	(0.0032)	(0.0010)	(0.0005)	(0.0010)	(0.0004)
Mean	0.137	0.035	0.057	0.015	0.006	0.015	0.005
N Schools	10,926	10,926	10,926	10,926	10,926	10,926	10,926
N Students	3,235,475	3,235,475	3,235,475	3,235,475	3,235,475	3,235,475	3,235,475

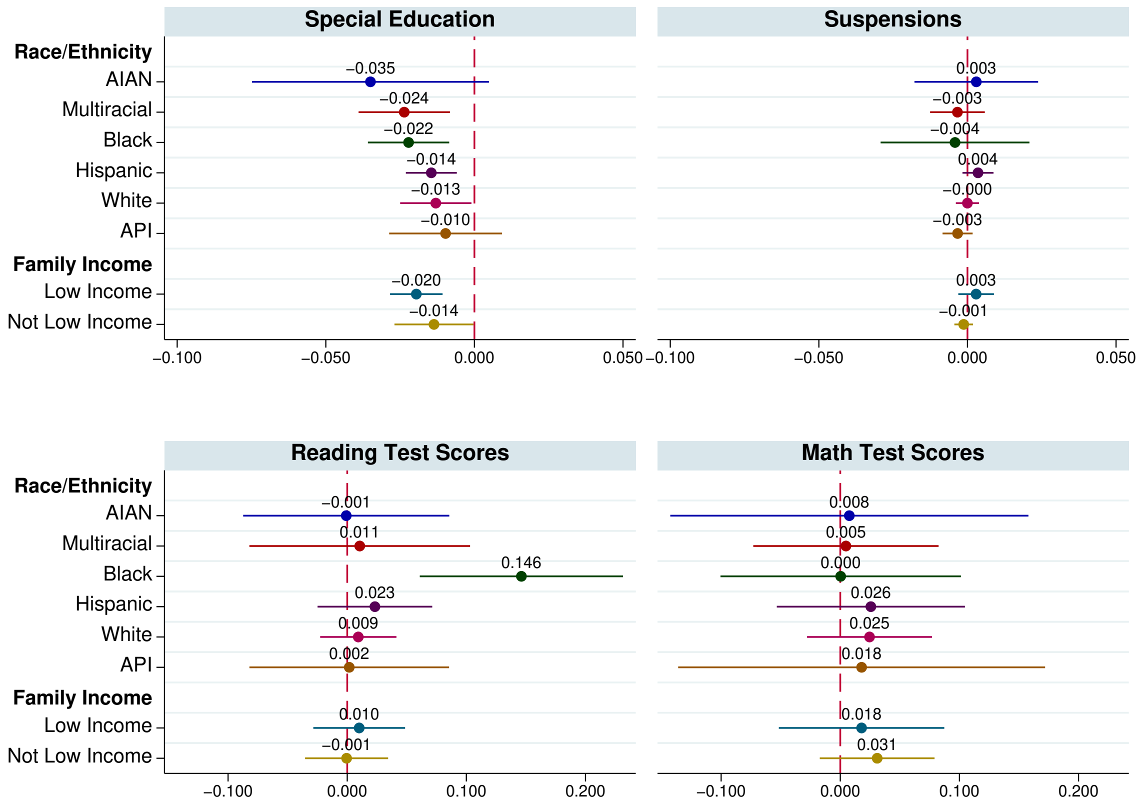
Note. Models are estimated using the estimator described in Callaway and Sant'Anna (2021). To implement this estimator, data are collapsed at the school level and weighted by student enrollment. Never treated units are specified as the comparison group in all models. Standard errors clustered at the district level are presented in parentheses. All outcomes include students in grades K-5. Schools from always treated districts are dropped from all models. Disability categories are abbreviated as follows: "SpEd" is overall special education identification; "SLD" is specific learning disability; "SLI" is speech or language impairment; "OHI" is other health impairment; "EBD" is emotional behavior disability; "ASD" is autism; and "ID" is intellectual disability. Statistically significant coefficients are indicated as follows: $+p < 0.10$, $*p < 0.05$, $**p < 0.01$, $***p < 0.001$.

Table 3: Impact of RTI on Academic and Disciplinary Outcomes

	(1) Reading	(2) Math	(3) Suspensions
RTI	0.0033	0.0172	0.0011
	(0.0164)	(0.0285)	(0.0022)
Mean	-0.127	-0.122	0.033
N Schools	9,037	9,037	10,210
N Students	1,341,768	1,345,571	3,038,304

Note. Models are estimated using the estimator described in Callaway and Sant'Anna (2021). To implement this estimator, data are collapsed at the school level and weighted by student enrollment. Never treated units are specified as the comparison group in all models. Standard errors clustered at the district level are presented in parentheses. Test score outcomes include students in grades 3-5 and are standardized within grade, subject, and year. The suspensions outcome represents the school-level rate of students receiving one or more suspensions in a given school year. This outcome includes students in grades K-5. Schools from always treated districts are dropped from all models. Statistically significant coefficients are indicated as follows: $+p < 0.10$, $*p < 0.05$, $**p < 0.01$, $***p < 0.001$.

Figure 3: Heterogeneous Effects of RTI by Race/Ethnicity and Income



Note. Each coefficient comes from models that are separately estimated for each student subgroup using the estimator described in Callaway and Sant'Anna (2021). To implement this estimator, data are collapsed at the school level and weighted by student enrollment. Never treated units are specified as the comparison group in all models. Ninety-five percent confidence intervals are displayed around each point estimate. The dependent variable is listed above each graph. Student subgroups are abbreviated as follows: American Indian/Alaska Native (AIAN) and Asian/Pacific Islander (API).

A Appendix Tables

Table A1: Counts of Treated Students, Schools and Districts by Time to RTI Adoption

Years to RTI Adoption	N Students	N Schools	N Districts
-10	24,561	79	12
-9	25,012	80	12
-8	25,398	77	12
-7	37,635	135	32
-6	48,296	180	48
-5	54,097	204	59
-4	65,683	241	67
-3	72,565	268	75
-2	87,541	316	81
-1	91,383	334	85
0	91,150	329	85
1	91,661	336	85
2	91,975	333	85
3	90,176	329	84
4	90,282	334	84
5	67,363	254	74
6	66,356	257	74
7	65,622	257	73
8	54,513	200	54
9	44,200	157	37
10	37,828	127	25
11	27,647	90	17
12	19,813	62	10
13	3,861	10	4

Note. Column (1) represents the number of years relative to RTI adoption with year 0 indicating the first school year in which RTI was adopted. The following columns report the number of treated students, schools, and districts included in each year relative to RTI adoption. Data come from school years 2008-2022 and drop always treated districts.

Table A2: Heterogeneous Effects of RTI by Race/Ethnicity

	(1) SpEd	(2) SLD	(3) SLI	(4) Reading	(5) Math	(6) Suspensions
Panel A: American Indian/Alaska Native						
RTI	-0.0350+	-0.0055	-0.0108	-0.0008	0.0076	0.0030
	(0.0203)	(0.0120)	(0.0089)	(0.0441)	(0.0767)	(0.0106)
Mean	0.179	0.042	0.087	-0.396	-0.394	0.039
N Schools	7,774	7,774	7,774	5,191	5,198	7,310
N Students	46,764	46,764	46,764	19,969	20,004	44,590
Panel B: Asian/Pacific Islander						
RTI	-0.0097	-0.0008	-0.0038	0.0017	0.0179	-0.0033
	(0.0097)	(0.0021)	(0.0061)	(0.0428)	(0.0786)	(0.0026)
Mean	0.079	0.014	0.039	-0.130	-0.036	0.018
N Schools	7,970	7,970	7,970	5,985	5,996	7,508
N Students	147,930	147,930	147,930	62,752	63,664	139,869
Panel C: Black						
RTI	-0.0222**	-0.0032	-0.0046	0.1462***	0.0003	-0.0042
	(0.0070)	(0.0051)	(0.0081)	(0.0435)	(0.0515)	(0.0128)
Mean	0.151	0.036	0.054	-0.445	-0.513	0.073
N Schools	7,264	7,264	7,264	4,987	4,979	6,820
N Students	86,187	86,187	86,187	34,503	34,604	80,942
Panel D: Hispanic						
RTI	-0.0145***	-0.0058+	-0.0066	0.0232	0.0257	0.0035
	(0.0044)	(0.0035)	(0.0047)	(0.0246)	(0.0403)	(0.0027)
Mean	0.123	0.037	0.050	-0.514	-0.425	0.031
N Schools	10,183	10,183	10,183	8,252	8,263	9,522
N Students	767,315	767,315	767,315	311,464	313,415	716,671
Panel E: White						
RTI	-0.0130*	-0.0053**	-0.0069	0.0093	0.0245	-0.0000
	(0.0061)	(0.0019)	(0.0039)	(0.0163)	(0.0268)	(0.0020)
Mean	0.143	0.035	0.059	0.041	0.014	0.034
N Schools	10,894	10,894	10,894	8,996	9,000	10,182
N Students	2,017,712	2,017,712	2,017,712	844,874	845,738	1,900,195
Panel F: Multiracial						
RTI	-0.0236**	-0.0097*	-0.0130**	0.0105	0.0047	-0.0034
	(0.0078)	(0.0038)	(0.0041)	(0.0473)	(0.0397)	(0.0047)
Mean	0.148	0.038	0.057	-0.031	-0.082	0.042
N Schools	8,597	8,597	8,597	6,634	6,637	7,988
N Students	161,134	161,134	161,134	63,146	63,106	148,690

Note. Models are estimated using the estimator described in Callaway and Sant'Anna (2021). To implement this estimator, data are collapsed at the school level and weighted by student enrollment. Models are run separately for each student subgroup. Never treated units are specified as the reference group in all models. Statistically significant coefficients are indicated as follows: +$p < 0.10$, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

Table A3: Heterogeneous Effects of RTI by Family Income

	(1) SpEd	(2) SLD	(3) SLI	(4) Reading	(5) Math	(6) Suspensions
Panel A: Not Low Income						
RTI	-0.0136*	-0.0042**	-0.0069	-0.0005	0.0310	-0.0013
	(0.0068)	(0.0014)	(0.0047)	(0.0178)	(0.0246)	(0.0016)
Mean	0.106	0.023	0.049	0.236	0.204	0.019
N Schools	8,756	8,756	8,756	7,370	7,372	8,391
N Students	1,299,264	1,299,264	1,299,264	551,696	552,455	1,241,273
Panel B: Low Income						
RTI	-0.0195***	-0.0083**	-0.0073*	0.0101	0.0179	0.0029
	(0.0045)	(0.0026)	(0.0033)	(0.0197)	(0.0354)	(0.0031)
Mean	0.155	0.043	0.061	-0.356	-0.326	0.042
N Schools	10,538	10,538	10,538	8,691	8,695	9,848
N Students	1,926,604	1,926,604	1,926,604	784,746	787,788	1,787,969

Note. Models are estimated using the estimator described in Callaway and Sant'Anna (2021). To implement this estimator, data are collapsed at the school level and weighted by student enrollment. Models are run separately for each student subgroup. Never treated units are specified as the reference group in all models. Statistically significant coefficients are indicated as follows: $+p < 0.10$, $*p < 0.05$, $**p < 0.01$, $***p < 0.001$.

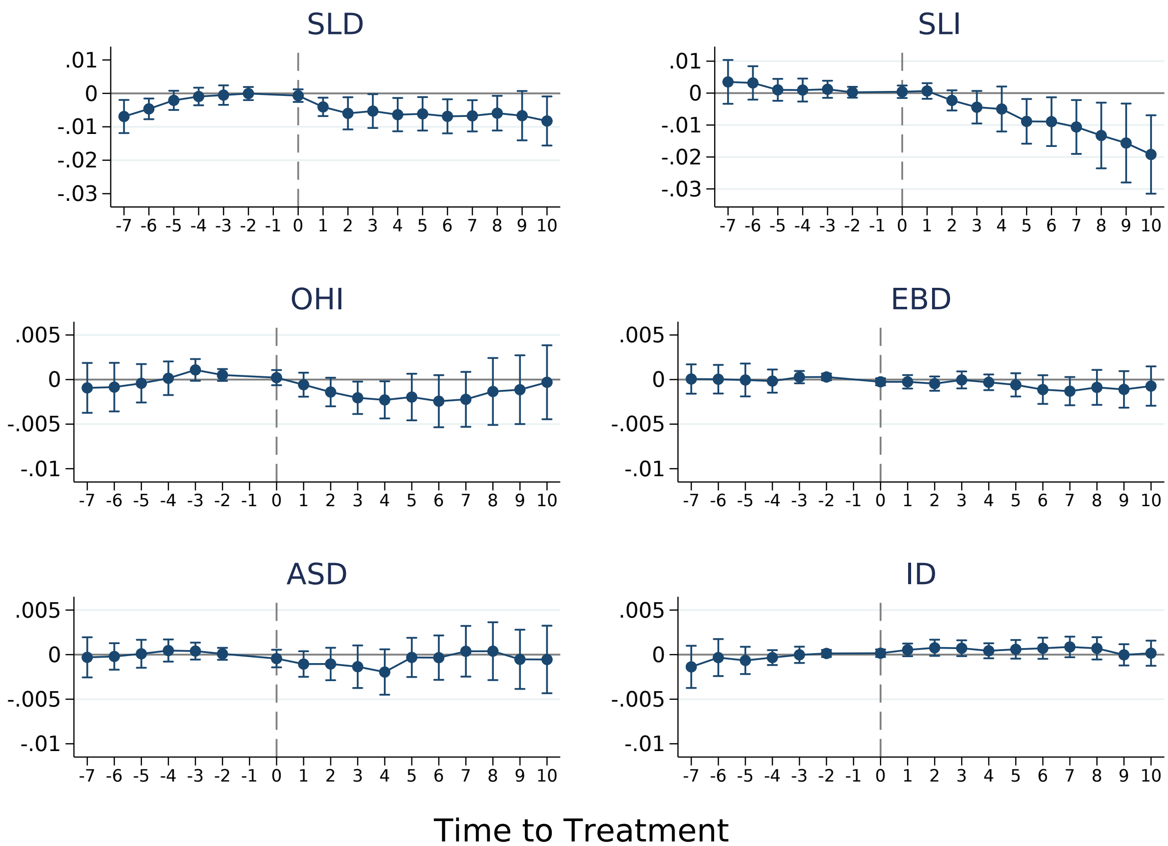
Table A4: Sensitivity Tests for Main Effect Estimates

	(1) CS-School	(2) CS-School	(3) CS-District	(4) CS-District	(5) TWFE	(6) 2s-DiD
Panel A: SpEd ID						
RTI	-0.0144**	-0.0147**	-0.0136**	-0.0141**	-0.0078**	-0.0161**
	(0.0051)	(0.0051)	(0.0051)	(0.0051)	(0.0025)	(0.0055)
N	10,925	11,030	2,786	2,786	11,345	11,345
N Students	3,235,383	3,244,821	3,297,256	3,297,256	3,294,214	3,294,214
Panel B: SLD ID						
RTI	-0.0054**	-0.0058**	-0.0053**	-0.0057**	-0.0038*	-0.0033+
	(0.0019)	(0.0020)	(0.0020)	(0.0021)	(0.0016)	(0.0020)
N	10,925	11,030	2,786	2,786	11,345	11,345
N Students	3,235,383	3,244,821	3,297,256	3,297,256	3,294,214	3,294,214
Panel C: SLI ID						
RTI	-0.0071*	-0.0069*	-0.0072*	-0.0070*	-0.0023	-0.0086*
	(0.0032)	(0.0033)	(0.0033)	(0.0033)	(0.0021)	(0.0038)
N	10,925	11,030	2,786	2,786	11,345	11,345
N Students	3,235,383	3,244,821	3,297,256	3,297,256	3,294,214	3,294,214
Panel D: Reading Test Scores						
RTI	0.0033	0.0000	-0.0043	-0.0076	0.0054	0.0129
	(0.0164)	(0.0157)	(0.0191)	(0.0182)	(0.0228)	(0.0224)
N	9,037	9,125	2,345	2,345	9,305	9,305
N Students	1,341,768	1,346,302	1,364,940	1,364,940	1,362,906	1,362,906
Panel E: Math Test Scores						
RTI	0.0172	0.0154	0.0092	0.0072	0.0033	0.0331
	(0.0285)	(0.0277)	(0.0306)	(0.0297)	(0.0266)	(0.0305)
N	9,037	9,123	2,347	2,347	9,305	9,305
N Students	1,345,571	1,350,122	1,368,710	1,368,710	1,366,670	1,366,670
Panel F: Suspensions						
RTI	0.0011	0.0009	0.0006	0.0003	-0.0018	0.0022
	(0.0022)	(0.0022)	(0.0022)	(0.0022)	(0.0038)	(0.0028)
N	10,210	10,315	2,600	2,600	10,566	10,566
N Students	3,038,304	3,047,742	3,090,655	3,090,655	3,085,682	3,085,682
School FE	✓	✓			✓	✓
District FE			✓	✓		
Never Treated	✓		✓		✓	✓
Not Yet Treated		✓		✓	✓	✓

Note. The estimator used in each column is referenced in the column title, with "CS" indicating the Callaway and Sant'Anna (2021) estimator, "TWFE" indicating a two-way fixed effects estimator and "2s-DiD" indicating the two-stage difference-in-differences estimator described in Gardner (2022). The specific unit fixed effect used in each model is indicated with a check mark at the bottom of the table, with "FE" standing for fixed effect. "Never treated" and "not yet treated" signify which units are used as the control group. Statistically significant coefficients are indicated as follows: $+p < 0.10$, $*p < 0.05$, $**p < 0.01$, $***p < 0.001$.

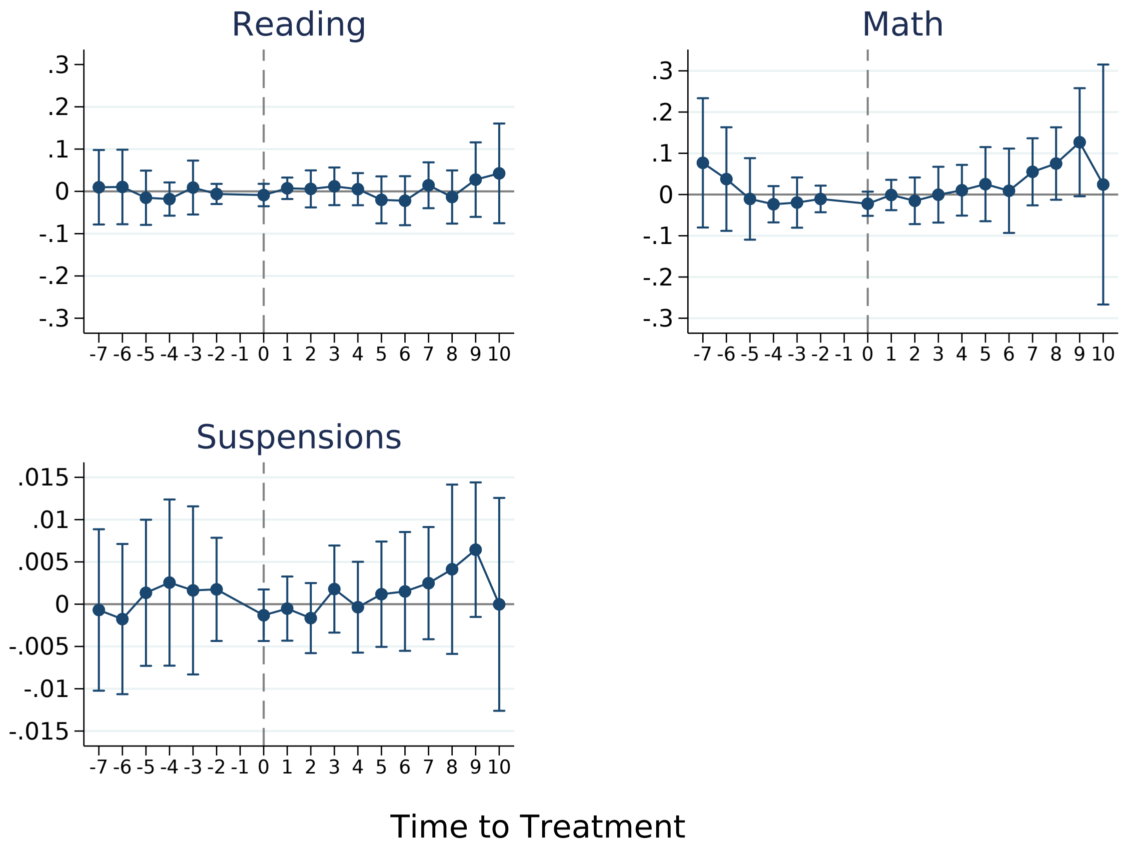
B Appendix Figures

Figure A1: Event Studies of RTI Impacts on Disability Identification



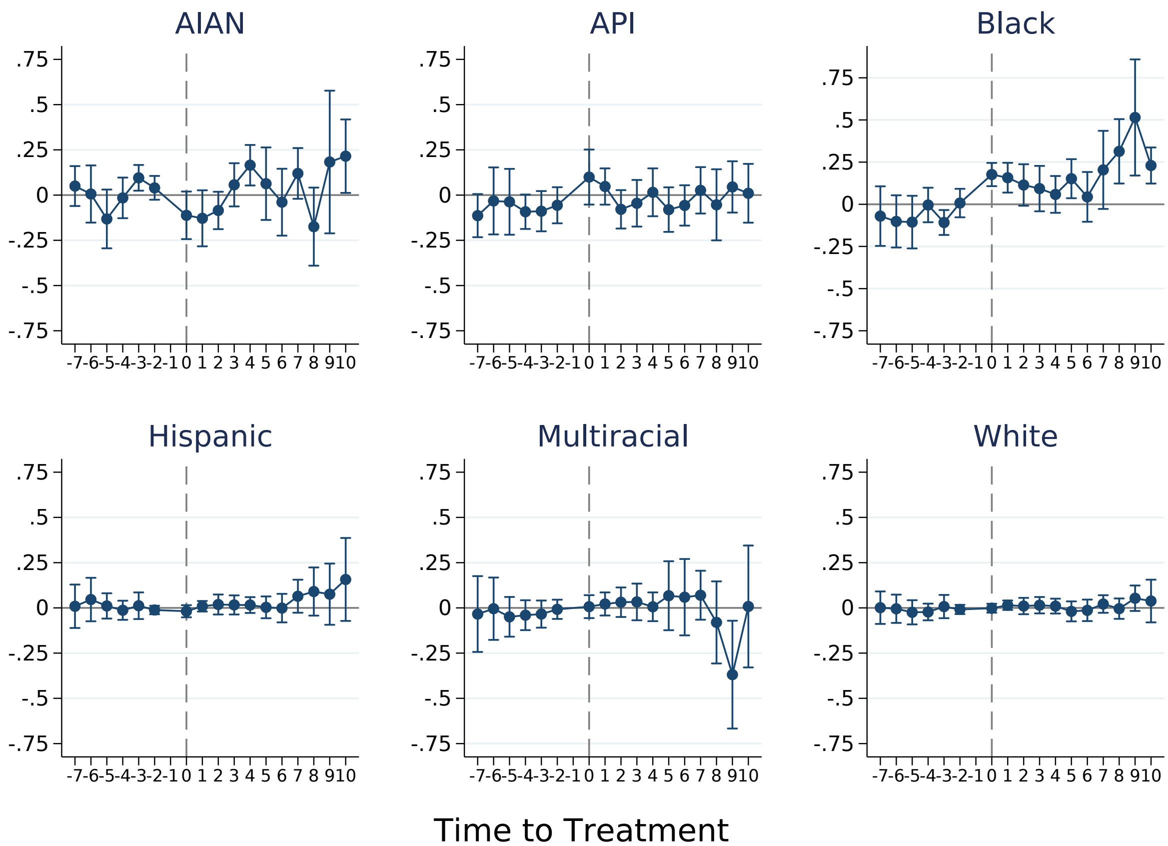
Note. Event study models are estimated using the estimator described in Callaway and Sant'Anna (2021). To implement this estimator, data are collapsed at the school level and weighted by student enrollment. Never treated units are specified as the comparison group in all models. Ninety-five percent confidence intervals are presented for each point estimate, which are calculated using standard errors clustered at the district level. Disability identification outcomes include students in grades K-5 and are reported as weighted school-level identification rates. Disability categories are abbreviated as follows: "SLD" is specific learning disability; "SLI" is speech or language impairment; "OHI" is other health impairment; "EBD" is emotional behavior disability; "ASD" is autism; and "ID" is intellectual disability.

Figure A2: Event Studies of RTI Impacts on Academic and Disciplinary Outcomes



Note. Event study models are estimated using the estimator described in Callaway and Sant'Anna (2021). To implement this estimator, data are collapsed at the school level and weighted by student enrollment. Never treated units are specified as the comparison group in all models. Ninety-five percent confidence intervals are presented for each point estimate, which are calculated using standard errors clustered at the district level. Test score outcomes include students in grades 3-5 and are reported in standard deviation units. Suspensions include students in grades K-5 and are reported as weighted school-level suspension rates.

Figure A3: Event Studies of RTI Impacts on Reading Test Scores by Race/Ethnicity



Note. Event study models are estimated using the estimator described in Callaway and Sant'Anna (2021). To implement this estimator, data are collapsed at the school level and weighted by student enrollment. Never treated units are specified as the comparison group in all models. Ninety-five percent confidence intervals are presented for each point estimate, which are calculated using standard errors clustered at the district level. Reading test score outcomes include students in grades 3-5 and are reported in standard deviation units. Student subgroups are abbreviated as follows: American Indian/Alaska Native (AIAN) and Asian/Pacific Islander (API).