The Use and Efficacy of Capacity-Building Assistance for Low-Performing Districts: The Case of California's District Assistance and Intervention Teams

Katharine O. Strunk
Andrew McEachin
Theresa N. Westover

Abstract

The theory of action upon which high-stakes accountability policies are based calls for systemic reforms in educational systems that will emerge by pairing incentives for improvement with extensive and targeted technical assistance (TA) to build the capacity of low-performing schools and districts. To this end, a little discussed and often overlooked aspect of the No Child Left Behind Act (NCLB) mandated that, in addition to sanctions, states were required to provide TA to build the capacity of struggling schools and Local Education Agencies (LEAs, or districts) to help them improve student achievement. Although every state in the country provides some form of TA to its lowest performing districts, we know little about the content of these programs or about their efficacy in improving student performance. In this paper, we use both quantitative and qualitative analyses to explore the actions taken by TA providers in one state—California—and examine whether the TA and support tied to California's NCLB sanctions succeeds in improving student achievement. Like many other states, California requires that districts labeled as persistently failing under NCLB (in Program Improvement year 3, PI3) work with external experts to help them build the capacity to make reforms that will improve student achievement. California's lowest performing PI3 districts are given substantial amounts of funding and are required to contract with state-approved District Assistance and Intervention Teams (DAITs), whereas the remaining PI3 districts receive less funding and are asked to access less intensive TA from non-DAIT providers. We use a five-year panel difference-in-difference design to estimate the impacts of DAITs on student performance on the math and English language arts (ELA) standardized tests relative to non-DAIT TA during the two years of the program intervention. We find that students in districts with DAITs perform significantly better on math California Standards Tests (CSTs) averaged over both treatment years and in each of the first and second years. We do not find evidence that students in districts with DAITs perform higher on ELA CSTs over the combined two years of treatment, although we find suggestive evidence that ELA performance increases in the second year of treatment relative to students in districts with non-DAIT TA. Ordinary least squares (OLS) regressions that explore the association between specific activities fostered by DAITs and changes in districts' gains in achievement over the two years of treatment show that DAIT districts that report increasing their focus on using data to guide instruction, shifting district culture to generate and maintain high expectations of students and staff, and increasing within-district accountability for student performance, have higher math achievement gains over the course of the DAIT treatment. In addition, DAIT districts that increase their focus on ELA instruction and

Journal of Policy Analysis and Management, Vol. 33, No. 3, 719–751 (2014)

Published by Wiley Periodicals, Inc. View this article online at wileyonlinelibrary.com/journal/pam
DOI:10.1002/pam.21658

shift district culture to one of high expectations have higher ELA achievement gains than do DAIT districts that do not have a similar focus.

INTRODUCTION

High-stakes accountability policies have been one of the longest lasting education reforms in recent history. Individual states have been implementing high-stakes accountability programs for two decades. The federal government instituted its first broad-reaching standards-based education reform in 1994 with the reauthorization of the Elementary and Secondary Education Act (ESEA). Greater incentives for reaching increasing performance targets were added in the reauthorization of ESEA as The No Child Left Behind Act (NCLB) in 2001. The theory of action upon which high-stakes accountability policies are based calls for systemic reforms in educational systems that will emerge by pairing incentives for improvement (in the form of sanctions for failure and rewards for success) with extensive and targeted technical assistance (TA) to build the capacity of low-performing schools and districts (O'Day & Smith, 1993; Smith & O'Day, 1991). Although early discussions of educational accountability policies highlighted capacity-building mechanisms as central to their design (Elmore & Fuhrman, 1994; Smith & O'Day, 1991), concerns about the provisions of supports to build district and school capacity have been largely overshadowed by attention to the attainment of performance goals and sanctions associated with failure.

However, attention to capacity-building supports has not entirely fallen by the wayside. A little-discussed and often-overlooked aspect of NCLB mandated that, in addition to sanctions, states were required to provide TA to build the capacity of struggling schools and Local Education Agencies (LEAs, or districts) to help them improve student achievement. Even greater attention is being paid to the provision of supports and assistance to struggling districts and schools under the newest iteration of federal accountability policy. Specifically, the proposals for ESEA reauthorization and the Department of Education's (DOE) guidance for and granting of ESEA waivers include increased attention to the provision of supports and capacity-building efforts targeted at the lowest performing schools and districts. All State Education Agencies (SEAs) applying for federal waivers to ESEA are required to address "the SEA's process for building SEA, LEA and school capacity to improve student learning in all schools" (United States Department of Education, 2012b, p. 17).

Under NCLB, many states implemented TA policies that aimed to build district-level capacity in their worst performing districts (those in Program Improvement year 3 (PI3) and beyond) (Gottfried et al., 2011; Weinstein, 2011). In California, where this study is located, the lowest performing PI3 districts were given substantial amounts of funding and were required to contract with state-approved external experts, called District Assistance and Intervention Teams (DAITs), to help them build district capacity. The DAITs worked closely with district administrators to assess why the district was failing to increase student achievement and to develop and implement targeted reforms to improve student outcomes. They also conducted comprehensive needs assessments of their districts, made recommendations for improvement that the districts were required to adopt, and provided two years of support as the districts worked to implement the recommendations. The remaining PI3 districts received a lesser amount of funding and were required to access non-DAIT TA, the providers of which were not obligated to provide the same level of support to districts and were not regulated by the California Department of

Education (CDE). Districts in PI2 and below were not given any additional funding for, and were not required to access, any external TA.

Many of the states that have been granted or have requested waivers to be exempted from NCLB are planning to use support structures that are similar to DAITs. For instance, the Colorado Department of Education's ESEA flexibility request assigns all low-performing schools an "Improvement Manager" whose job it is to act as a liaison between the Colorado Department of Education and district leadership teams to help districts identify needs, plan and implement targeted interventions, and monitor progress (United States Department of Education, 2012a). The Florida Department of Education proposed a unified "Differentiated Accountability" system that provides differentiated supports to LEAs through five regional teams that work with low-performing LEAs to build their capacity for reform (United States Department of Education, 2012c).

Although TA programs are provided in every state under NCLB legislation, and will be even more central to education policy in this new era of federal accountability policy, we know surprisingly little about the actions or efficacy of these interventions in building district capacity and improving student performance. In this paper, we examine whether or not the TA and support tied to NCLB accountability sanctions in one state—California—succeeded in improving student achievement. We ask three main questions. First, we ask: *Do students in districts that receive technical assistance (either from DAITs or non-DAIT providers) show greater achievement gains in math or English language arts than students in districts that do not receive technical assistance?* To assess the relationship between the provision of TA and student performance in math and English language arts (ELA) on California Standards Tests (CSTs) in the first two years of program implementation, we use a use a panel difference-in-difference regression design with a five-year panel of data from California's student-level administrative data set (from 2005/2006 to 2009/2010) that tracks approximately 4.9 million students enrolled in about 9,000 schools and 1,000 districts across each of the five years, for approximately 24.4 million student-year observations. We find no evidence of a uniform relationship between the receipt of TA and student achievement in either math or ELA standardized tests. However, we do find evidence that the support provided by DAITs is associated with improved student achievement in math, which brings us to our second question: *Do students in districts with DAITs show greater achievement gains in math and English language arts than students in districts with non-DAIT technical assistance?* This is a particularly policy-relevant question, as states continue to implement TA programs to help districts and schools meet accountability requirements, yet we have found no study to date that assesses the effectiveness of one form of TA relative to another.

To evaluate the efficacy of DAITs on short-run (two year) student achievement we use a five-year panel of data from California's student-level administrative data set (from 2005/2006 to 2009/2010) that tracks approximately 4.9 million student-year observations for students enrolled in the 95 districts designated as PI3, and thus receiving some form of TA. We again use a panel difference-in-difference regression design, this time estimating the impact of DAITs on student performance on the math and ELA CSTs relative to non-DAIT TA. We find that students in districts with DAITs perform significantly better on math CSTs averaged over both treatment years and separately in each of the first and second years. We do not find evidence that students in districts with DAITs perform higher on ELA CSTs over the combined two years of treatment, although we do find some suggestive evidence that ELA performance increases in the second year of treatment relative to students in districts with non-DAIT TA. All of these results are robust to multiple validity checks.

We next use data from our complementary qualitative study of districts with DAITs to explore what actions DAITs encourage and foster that may be associated with the achievement gains found in the first set of analyses. Specifically, our third question asks: *How are DAIT-induced changes in districts' practices and priorities associated with district achievement growth?* We run ordinary least squares regressions that explore the association between changes in district implementation of specific activities fostered by DAITs and changes in the districts' value-added gains in achievement over the two years of treatment. In accordance with previous literature on accountability-induced reforms, we find that DAIT districts that report increasing their focus on using data to guide instruction, shifting district culture to generate and maintain high expectations of students and staff, and increasing within-district accountability for student performance, have higher math achievement gains over the course of the DAIT treatment. In addition, we find that DAIT districts that increase their focus on ELA instruction and shift district culture to one of high expectations have greater ELA achievement gains than do DAIT districts that do not have a similar focus.

The paper proceeds as follows. The next section briefly reviews the limited literature on district-level capacity-building reforms. The following section describes the TA provided to California PI3 districts, focusing in greater detail on the DAIT intervention. The fourth section outlines the data and methods used to assess the relationship between TA and student outcomes, and the effect of DAITs relative to non-DAIT TA on student achievement, and provides results. The fifth section examines threats to the validity of our estimates, followed by description of the relationship between DAIT-induced reforms and districts' achievement growth. The final section discusses policy implications of our findings as well as opportunities for future research.

BACKGROUND AND PREVIOUS RESEARCH

NCLB required state educational agencies to provide TA to all LEAs identified for improvement. This TA was intended to enable low-performing districts to develop and implement their LEA improvement plans and work with schools needing improvement (NCLB, PL 107–110, Title I, Sec. 1116(c)). This element of NCLB was based on the assumption that, to improve student achievement, states need to help build district capacity so that districts may in turn improve schools' capacities for instructional reform.

This focus on building district-level capacity for reform was not particularly new; although, some research previous to the onset of high-stakes accountability policies judged school districts as ineffective and possibly harmful to instructional reforms (Elmore, 1993; Hill, 1995; Rowan & Miskel, 1999), scholars of education reform have long recognized the potential for districts to be instrumental in educational change (Berman & McLaughlin, 1978; Elmore & McLaughlin, 1988; McLaughlin, 1987). In fact, early work on high-stakes accountability reforms considered their success to be especially dependent on the capacities of local school districts to foster and engender instructional reforms that would lead to increased student achievement (Elmore & Fuhrman, 1994; Smith & O'Day, 1991). Studies that have since examined the implementation of high-stakes accountability policies such as NCLB have stressed the need for state governments to build school and district capacities to improve student achievement (Hamilton, Berends, & Stecher, 2005; Opper, Henry, & Mashburn, 2008; Stecher et al., 2008).

However, although much of the literature on instructional coherence and systemic reforms identifies the district as an important actor, it does not focus on how the state can help districts build capacity to foster coherence, and thus improve performance (Gottfried et al., 2011). Rather, the literature has for the most part focused on how districts themselves can provide professional development assistance to schools and teachers to foster instructional change (Cohen & Hill, 2001; Elmore

& Burney, 1998; Firestone et al., 2005; Spillane & Thompson, 1997). This research largely focuses on districts as supporters of instructional reform, designing and delivering professional development for teachers and school administrators, rather than on how districts can work with TA providers to assess and improve student achievement.

With the onset of NCLB and its requirements that states provide TA to districts and schools that are failing to make adequate yearly progress (AYP), this is slowly changing. However, scholars have noted that SEAs may themselves lack the capacity to help districts and schools implement instructional reforms. They lack experience with direct interventions into schools as well as the local organizational networks that can help districts work with schools and other local education actors (Slotnick, 2010; Sunderman & Orfield, 2007). Given this inexperience, many states are turning to independent TA providers or intermediary organizations to help them work with school districts to build district capacity for reform.

Intermediary organizations are a specific type of TA provider. They are independent organizations that work in between two levels of government and provide assistance and essential functions to both parties (Honig, 2004). Rather than simply providing assistance or expertise about one specific area in need of improvement, intermediary organizations provide a broader expanse of knowledge and skills to their work with the lower level of government (i.e., the district), and inform the higher level of government (i.e., the state) about their progress and improvements. There is again little research on the success of intermediary organizations in building district capacity for reform—most of the work on these organizations has focused on the efficacy and operations of intermediary organizations that are hired by LEAs to help schools (Coburn, 2001; Honig, 2004; Honig & Ikemoto, 2008; Marsh et al., 2006; Marsh & Robyn, 2006).

Twelve states including California have made contracting with an intermediary organization (such as a DAIT) a mainstay of their plans to help school districts make instructional reforms (Weinstein, 2011). Although the details of these plans differ by state, the main idea is that the SEAs require or encourage districts in need of improvement to contract with an intermediary organization that will help the district assess their needs, generate improvement plans, and implement improvement strategies. A key element of these state plans is their focus on working with intermediary organizations to build district-level capacity to address problems and issues, not just to assist in solving a specific problem or problems. The capacity-building activities of these intermediaries can include working with district leadership teams to target their specific developmental needs and to develop and focus their improvement plans, providing consulting services to districts based on the needs identified in their improvement plans, mediating relationships between and within district and school administrations, assisting with curricular and instructional alignment, coaching and modeling productive strategies for district administrators and other members of the leadership team, assisting with and monitoring implementation of improvement plans, and providing or modeling specific professional development or coaching strategies. SEAs often fully or partially fund districts' work with these external organizations and the intermediaries are often tasked with providing feedback and recommendations to the state agencies.

Although these intermediary organizations appear to be widely used and are becoming important actors in the provision of TA and capacity-building services for low-performing districts, we have found no peer-reviewed research and few rigorous evaluations of the actions and efficacy of these external service providers. Two studies have examined the supports provided by such intermediary organizations to districts. The first was commissioned by the California County Superintendents' Educational Services Association (CCSESA), and examined the pilot phase of the DAIT intervention discussed in this paper (Padilla et al., 2009). The pilot intervention differed markedly from the final DAIT intervention. Most importantly, it allowed low-performing California districts to volunteer for DAIT assistance, and the intervention itself differed in intensity and structure. The pilot evaluation authors found that there was no significant difference in student achievement in DAIT districts, but caution that the findings should be understood as preliminary and that the lack of consistent achievement results may have been due to too short of a time frame by which to judge results (Padilla et al., 2009). In addition, this study is limited by the small sample of the pilot study, and is unable to use a causal design to assess the impact of the pilot intervention on student outcomes. The second study, by Mass Insight, outlines some difficulties districts face when attempting to work with intermediary organizations. They find that districts working with intermediary organizations have difficulty finding qualified lead providers, and once they find adequate intermediary organizations that might assist them in building capacity, they still need extra assistance in selecting and utilizing these service providers (Mass Insight Education, 2010). Given the fairly recent onset of much of this work (many states began working with and requiring districts to work with these organizations in the last few years), this general lack of research is not surprising. Nonetheless, given their growing prevalence, the lack of attention to the services provided by these intermediary organizations and to their efficacy in improving student performance is alarming. We attempt to address this gap in the research by examining the actions and efficacy of external intermediary organizations in one state—California—relative to the impact of more traditional TA or no assistance at all.

CALIFORNIA'S PROVISION OF TA TO LOW-PERFORMING DISTRICTS UNDER NCLB

By the start of the 2008/2009 school year, California had 248 districts in Program Improvement, and of these 95 (approximately 10 percent of all California school districts) were classified as PI3 or higher—signifying districts that had failed to make AYP under NCLB for at least four years. All PI3 districts received the same overall sanction: Corrective Action Six, which requires them to "institute and fully implement a new curriculum that is based on state academic content and achievement standards" (California State Senate, 2004). In addition, the California Legislature passed Assembly Bill 519 in 2008, which supplied funding to allow districts to access TA based on the severity of the districts' achievement problems. Districts in PI1 and PI2 are not required to access TA, and no funding is provided to help them to do so. Districts in PI3 receive funding to pay for TA (or part of the costs of TA), based on the number of schools in PI3 or higher in the district. Districts receiving these funds must use them to pay for TA, and reports from our qualitative work with districts and providers indicate that the funding is often insufficient to cover the total expense of the TA. The funding for DAITs and non-DAIT TA was paid to the districts in one lump sum at the beginning of the two-year intervention, and districts were instructed to use it all within that time frame to pay for the DAIT or non-DAIT TA.

Within the PI3 category, the CDE has classified districts as in need of "intensive," "moderate," or "light" assistance. These distinctions are made based on an algorithm that takes into account the districts' Academic Performance Index (API) score and relative growth over time, AYP indicators, and the number of PI schools. The lowest performing districts based on the performance algorithm are deemed in intensive need of assistance. These districts are given the most funds per PI3 school and are required to contract with a state-assigned DAIT. The mid-ranking PI3 districts are given slightly less money per PI3 school and are required to contract with their choice of DAIT provider, selected from a list of state-certified organizations. The highest ranking PI3 districts are given even less money per PI3 school and are required to contract with their choice of non-DAIT TA provider. In the first year of implementation of the TA intervention (2008/2009), 43 PI3 districts were classified as in intensive or moderate need of assistance based on their 2007/2008 performance and were required to work with DAIT providers. The remaining 52 PI3 districts received non-DAIT TA.[1]

The intervention is only intended to last for two years, giving DAITs and other assistance providers an aggressive timeline for reform. At the end of each year of the intervention, the CDE and the State Board of Education review the progress made by the PI3 districts that received both DAITs and non-DAIT TA, and make decisions about further sanctions for districts that do not make adequate progress during this time. For the moderate and intensive PI3 districts, the CDE and the SBE are making policy decisions based on short-run impact data and reports from the DAITs regarding the districts' progress.

DAITs

DAITs are state-approved intermediary organizations that work to provide support to help build the capacity of low-performing PI3 districts. Importantly, it is not intended that DAITs provide TA just to solve the districts' specific problems, but rather that they build districts' capacity to assess and solve their own problems in the future. To this end, DAITs provide resources in the form of additional knowledge needed by districts to implement reforms, political and social ties to other organizations and networks, and assistance with shaping the administrative infrastructures of the districts to leverage reforms. In order to be placed on the state-approved list of providers, DAITs must demonstrate expertise in leadership, academic subject areas, meeting the needs of English language learners (ELLs) and students with disabilities (SWDs), and building district capacity. Government agencies, primarily County Offices of Education (COEs), as well as for profit and nonprofit organizations, were approved as DAIT providers and participated in state training events designed to facilitate and inform their work with districts. In the first years of the intervention, 24 districts worked with COEs and 19 worked with private organizations. A survey of DAIT leads indicated that the over half of the DAITs had members with expertise in curriculum; ELL student needs; ELA; math, finance, or budgets; the needs of SWDs; and technology or data systems.[2] In addition, 56 percent of DAIT providers reported that there were problems that required the team to seek additional expertise to supplement the DAIT members.

DAITs are expected to assess district needs in nine "essential program components" as well as eight overarching areas relating to district governance, operations, instruction, and culture. Once the DAITs have assessed district performance in all of

[1] We note that the CDE generated their assignment algorithm (the Priority Assistance Index, PAI) in a way that they believed best-captured districts' need for capacity-building technical assistance, rather than any likely district capacity or readiness for reform or improvement. Although the CDE gave considerable time and attention to generating the algorithm in the way that best used available data to identify the districts most in need of assistance, many districts and county offices of education across the state have shown that, using the same data elements and a slightly different algorithm, the ordering would have been substantially different, resulting in a different set of districts being required to work with a DAIT. This indicates that the differences among intensive, moderate, and light districts are somewhat arbitrary. Importantly, conversations with CDE officials indicate that no districts or technical assistance providers impacted the generation of the PAI.

[2] Information about our qualitative data collection that inform our description of DAITs and their activities, as well as our descriptive analyses later in this paper, can be found in the Appendix section entitled "DAIT Assistance and Support." All appendices are available at the end of this article as it appears in JPAM online. Go to the publisher's Web site and use the search engine to locate the article at http://www3.interscience.wiley.com/cgi-bin/jhome/34787.

these areas, they are then tasked with providing recommendations for improvement. These assessments and recommendations are conveyed in a capacity study, which is submitted to the CDE. Districts are required to implement the recommendations made in these capacity studies. DAITs are then tasked with assisting the district in revising its LEA plan to document steps to implement the DAITs recommendations, and in implementing the revised LEA plan with the goal of accelerating and increasing student achievement (California County Superintendents Educational Services Association, 2008). During both years 1 and 2 of the intervention (2008/2009 and 2009/2010), the DAITs worked with the district leadership teams to continue to understand areas in need of reform and to implement the recommended policy and practice changes.

Results from our qualitative data collection indicate that the DAIT intervention is highly context specific: We find that DAIT activities and implementation varied widely by district, apparently targeting their services and activities to meet the specific needs and contexts of each low-performing district. Some commonalities do appear to exist across districts, however. We find that, overall, DAITs focused to a great extent on building district capacity in governance, professional development for teachers and principals, improving district interactions with and support of school sites to insure consistent instructional practice, and providing additional or improved instructional interventions, particularly in math. We provide greater detail on the activities and priorities of the DAIT providers in the Appendix section entitled “DAIT Assistance and Support.”[3]

Non-DAIT TA

PI3 districts that are classified as in need of light assistance are given less funding per PI3 or higher school and are required to use the funds to work with one or more TA providers of their choosing. The tasks expected of the providers are much less clear than those set forth for DAITs, and districts can choose to hire TA providers to address specific district-identified issues. The state exercises minimal oversight over these providers and over the PI3 districts working with them. Both because of this reduced oversight and because of funding constraints in our study, we know far less about the qualifications and actions of these non-DAIT TA providers than we do about DAIT providers. Although we surveyed all the non-DAIT TA districts, they were not required or particularly encouraged by the CDE to respond, and the survey yielded a low (and likely unrepresentative) response rate of 52 percent. Fifty-two percent of the responding districts reported contracting with a single TA provider, 35 percent reported working with multiple TA providers, and 13 percent reported not contracting with an outside TA provider at all. Responding districts most frequently reported that TA providers supplied the following services “to a great extent”: (1) providing professional development for teachers (55 percent), and (2) training to increase the use of student data to improve instructional practices (48 percent).

THE EFFECT OF TA, AND SPECIFICALLY DAITs, ON STUDENT ACHIEVEMENT

Given the important differences between the kind of work done by intermediary organizations such as DAITs and more traditional TA providers, and the rising prevalence of states’ use of DAIT-like organizations to help build district capacity,

[3] All appendices are available at the end of this article as it appears in JPAM online. Go to the publisher’s Web site and use the search engine to locate the article at http://www3.interscience.wiley.com/cgi-bin/jhome/34787.

the main intent of our analyses is to determine the impact of the DAIT intervention on student achievement outcomes. However, it is first important to understand if there is an overarching effect of TA of any sort—whether supplied by DAITs or by non-DAIT providers—on student outcomes. In this section we first discuss the data used to address both of these questions. We next outline the panel difference-in-difference estimation strategy employed to answer the larger question regarding the relationship between any TA and student outcomes, noting that the results of this analysis may be biased by the noncomparability of the treatment and control groups. We then focus in on the panel difference-in-difference estimation strategy we use to estimate the causal impact of DAITs vs. non-DAIT TA on student achievement.

Data

We start with data on all students in California public schools in grades 2 to 11 from 2005/2006 through 2009/2010 for whom test scores are available in either math or ELA. These data were made available by the CDE specifically for this study. The complete panel data set includes approximately 24.4 million student-year observations, with approximately 4.9 million unique students observed in each of the five years of the panel. Given data entry errors such as duplicate or missing student identifiers and student grade progression patterns, we retain approximately 83 percent of the data in our full five-year sample, for a total of approximately 20.3 million student-year observations. Because we include lagged test score data in our analysis, we further restrict our sample to approximately 13.9 million student-year observations in grades 3 through 11 with lagged test scores in the 2006/2007 to 2009/2010 school years. These data include current and lagged student ELA and math CST scores, as well as student demographic characteristics, and are tied to students' school and district identification codes. As explained in the online Appendix section entitled "Data Management," students are not missing to a substantially greater or lesser extent in districts that receive different kinds of TA (DAIT or non-DAIT), or that are in various stages of Program Improvement. The Appendix section provides a more in-depth description of the specific processes and rationales for dropping observations from the data set.[4]

To answer our first question, we include indicators for districts' placement along the NCLB program improvement spectrum. In these models, we simply include indicators that are set equal to 1 if a district is in that stage of PI (PI1, PI2, PI3-DAIT, or PI3-non-DAIT-TA), using districts that are not in any level of PI as our reference category. In other specifications, we split the PI3-DAIT group into those districts that were labeled in need of intensive vs. moderate DAIT assistance. We attribute the receipt of the DAIT or non-DAIT TA interventions to students who attend a district identified for the intervention in 2008/2009, in the first year of implementation. To remain consistent across all stages of PI, we set each PI indicator equal to 1 in accordance with its PI stage at the start of the 2008/2009 school year. For example, the DAIT variable is set equal to 1 for districts identified for DAIT in 2008/2009 and 2009/2010 and to 0 for such districts in 2005/2006, 2006/2007, and 2007/2008. When we narrow in on the impact of DAIT vs. non-DAIT TA, we reduce the sample to the 95 PI3 districts.

In our analyses of the impact of DAITs for students across the performance distribution, we include indicators of students' performance levels from the 2007/2008

[4] All appendices are available at the end of this article as it appears in JPAM online. Go to the publisher's Web site and use the search engine to locate the article at http://www3.interscience.wiley.com/cgi-bin/jhome/34787.

school year. These performance levels follow California's NCLB designations: Far Below Basic, Below Basic, Basic, Proficient, and Advanced. Throughout the analyses, we also include student-level covariates taken from the CDE's proprietary student-level data set. We also include public-school and district-level data in our models, available from the CDE or from the National Center for Education Statistics Common Core of Data.

Estimation Strategy

We first ask whether or not districts that receive any sort of TA (DAIT or non-DAIT TA) have higher student achievement outcomes than districts without TA. Ideally, we would want to compare student achievement outcomes in districts that received TA to some untreated set of students. One way to do this would be to compare the performance of districts that received the TA intervention in 2008/2009 and 2009/2010 to the same districts' performance before they were assigned the intervention, in the years previous to 2008/2009.

However, it is possible and even likely that some common factor impacted all California districts, or all California districts in PI3, over that period of time. If this is the case, then our simple interrupted time series analysis could attribute some positive or negative trend in student performance over the time period to the intervention, rather than to the secular California-wide trend. Because of this, we would also like to compare students in districts that received TA to students who were likely similar to these students, but were enrolled in districts that did not receive TA. To do this, we utilize a difference-in-difference methodology that compares students in treated (PI3 districts) to students in untreated districts (non-PI, PI1, and PI2 districts) both before the onset of the intervention (before the 2008/2009 school year) and after (during the 2008/2009 and 2009/2010 school years). We use a set of panel difference-in-difference regressions with controls for pertinent student, school, and district characteristics, as well as district and time fixed effects, to isolate the relationship between TA and students' ELA and math CST scores before and after the implementation of the intervention (Angrist & Krueger, 1999; Angrist & Pischke, 2009; Ashenfelter & Card, 1985; Imbens & Wooldridge, 2009; Reback, 2010).

This difference-in-difference approach relies on the comparability of the treated and nontreated groups. However, our nontreated group, which consists of districts in PI1, PI2, or not in program improvement status at all, is likely not particularly comparable to the PI3 districts that received TA. Students in PI3 districts differ significantly in many ways from those in other districts, including their prior achievement, minority status, and ELL status. In addition, PI3 districts received different NCLB policy treatments other than the provision of TA than did PI2 districts and below. Given these factors, comparing students in PI3 districts, with TA, to students in non-PI, PI1, and PI2 districts, may yield biased estimates of a TA treatment effect. The noncomparability of the treatment and control groups diminishes our ability to interpret the relationship between TA and student achievement outcomes as causal.

Nonetheless, to establish whether there may be a relationship between TA treatment and student outcomes, we take advantage of pre- and post-intervention student achievement, controlling for district fixed effects and year fixed effects to find the differences-in-differences estimates:

$$\begin{aligned} Y_{isdt} = {} & \theta Y_{isd,t-1} + \beta_1 PI1_{dt} + \beta_2 PI2_{dt} + \beta_3 \text{non - DAIT TA}_{dt} + \beta_4 \text{DAIT}_{dt} + X_{isdt}\beta_5 \\ & + S_{sdt}\beta_6 + Z_{dt}\beta_7 + \delta_d + \tau_\tau + \varepsilon_{isdt}, \end{aligned} \tag{1}$$

where Y_{isdt} is the standardized ELA or math CST test score for student i in school s in district d in year t,[5] and the indicators $PI1_{dt}$, $PI2_{dt}$, non-DAIT TA_{dt}, and DAIT_{dt} are district d's assignment status to PI1, PI2, non-DAIT TA, or DAIT in year t.

X_{isdt} is a vector of student control variables, including binary indicators for minority and special education students and for ELLs. We also includes a lagged student ELA or math test score, $Y_{isd,t-1}$. We include these measures to help adjust for potential biases that emerge from using nontreated groups that differ on these observable characteristics. S_{sdt} is a vector of school controls, including the natural log of school enrollment, the percent of minorities within the school, and indicators for high and middle schools (elementary schools are the reference category). Z_{dt} is a vector of time-variant district control variables, including measures of the percent of minority students enrolled in districts and the district's per-pupil expenditures. Other potentially important observable district-level variables are time invariant, and as such are accounted for by our district fixed effect. We do not include measures of student poverty, as they are highly correlated with students' race/ethnicity at the individual, school, and district levels (with correlation coefficients of approximately 0.70). In addition, we do not include the exact amount of the funding associated with DAITs and non-DAIT TA because this amount is included in the per-pupil expenditure amount and the funding amount is highly correlated with student enrollment given the way the CDE allocated the funds ($\rho = 0.92$). ε_{isdt} is an idiosyncratic error term. All errors are clustered to the district level.

To answer our second question regarding the impact of DAITs on student outcomes, we again want to compare student achievement outcomes in districts that received a treatment—this time specifically the DAIT intervention—to some untreated set of students. Fortunately, for this analysis we have access to a more comparable control group: the students in PI3 districts that received non-DAIT TA. This is an appropriate comparison group because both sets of districts are designated as PI3 under NCLB, thus facing the same accountability threat and the same sanctions. The only policy difference between these two groups is the one in which we are interested—the level of support and assistance provided them in the form of DAITs vs. non-DAIT TA.

We use a similar set of panel difference-in-difference regressions with controls for pertinent student, school, and district characteristics to isolate the effect that the DAITs had on students' ELA and math CST scores relative to students in districts with non-DAIT TA before and after the implementation of the DAIT intervention. We run essentially the same model as outlined in equation ((1)), this time limiting the sample just to those students in the 95 PI3 districts (all of whom either received the DAIT treatment or the non-DAIT TA control), again controlling for district and year fixed effects to find the difference-in-difference estimates:

$$Y_{ijdt} = \theta Y_{isd,t-1} + \beta_1 \text{DAIT}_{dt} + X_{isdt}\beta_2 + S_{sdt}\beta_3 + Z_{dt}\beta_4 + \delta_d + \tau_\tau + \varepsilon_{isdt}. \tag{2}$$

[5] As in most state achievement test score data sets, California test scores have comparability problems due to the different tests students take as they progress through each grade. In order to make the scores comparable across grades and over time, we standardize all scale scores by subject, grade level, and year. Because California's rules about which students are included in AYP calculations for NCLB and in API calculations for California's own Public Schools Accountability Act (PSAA) are difficult to follow, and we want to be sure that our results are not sensitive to which students are included in reporting under NCLB or PSAA, we standardize our CST scores to 12 different sets of students. These are reported in the Appendix section entitled "Standardization of Outcome Variables," along with Tables A3 and A4 that show results from our main analyses using the outcomes standardized to each of the different populations. Our results are consistent across all standardization structures. All appendices are available at the end of this article as it appears in JPAM online. Go to the publisher's Web site and use the search engine to locate the article at http://www3.interscience.wiley.com/cgi-bin/jhome/34787.

In alternate specifications, we isolate the DAIT treatment effect in each of the first and second years of the intervention. Specifically, we estimate the following:

$$Y_{ijdt} = \theta Y_{isd,t-1} + \beta_1 \text{DAIT Year1}_{dt} + \beta_2 \text{DAIT Year2}_{dt} + X_{isdt}\beta_3 + S_{sdt}\beta_4 + Z_{dt}\beta_5 + \delta_d + \tau_\tau + \varepsilon_{isdt}, \tag{3}$$

where all of the covariates are the same as in model (2), except that we have split the DAIT two-year average effect into its first and second years.[6]

Again, Y_{isdt} is the standardized ELA or math CST test score for student i in school s in district d in year t. DAIT_{dt} is the district d's assignment (treatment) status to DAIT or non-DAIT TA in year t. In this specification, we are identifying our estimates off of students in districts that change status from non-DAIT to DAIT, controlling for secular time trends that may impact student achievement in both DAIT and non-DAIT districts. Because the CDE assigned the DAIT intervention to districts with the lowest aggregate student achievement, districts that were required to contract with DAITs have lower performing students, on average, in both math and ELA than do districts that contracted with non-DAIT TA providers. To account for this in our models, we include the lagged student test score ($Y_{isd,t-1}$) so that we are effectively predicting the increases in student achievement in DAIT vs. non-DAIT TA districts, controlling for previous student performance. These difference-in-difference estimates should provide unbiased estimates of the effect of the DAIT intervention if omitted district-level variables are time invariant. We provide falsification tests below to show that these estimates are not biased due to persistent trends.

RESULTS

Columns 1 and 2 of Tables 1 and 2 show our results from equation ((1)) for both ELA and math CST outcomes. Columns 1 and 2 of Table 1 show that students in PI1, PI2, and non-DAIT TA districts all perform significantly worse on ELA achievement outcomes than do students in non-PI districts, and students in districts with DAITs do not perform significantly differently than students in non-PI districts. Wald tests that assess whether significant differences exist between student ELA achievement in districts that receive any TA (DAIT or non-DAIT) and districts without TA confirm that districts receiving any TA do not perform significantly differently than districts in PI2 status (without TA) ($p = 0.41$), or than districts in PI2 or PI1 status (both without TA) ($p = 0.64$). In addition, Wald tests indicate that there are no significant differences in ELA achievement outcomes between students in districts that received DAITs vs. those in non-DAIT TA districts ($p = 0.33$), or between districts with DAITs and PI2 districts ($p = 0.54$).

Columns 1 and 2 of Table 2 show that students in PI1, PI2, and non-DAIT TA districts do not perform significantly differently on math achievement tests than do students in non-PI districts, but that students in districts with DAITs (both moderate and intensive) perform significantly better than students in non-PI districts. Wald tests that examine whether there are differences in the relationships between any form of TA (DAIT or non-DAIT TA) and math achievement outcomes and no TA (PI2, or PI1 and PI2 combined) and student achievement outcomes have p-values of 0.02 and 0.03, respectively, suggesting that there may be an overall TA effect. However, when we perform additional Wald tests it becomes clear that the overall

[6] We also run analyses that use the change in test score between two years t and $t-1$ as the outcome variable, no longer controlling for the lagged test score on the right-hand side. We find quite similar results in both significance and magnitude. These results are available from the authors upon request.

Table 1. Difference-in-difference estimates of the impact of DAIT and non-DAIT TA on student ELA achievement gains.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Lagged ELA CST	0.785***	0.785***	0.791***	0.791***	0.791***		
	(0.001)	(0.001)	(0.005)	(0.005)	(0.005)		
PI1 district	−0.011**	−0.011**					
	(0.004)	(0.004)					
PI2 district	−0.010+	−0.010+					
	(0.006)	(0.006)					
Non-DAIT TA district	−0.011**	−0.011**					
	(0.004)	(0.004)					
DAIT district (two-year avg. effect)	−0.006		0.007			0.009	
	(0.004)		(0.006)			(0.011)	
DAIT year 1 effect				0.004			0.003
				(0.010)			(0.010)
DAIT year 2 effect				0.010*			0.014
				(0.005)			(0.012)
Moderate DAIT district		−0.007			0.016		
		(0.004)			(0.017)		
Intensive DAIT district		0.002			0.006		
		(0.012)			(0.006)		
Constant	0.054	0.054	0.25	0.232	0.244	0.777**	0.761**
	(0.078)	(0.078)	(0.174)	(0.187)	(0.175)	(0.273)	(0.271)
Student, school, and district controls	X	X	X	X	X	X	X
Adj. R-squared	0.696	0.696	0.673	0.673	0.668	0.151	0.151
No. of students	13,512,375	13,512,375	3,866,090	3,866,090	3,866,090	4,890,786	4,890,796
No. of districts	919	919	95	95	95	95	95

$^{+}p < 0.10$; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

Note: District-clustered standard errors in parentheses. All models include all student, school, and district covariates outlined in the paper, as well as district and year fixed effects.

Table 2. Difference-in-difference estimates of the impact of DAIT and non-DAIT TA on math student achievement gains.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Lagged math CST	0.697***	0.697***	0.693***	0.693***	0.693***		
	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)		
PI1 district	−0.005	−0.005					
	(0.010)	(0.010)					
PI2 district	−0.009	−0.009					
	(0.008)	(0.008)					
Non-DAIT TA district	−0.005	−0.005					
	(0.005)	(0.005)					
DAIT district (two-year avg. effect)	0.025**		0.032***			0.044*	
	(0.009)		(0.009)			(0.017)	
DAIT year 1 effect				0.033*			0.035*
				(0.013)			(0.015)
DAIT year 2 effect				0.031**			0.052**
				(0.010)			(0.020)
Moderate DAIT district		0.023*			0.059***		
		(0.009)			(0.015)		
Intensive DAIT district		0.058***			0.030**		
		(0.014)			(0.010)		
Constant	0.215+	0.214+	0.768**	0.777**	0.751**	1.564***	1.542***
	(0.114)	(0.114)	(0.256)	(0.273)	(0.258)	(0.360)	(0.365)
Student, school, and district controls	X	X	X	X	X	X	X
Adj. R-squared	0.589	0.589	0.545	0.545	0.548	0.128	0.128
No. of students	13,009,706	13,009,706	3,734,900	3,734,900	3,734,900	4,767,783	4,767,783
No. of districts	919	919	95	95	95	95	95

$^{+}p < 0.10$; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

Note: District-clustered standard errors in parentheses. All models include all student, school, and district covariates outlined in the paper, as well as district and year fixed effects.

TA effect is actually driven by a very strong relationship between DAITs and student math achievement outcomes. Specifically, a Wald test between the DAIT and non-DAIT TA effects is significant ($p = 0.001$), whereas Wald tests between non-DAIT TA and PI2 districts, non-DAIT TA and PI1 districts, and non-DAIT TA and PI1 and PI2 districts combined are not at all significant ($p = 0.71$, $p = 0.99$, and $p = 0.51$, respectively).

We further explore the relationship between being enrolled in a DAIT district and achievement gains using the difference-in-difference estimation strategy outlined in equations ((2)) and ((3)). Results are shown in columns 3 through 5 of Tables 1 (ELA) and 2 (math). We again find no evidence of a positive or negative impact of DAITs on the average two-year student achievement on the ELA CSTs. Column (4) shows that there does appear to be a small (1 percent of a standard deviation) increase in students' ELA scores in DAIT districts in the second year of the intervention. However, a Wald test that compares the coefficients between the first- and second-year DAIT effects shows no significant difference between the two years ($p = 0.53$). Therefore, we take the second-year ELA effect as suggestive of a potential longer-term effect and do not consider it to be conclusive.

Table 2, column 3 shows that students in districts with DAITs did see a significant, albeit small, improvement in math CST scores over the two years of treatment, on the order of 3 percent of a standard deviation, or approximately 1.8 points on the 2008/2009 fifth-grade math scale score distribution, relative to students in districts with non-DAIT TA. Column 4 shows that the DAIT effect on students' math CST scores appears consistent across both the first and second years, remaining steady at approximately 3 percent of a standard deviation, and column 5 shows that the impact of DAITs is significant for students in both moderate and intensive DAIT districts.

The inclusion of a lagged CST score on the right-hand side of our models may cause two concerns regarding our estimation strategy. First, the reliance on a lagged score requires that we use a smaller sample of students who have previous CST scores from the year prior. It is possible that the exclusion of these students biases our results. Second, the inclusion of a lagged test score effectively nets out changes in student performance that are attributable to the treatment in the previous year. To insure that we are not biasing our results, we include specifications without the lagged score. Columns 6 and 7 of Tables 1 and 2 present these results. We show that our results remain consistent even with the larger sample without the lagged test score for both ELA and math in the average treatment effect as well as in each year.

In addition, by including the year fixed effect we are assuming a consistent time trend across all districts in our sample. However, it is possible that there are individual time trends occurring in each district, and that a specific district's or set of districts' time trend is driving our main result. We therefore run equations ((2)) and ((3)), this time including district-specific time trends. This makes little difference to the magnitude of our findings, although the DAIT coefficient is no longer always significant at conventional levels ($p = 0.115$ for the two-year average treatment effect, and $p = 0.084$ and $p = 0.087$ for each of the separate year treatment effects, respectively) given the reduced power of the model with the addition of all of the separate district time trend variables. In the interest of space, we do not include these analyses in our tables, but they are available from the authors upon request.

Because our findings show that there is a significant and positive impact of DAITs on student math achievement relative to similar students enrolled in PI3 districts that receive non-DAIT TA, but that there is little evidence of an overall TA effect, the remainder of our paper focuses on the exploring and confirming the DAIT effect. To begin with, given recent focus on the impacts of accountability policies on students along the performance distribution, we model equations ((2)) and ((3)) for each subsample of students along the NCLB performance continuum. This is important for two reasons. First, TA interventions may be more or less effective for students at particular areas of the performance distribution. It will be helpful to understand whether or not the DAIT intervention is more effective for students with specific proficiency levels. Second, some policymakers, practitioners, and researchers have raised the concern that high-stakes accountability policies and their associated interventions incentivize districts and schools to focus on students around the all-important proficiency cut point. Such an accountability-induced focus has been shown to occur in quantitative studies of the Chicago Public Schools (Neal & Schanzenbach, 2010), Washington state (Krieg, 2008), and Texas (Reback, 2008), and qualitatively in a single school district in Texas (Booher-Jennings, 2005). However, other recent research has not found this singular focus on students at the proficiency cut-point (Ballou & Springer, 2009; Ladd & Lauen, 2010). No research that we know of has examined whether the supports associated with accountability policies help districts to improve the achievement of students at specific areas of the performance continuum.[7]

In addition, this analysis provides an additional benefit to our study, serving as an initial validity check. One concern with our analyses is that our results may be biased by *mean reversion*, or regression to the mean. This common phenomenon occurs when students who score at the extreme ends of the distribution (e.g., students scoring Far Below Basic or Advanced) are more likely to score closer to the mean in future test administrations. This might occur because students who score at the lower tail of the distribution have more room to grow, as opposed to students who score at the upper tail of the distribution, who may have little opportunity for improvement. We check for this trend in the California data set and find that students scoring Far Below Basic in one year are more likely to make larger gains the following year than are students who scored higher along the distribution, with Advanced students making negative progress, on average (of approximately −0.3 standard deviation units) between the 2007/2008 and 2008/2009 school years. Therefore, the significant treatment effect found in our initial analyses could be caused by the larger Far Below Basic and Below Basic populations and smaller Advanced and Proficient populations in districts that received DAITs as compared to those that received non-DAIT TA.[8] If this is the case, then equations ((2)) and ((3)) may provide an upwardly biased estimate of the DAIT treatment effect.

Table 3 shows our results from the specification of equations ((2)) and ((3)) for students who are labeled Far Below Basic, Below Basic, Basic, Proficient, and Advanced in ELA or math in the year previous to the implementation of the DAIT intervention (2007/2008). We see that the DAIT impact reported in Table 2 appears at least in part to be driven by gains in the achievement of students at the Basic and Proficient levels. The significant math coefficients from the two-year average treatment effect model for Basic and Proficient students are of the same magnitude as the main treatment effect (0.036, $p < 0.05$), while the insignificant coefficients for Far Below Basic, Below Basic, and Advanced students are 0.007, 0.027, and 0.019, respectively. Although the separate year DAIT impact estimates indicate that Below Basic students in the year previous to the DAIT implementation also significantly improve

[7] We note that there are also additional concerns about the impacts of accountability-associated interventions on closing the achievement gaps between important subgroups of students. In analyses reported in Strunk and McEachin (2012), we show that the DAIT reform appears to diminish the achievement gap between English Language Learners and non-English Language Learners, between white and minority students, and between students who qualify for the federal free- and reduced-price lunch program and those who do not.

[8] The proportions of students classified at each proficiency level in math and ELA in the 2007/2008 school year can be found in Appendix Table A5. All appendices are available at the end of this article as it appears in JPAM online. Go to the publisher's Web site and use the search engine to locate the article at http://www3.interscience.wiley.com/cgi-bin/jhome/34787.

Table 3. Treatment coefficient from difference-in-difference estimates of the impact of DAITs on ELA and math achievement gains for students across the CST proficiency distribution.

		Far Below Basic	Below Basic	Basic	Proficient	Advanced
Math	DAIT two-year ATE	0.007	0.027	0.036*	0.036*	0.019
		(0.017)	(0.019)	(0.017)	(0.017)	(0.016)
	Adj. *R*-squared	0.277	0.212	0.216	0.257	0.321
	N (students)	335,191	899,154	915,788	792,187	470,161
	DAIT year 1 effect	0.005	0.018	0.028+	0.031+	0.012
		(0.022)	(0.021)	(0.016)	(0.017)	(0.019)
	DAIT year 2 effect	0.011	0.037*	0.045*	0.042*	0.025
		(0.013)	(0.019)	(0.020)	(0.019)	(0.018)
	Adj. *R*-squared	0.277	0.212	0.216	0.257	0.321
	N (students)	335,191	899,154	915,788	792,187	470,161
ELA	DAIT two-year ATE	0.008	0.015	0.009	0.008	0.000
		(0.012)	(0.009)	(0.007)	(0.006)	(0.009)
	Adj. *R*-squared	0.45	0.528	0.549	0.543	0.489
	N (students)	352,124	923,421	927,088	798,163	473,028
	DAIT year 1 effect	0.003	0.011	0.005	0.005	−0.007
		(0.013)	(0.012)	(0.009)	(0.009)	(0.012)
	DAIT year 2 effect	0.014	0.019*	0.013+	0.011	0.007
		(0.015)	(0.009)	(0.008)	(0.007)	(0.008)
	Adj. *R*-squared	0.450	0.528	0.549	0.543	0.489
	N (students)	352,124	923,421	927,088	798,163	473,028

$^{+}p < 0.10$; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

Note: District-clustered standard errors in parentheses. All models include all student, school, and district covariates outlined in the paper, as well as district and year fixed effects. Samples selected based on a student's 2007/2008 (pre-intervention) proficiency level.

their test scores in the second year of the intervention, they do not do so in the first year of the intervention. These results are in line with previous research that finds that schools and districts under accountability pressures focus on the students close to the proficiency threshold, or students who are more likely to increase schools' or districts' accountability. However, we cannot discern if DAIT districts are triaging their assistance to improve the performance of students at the proficiency threshold or if the DAIT intervention simply is particularly effective for these students.

We do not see the same patterns with regards to any effect of DAITs on student ELA scores. Specifically, the ELA results in Table 1 that show a small increase in student achievement in the second year of the DAIT intervention appear to be largely driven by students scoring Below Basic and Basic in 2007/2008. While the two-year average effect remains insignificant across the five ELA proficiency levels, we see that the students who scored Below Basic and Basic in 2007/2008 made moderately significant gains in the second year of the intervention compared to the students in the non-DAIT TA districts, with coefficients of 0.019 ($p < 0.05$) and 0.013 ($p < 0.10$), respectively.[9]

[9] We also perform analyses in which the impact of DAITs is allowed to vary by prior student achievement by including an interaction term between the DAIT indicator and prior year student test scores. We find no evidence that the effect of DAITs differs across students with different prior year achievement scores. These results are available from the authors upon request.

Although the math results shown in Table 3 may be important from a policy perspective if they do indeed indicate that the DAIT intervention is only helpful for certain students or that districts with DAITs focused on high-stakes Basic and Proficient students in order to meet NCLB goals, they, along with the ELA results from Table 3, do help to relax our concern that the significant effects found in Table 2 are driven by the regression to the mean phenomenon. In other words, the significant two-year average treatment effect, and the significant separate year effects, do not appear to be driven by a larger proportion of low-performing students and a smaller proportion of high-performing students in the DAIT districts compared to the non-DAIT TA districts. This alleviates one concern regarding the internal validity of our results. We discuss the possibility of other threats to the validity of our findings in the next section.

ASSESSING THE VALIDITY OF THE DAIT EFFECT

In addition to concerns about mean reversion, there are a number of other factors that may cause us to find a positive impact of DAITs on student achievement that should not be attributed to the DAITs themselves. In this section, we test for these sources of bias. One potential threat to the internal validity of our results is the simple misattribution of the effect of other trends to the DAIT treatment effect. As noted above, we estimate equations ((2)) and ((3)) to compare districts that received DAITs in the 2008/2009 and 2009/2010 school years to those that received non-DAIT TA in those years, and we assumed that the impact estimate captures the effect of DAITs on student achievement. However, it is possible that some other force or policy change impacted student achievement in districts that received DAITs relative to those in districts that received non-DAIT TA over the same time period. If this is the case, then our estimates will capture the other impact as well as the impact of the DAITs.

We run our models again, this time examining achievement outcomes in the years before the DAIT and non-DAIT TA treatments were provided. Such a specification check tests to make sure that the cause (i.e., the DAIT intervention) to which we are attributing effects occurs before the resulting outcomes. Table 4 shows our results. This model uses the same data as described above, but rather than having the DAIT indicator equal 1 only in the years of the intervention (2008/2009 and 2009/2010), we now allow the DAIT indicator also to equal 1 in the years preceding the intervention (2006/2007 or 2007/2008). We show separate year DAIT effects (the interactions shown in Table 4), alternately using 2006 to 2007 and 2007 to 2008 as reference years. If we were to find a positive and significant treatment estimate of DAITs on student achievement in the years preceding the intervention, we would be concerned that there is another, non-DAIT, reason for the difference in outcomes between students in districts with DAIT and non-DAIT TA over the time period. However, the results presented in Table 4 show no significant treatment effect in the years previous to the DAIT intervention. This indicates that it is not likely that a force unrelated to the DAIT intervention impacted the achievement levels of students in DAIT vs. non-DAIT TA districts.

Another possible source of bias stems from the possible existence of general accountability pressures associated with being labeled as a lowest performing district and being required to work with a DAIT. Research has shown that accountability systems can increase aggregate student achievement (see, e.g., Carnoy & Loeb, 2002; Dee & Jacob, 2011; Hanushek & Raymond, 2005; Rockoff & Turner, 2010). However, it is unclear if all districts that feel increased accountability pressures improve student achievement. In fact, Carnoy and Loeb's (2002) findings indicate that districts in stronger accountability states before NCLB saw greater increases

Table 4. Misattribution robustness checks for the impact of DAITs on ELA and math achievement gains.

	ELA		Math	
	(1)	(2)	(3)	(4)
Lagged ELA (math CST)	0.791***	0.791***	0.693***	0.693***
	(0.005)	(0.005)	(0.004)	(0.004)
Time effect 2007		0.024***		−0.003
		(0.007)		(0.013)
Time effect 2008	−0.024***	−0.002	0.003	
	(0.007)	(0.006)	(0.013)	
Time effect 2009	−0.039***	−0.016	−0.006	−0.008
	(0.007)	(0.012)	(0.011)	(0.014)
Time effect 2010	−0.025***	−0.002	0.007	0.004
	(0.006)	(0.006)	(0.013)	(0.008)
DAIT × (time effect 2007)		−0.006		0.009
		(0.009)		(0.014)
DAIT × (time effect 2008)	0.006		−0.009	
	(0.009)		(0.014)	
DAIT × (time effect 2009)	0.007	0.001	0.028*	0.038*
	(0.009)	(0.012)	(0.013)	(0.015)
DAIT × (time effect 2010)	0.013+	0.007	0.026+	0.035***
	(0.007)	(0.006)	(0.013)	(0.010)
Constant	0.240	0.218	0.765**	0.765**
	(0.187)	(0.190)	(0.260)	(0.267)
Adj. R-squared	0.668	0.668	0.548	0.548
Number of student/year	3,866,090	3,866,090	3,734,900	3,734,900

$^{+}p < 0.10$; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

Note: District-clustered standard errors in parentheses. All models include all student, school, and district covariates outlined in the paper, as well as district and year fixed effects.

in student achievement. Given the assumed severity of districts' achievement problems that resulted in the requirement to work with a DAIT, these districts may feel stronger accountability pressures than PI3 districts that were considered functional enough to not need DAIT support. We note that the mandate to work with a DAIT is not particularly public—although districts and their school boards know if they are required to access DAIT assistance, it does not appear that the general public is any more aware of a district's label as a PI3-DAIT district than it is aware of a PI3-Non-DAIT-TA district. Nonetheless, if there is extra accountability pressure and it spurs them to enhance student achievement, even in the absence of DAITs, then the observed gains in math CST scores between students in districts that received DAITs vs. non-DAIT TA may be caused by a general accountability threat, and not the DAITs themselves.

We of course cannot test this theory completely. As in other work, it is impossible to separate out the accountability threat from the supports provided that are associated with the accountability threat (Dee & Jacob, 2011). However, the results from equation ((1)), shown in Table 2 column 1, provide an initial test of this hypothesis. In theory, districts in higher levels of PI should be facing increased accountability threat. The results in Table 2 column 1 show that there are no significant differences in student math achievement for students enrolled in districts in different levels of PI (thus facing more or less accountability threat) and students enrolled in non-PI districts, except between DAIT and non-PI districts. Using F-tests that compare the differences between the effects of each level of PI on student math achievement, we find no significant differences between the (null) effect on student math achievement of being enrolled in a non-DAIT TA district vs. a PI2 district ($p = 0.71$), or of being enrolled in a PI2 vs. PI1 district ($p = 0.79$). If the observed treatment effect in Table 1 is solely attributable to some accountability effect, we might expect one or both of these comparisons to show a similar treatment effect. However, we see no evidence of an accountability pressure effect for either of the comparisons listed above.[10]

Another set of robustness checks stems from our original methodological desire to estimate the impact of DAITs on student outcomes relative to non-DAIT TA through the use of a regression discontinuity approach. The setup of the assignment to DAIT vs. non-DAIT TA appears at first glance to be perfect for a regression discontinuity design, because as explained above, the CDE generated an algorithm (the Priority Assistance Index, PAI) that allocated a score on the index to every California district in PI3, and then drew the cut point for DAIT treatment and non-DAIT TA rather arbitrarily, based on the amount of funds the state had available to spend. However, power analyses indicate that, given the district-level sample size of only 95 districts, we would need at least four times the amount of districts in the sample to have sufficient power to perform regression discontinuity analyses. However, we can employ one specification check that is common in the use of regression discontinuity designs—the estimation of threshold models. In these models, the researcher narrows the thresholds around the cut point, or discontinuity, to insure that comparisons between districts farther away from the cut point are not driving the results.

Table 5 shows our results from these threshold models. We run the two-year average effect and separate year effect models from equations ((2)) and ((3)), including only districts that are within consecutively tighter thresholds around the PAI cut point. Originally, we had 43 treated DAIT districts compared to 52 control non-DAIT TA districts. These original full models are replicated in columns 1 and 2. We then consecutively narrow the threshold bands so that columns 3 and 4 show the estimated effect of the DAIT intervention on the districts within 25 percent of the distribution from the cut point. Columns 5 and 6 show the treatment effect when the sample is limited to just the districts 20 percent of the distance from the cut point, columns 7 and 8 show the treatment effect within a 15 percent threshold, and columns 9 and 10 show the treatment effect within a 10 percent threshold. We find that the effect size and significance of both the average two-year and separate year effects remain consistent across all bandwidths. The results of the bandwidth models suggest that it is not the achievement gains of the lowest performing DAIT districts that drive our math treatment effect found in Table 2.

In addition, we are concerned that perhaps the intensive DAIT districts—those that were the lowest performing seven districts according to the PAI, and thus

[10] It is important to again note that we still cannot test if there is a specific DAIT accountability threat that causes districts with DAITs to improve, regardless of the actual intervention. This can be thought of as similar to a Hawthorne Effect in randomized controlled experiments; the DAIT intervention represents the SBE's acknowledgement that these low-performing districts need serious assistance to move out of PI. They are required to provide their revised LEA plans to the CDE and SBE, and in intensive districts, the DAITs provide reports to the state on the districts' progress. In some cases, districts or DAITs are asked to present their plans for improvement to the SBE, and some districts have been visited by SBE members and CDE representatives. Thus, districts with DAITs may feel that they are particularly in the state eye, and our district informants indicate that being required to work with DAITs raises the sense of urgency in making changes to improve student achievement. This may increase accountability pressures in districts with DAITs relative to districts with non-DAIT TA to a greater extent than in other comparisons shown in Tables 1 and 2.

Table 5. Threshold robustness checks for the impact of DAITs on math achievement gains.

	Full		25%		20%		15%		10%	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Lagged math CST	0.693***	0.693***	0.693***	0.693***	0.693***	0.693***	0.692***	0.692***	0.690***	0.690***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)
DAIT avg. two-year effect	0.032***		0.033***		0.030**		0.031**		0.028*	
	(0.009)		(0.010)		(0.010)		(0.011)		(0.012)	
DAIT year 1 effect		0.033*		0.033*		0.032*		0.030*		0.032*
		(0.013)		(0.013)		(0.014)		(0.014)		(0.016)
DAIT year 2 effect		0.031**		0.030**		0.028**		0.029*		0.025+
		(0.010)		(0.010)		(0.010)		(0.011)		(0.013)
Constant	0.768**	0.777**	0.781**	0.793**	0.767**	0.780**	0.829**	0.834**	0.864**	0.882**
	(0.256)	(0.273)	(0.260)	(0.273)	(0.259)	(0.275)	(0.266)	(0.276)	(0.263)	(0.275)
Adj. R-squared	0.545	0.545	0.548	0.548	0.548	0.548	0.548	0.548	0.548	0.548
Number of student/year	3,734,900	3,734,900	3,662,538	3,662,538	3,647,248	3,647,248	3,516,148	3,516,148	3,311,710	3,311,710
Number of districts	95	95	87	87	84	84	78	78	66	66
Number of DAIT districts	43	43	36	36	34	34	30	30	24	24
Number of non-DAIT districts	52	52	51	51	50	50	48	48	42	42

$^{+}p < 0.10$; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

Note: District-clustered standard errors in parentheses. All models include all student, school, and district covariates outlined in the paper, as well as district and year fixed effects. The range of the PAI is 0 to 1 with a cutoff of 0.586. The bandwidths represent the districts within ± a given percent of the PAI. For example, the 25 percent bandwidth includes districts within ±0.25 of the 0.586 cutoff.

Table 6. Auxiliary regressions predicting changes in student characteristics between DAIT and non-DAIT TA districts.

	Minority	FRL	Disability	ELL
DAIT avg. two-year effect	−0.002	0.002	−0.002	0.008
	(0.002)	(0.015)	(0.004)	(0.005)
High school	0.007***	−0.092***	0.004	−0.016
	(0.002)	(0.009)	(0.006)	(0.016)
% minority in school	0.999***	0.969***	0.016***	0.750***
	(0.003)	(0.035)	(0.004)	(0.082)
Middle school	0.005***	−0.033***	0.011	−0.009
	(0.001)	(0.007)	(0.007)	(0.013)
ln(school enrollment)	−0.004***	0.003	−0.015**	0.056***
	(0.001)	(0.005)	(0.005)	(0.009)
% minority in district	−0.208***	0.23	−0.034	−0.294*
	(0.038)	(0.188)	(0.044)	(0.114)
District per-pupil expenditures	0.000	0.000	0.001	−0.001^{+}
	(0.000)	(0.001)	(0.001)	(0.001)
Constant	0.197***	−0.236	0.202***	−0.125
	(0.034)	(0.153)	(0.044)	(0.086)
Adj. R-squared	0.165	0.129	0.001	0.065
Number of student/year	4,961,690	4,959,333	4,961,882	4,963,232
Number of districts	95	95	95	95

$^{+}p < 0.10$; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

Note: District-clustered standard errors in parentheses. All models include all student, school, and district covariates outlined in the paper, as well as district and year fixed effects.

receive the greatest accountability threat and the largest amount of funding, along with a state-assigned rather than a self-selected DAIT provider—may drive the math achievement gains of students in DAIT districts. We return to Table 2 to explore this possibility. In column 5 of Table 2, we estimate equation ((2)) comparing the intensive DAIT districts to the moderate DAIT districts (the remaining 36 treated districts that were required to contract with a self-selected DAIT provider) and to the 52 districts that received non-DAIT TA. Although our threshold analyses showed consistent results with the full sample, the severity model still serves as a check to see if the intensive districts were driving the treatment effect found in our original models. We see that, in accordance with the threshold models, the intensive districts do not appear to be driving our main results.

We run one last set of robustness checks to assess the validity of the DAIT effect: a set of auxiliary regressions that uses nonachievement student traits as the outcome variables in models that are otherwise identical to equation ((2)). We are checking to make sure that the DAIT treatment does not impact characteristics that should not be associated with enrollment in a DAIT district relative to a district with non-DAIT TA. Specifically, we run models that examine the effect of DAITs on whether or not a student is (1) minority, (2) eligible for the federal free- or reduced-price lunch program, (3) classified as "disabled" (or special needs), and (4) classified as an ELL. We run these robustness checks as linear probability models, and they are shown in Table 6. We show that DAIT treatment status has no significant effect on students' classification in any of these groups, and that the point estimate is approximately equal to zero in all four estimations.

Limitations of the DAIT Impact Analyses

The main analyses described above suffer from a number of limitations, many of which we have attempted to address through our specification checks. However, it is important to recognize at the outset that the results from these analyses may still be biased or may not tell the complete story. Although we test for five specific potential threats to the validity of our results, there may still be other sources of bias that remain unchecked. In addition, we are able to examine only two years of outcomes data for the intervention. Although the intervention lasts only two years, more years of data will eventually allow us to better understand the longer-term impacts of DAITs on student achievement. Both of these limitations imply that more research is necessary to understand the true long-run and cohort-specific impacts of DAITs on student achievement.

In addition, the outcome measures we use are not intended for longitudinal assessments of achievement growth. Specifically, the CSTs are not norm-referenced or vertically aligned. Given this fact, it is difficult to compare student achievement on the CSTs over time. We attempt to address this issue by standardizing the outcome variables by year and grade/subject. However, we recognize that this is an imperfect measure of student achievement change over time. Moreover, as mentioned above, we cannot entirely separate out the impact of the accountability threat associated with being required to work with a DAIT provider, and thus labeled as one of the 43 lowest performing districts in California, from the impact of the supports provided to these districts by the DAITs. To begin to explicate how DAIT supports may have helped districts improve, the following section examines the associations between certain DAIT-induced policy changes and reforms and DAIT districts' increases in math and ELA achievement over the course of the intervention.

HOW ARE DAIT-INDUCED CHANGES IN DISTRICTS' PRACTICES AND PRIORITIES ASSOCIATED WITH DISTRICT ACHIEVEMENT GROWTH?

We used DAIT provider and district leader responses to our surveys (described in the Appendix) to generate measures of changes in practice and instructional priorities attributable to DAIT reforms.[11] In our surveys, we asked about DAIT and district activities and priorities in each of the two years of the intervention. We then collected together the items that asked about specific areas of reform in each year and generated measures of focus on each reform area using either principal factor analysis with varimax rotation or, when there were not enough items for this approach, by creating summative index variables. We generated seven measures from our survey data for each of the two years of intervention, each representing areas of reform that districts focused on in each year of the intervention. These areas of focus include math or ELA instruction, managing and collecting data for use in instruction and planning, using data to guide instruction, open communication between the district actors, fiscal responsibility, setting and maintaining high expectations for all students, and holding all personnel within the district accountable for student performance. Because we asked both district leaders and DAITs all survey questions, we used the lowest of the two responses on the Likert scale ratings in order to generate the most conservative response possible to questions pertaining to district focus due to the DAIT intervention. Appendix Table A6 lists these measures and their Cronbach alphas along with the mean values and standard deviations of

[11] All appendices are available at the end of this article as it appears in JPAM online. Go to the publisher's Web site and use the search engine to locate the article at http://www3.interscience.wiley.com/cgi-bin/jhome/34787.

each measure in each of the intervention years, as well as the difference in levels between 2008/2009 and 2009/2010. We use the change in levels between 2008/2009 and 2009/2010 in our analyses of the relationship between changes in district practice induced by the DAIT reform and value-added measures of district achievement across the two years. The value-added measures of district achievement were generated from a model similar to equation ((2)) with two changes: (1) we removed the DAIT treatment indicator; (2) we replaced the separate district and time fixed effects with a district-time fixed effect. We interpret the district-time fixed effects as measures of districts' value added to student achievement in each school year.

We used these data in district-level regression analyses to explore whether or not any particular DAIT-supported changes in district priorities and practices are associated with changes in districts' math and ELA achievement growth during the intervention. Tables 7 and 8 show results from regressions of the 43 DAIT districts' math and ELA change in value added between 2008/2009 and 2009/2010, respectively, on the districts' shifts in focus in each of the priority areas. In all regressions we control for a set of district covariates, as indicated in Tables 6 and 7. We note at the outset that the relationships addressed in these analyses are in no way causal. We simply aim to better understand the relationships between DAIT-induced priority/implementation changes in the 43 districts that worked with DAITs and increases in student achievement, but we cannot make any claims that these shifts in priorities caused increases or decreases in student achievement.[12]

We see in Table 7 that increases in attention to three different sets of priorities are associated with increases in districts' adjusted math achievement growth: a 1 standard deviation increase in a focus on using data to inform instruction (SD = 0.647) is significantly associated with a 0.037 standard deviation increase in adjusted math achievement, a 1 standard deviation increase in districts' focus on high expectations for all students (SD = 0.938) is associated with a 0.033 standard deviation increase in adjusted math achievement, and a 1 standard deviation increase in districts' attention to within-district accountability (SD = 1.390) is associated with a 0.021 standard deviation increase in adjusted math achievement.[13] Interestingly, two of the three factors that appear to be associated with increases in math achievement have more to do with changing the culture of the district than actual practices (although cultural shifts likely culminate in changes in practices). Because these are self-reported data, we do not have information on how the three factors above—using data to inform instruction, setting high expectations for students and staff, and creating a culture of strong within-district accountability—are related to specific behavioral changes within the DAIT districts.

Table 8 shows that shifting focus to foster a culture of high expectations for all students is also associated with increasing ELA achievement in DAIT districts. A 1 standard deviation (SD = 0.938) increase is associated with a 0.021 standard deviation increase in adjusted ELA achievement growth over the course of the intervention. In addition, we see that a shift in focus toward improving instruction in ELA is related to improvements in ELA achievement growth, such that a 1 standard deviation increase in focus on ELA instruction (SD = 0.616) is associated with a district

[12] In addition, we find no evidence that the type of DAIT provider (a private/nonprofit entity or a County Office of Education) is associated with performance gains.

[13] The three significant priority changes in the math achievement analyses are correlated at approximately 0.40. Because we are concerned that we are in fact capturing some similar underlying focus between the three priorities, we run a model that includes all three priorities at the same time. The changing focus on data instruction and high expectations remain significant, whereas the focus on within-district accountability is no longer significant at conventional levels.

Table 7. OLS regression of DAIT districts' adjusted math achievement growth on changes in reform priorities from 2008/2009 to 2009/2010.

	Math instruction (1)	Data management (2)	Data instruction (3)	Open communication (4)	Fiscal responsibility (5)	High expectations (6)	Within district accountability (7)
Change in focus	0.026	0.020	0.057**	−0.001	0.037	0.035**	0.015^{+}
	(0.018)	(0.014)	(0.020)	(0.014)	(0.022)	(0.012)	(0.008)
ln(enrollment)	−0.015	−0.014	−0.015	−0.016	−0.012	−0.016	−0.019^{+}
	(0.011)	(0.013)	(0.010)	(0.012)	(0.011)	(0.011)	(0.011)
% minorities	0.018	0.018	0.018	0.009	0.022	0.088	0.044
	(0.082)	(0.091)	(0.070)	(0.095)	(0.077)	(0.068)	(0.087)
High school	0.041	0.043	0.068	0.049	0.042	0.047	0.068
	(0.053)	(0.047)	(0.042)	(0.053)	(0.048)	(0.041)	(0.046)
Unified	0.016	0.014	0.037	0.022	0.016	0.027	0.031
	(0.030)	(0.034)	(0.028)	(0.032)	(0.029)	(0.031)	(0.029)
Urban	−0.008	−0.008	−0.011	−0.021	−0.013	−0.002	−0.013
	(0.032)	(0.030)	(0.028)	(0.030)	(0.030)	(0.032)	(0.029)
Rural	−0.104^{+}	−0.088^{+}	−0.129**	−0.096^{+}	−0.139*	−0.135***	−0.106^{+}
	(0.051)	(0.045)	(0.041)	(0.052)	(0.065)	(0.034)	(0.055)
Per-pupil expenditures	0.010	0.007	0.007	0.008	0.013	0.009	0.009
(in $1,000s)	(0.011)	(0.010)	(0.009)	(0.011)	(0.011)	(0.010)	(0.011)
Constant	0.044	0.058	0.066	0.074	−0.028	−0.042	0.037
	(0.144)	(0.149)	(0.129)	(0.167)	(0.157)	(0.134)	(0.152)
R-squared	0.026	0.022	0.192	−0.052	0.033	0.109	0.035
N	43	43	43	43	43	43	43

$^{+}p < 0.10$; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

Table 8. OLS regression of DAIT districts' adjusted ELA achievement growth on changes in reform priorities from 2008/2009 to 2009/2010.

	ELA instruction (1)	Data management (2)	Data instruction (3)	Open communication (4)	Fiscal responsibility (5)	High expectations (6)	Within district accountability (7)
Change in focus	0.036**	0.010	0.015	−0.006	0.005	0.022*	0.001
	(0.011)	(0.008)	(0.012)	(0.008)	(0.022)	(0.009)	(0.006)
ln(enrollment)	0.001	−0.002	−0.003	−0.004	−0.003	−0.003	−0.003
	(0.009)	(0.010)	(0.011)	(0.010)	(0.010)	(0.011)	(0.011)
% minorities	0.032	0.025	0.024	0.011	0.023	0.070	0.024
	(0.050)	(0.068)	(0.061)	(0.070)	(0.067)	(0.060)	(0.071)
High school	0.017	0.029	0.037	0.028	0.031	0.031	0.033
	(0.029)	(0.036)	(0.033)	(0.034)	(0.033)	(0.032)	(0.034)
Unified	−0.004	0.004	0.012	0.011	0.008	0.011	0.009
	(0.016)	(0.022)	(0.021)	(0.022)	(0.021)	(0.020)	(0.021)
Urban	−0.010	0.000	−0.004	−0.009	−0.006	0.005	−0.006
	(0.016)	(0.019)	(0.020)	(0.020)	(0.020)	(0.019)	(0.020)
Rural	−0.078	−0.063	−0.076	−0.063	−0.073+	−0.091+	−0.068
	(0.053)	(0.044)	(0.049)	(0.047)	(0.042)	(0.050)	(0.048)
Per-pupil expenditures	0.010+	0.008	0.008	0.009	0.009	0.009	0.009
(in $1,000s)	(0.005)	(0.006)	(0.006)	(0.007)	(0.007)	(0.006)	(0.007)
Constant	−0.106	−0.066	−0.061	−0.043	−0.073	−0.131	−0.062
	(0.077)	(0.102)	(0.106)	(0.107)	(0.110)	(0.113)	(0.104)
R-squared	0.166	−0.006	−0.010	−0.038	−0.045	0.097	−0.047
N	43	43	43	43	43	43	43

$^{+}p < 0.10$; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

value-added increase in ELA achievement of 0.022 of a standard deviation.[14] This is particularly interesting because our initial interviews and surveys with districts at the beginning of the intervention indicated that a greater proportion of DAITs were helping districts to focus on math instruction than ELA (35 percent of DAITs and 21 percent of District Leadership Teams (DLTs) reported an initial focus on math instruction in the first year of the intervention, whereas only 24 percent of DAITs and 9 percent of DLTs reported a focus on ELA instruction in this year). In the second year of the intervention, however, attention shifted to ELA instruction, with 41 percent of DAITs and 16 percent of DLTs reporting a focus on ELA instruction in year 2, as opposed to 22 percent of DAITs and 13 percent of DLTs describing placing a priority on math instruction.[15]

These descriptive analyses shed some light on our earlier findings. Specifically, our qualitative data indicate that DAITs worked closely with district leaders to develop highly context-specific reform strategies targeted at districts' unique management and instructional needs. Given the individualized nature of the DAIT reform, we may not have expected to see systematic relationships between particular practices and achievement gains. However, our regression analyses indicate that, in fact, certain changes in districts' reform priorities, reportedly brought on by the DAIT intervention, are associated with higher district-average achievement growth in both math and ELA. The ELA findings are also surprising, and suggest that the slight trend toward positive year 2 impacts on ELA outcomes may be capturing changes in practice accomplished by districts through their work with DAITs. Clearly, more work is needed to better understand these relationships.

DISCUSSION AND CONCLUSION

The use of TA providers to assist districts and schools to improve student outcomes was mandated in NCLB, and it appears that any reauthorization of the law will likely retain and enhance provisions requiring TA from states for low-performing districts and schools. Our results indicate that the DAIT model used by California for its lowest performing districts may be more useful than other, less-structured TA of the sort found in the districts that worked with non-DAIT TA providers. DAIT-like interventions using intermediary organizations to build district capacity for instructional reform are growing in popularity, as state departments of education struggle to help districts build their capacity to improve student achievement under pressure from severe accountability threats. Under NCLB, 12 states including California are using DAIT-like strategies. Moreover, the ESEA waiver applications submitted by 37 states and the District of Columbia show an increasing reliance on DAIT-like supports to build the capacity of low-performing districts. The results from this study may help inform policies in these states as they design and implement district-level interventions required by ESEA and other federal policies. Our analyses are especially relevant in California, where the CDE and SBE are considering whether or not to renew the DAIT intervention.

[14] The two significant priority changes in the ELA achievement analyses are not particularly correlated with each other, at approximately 0.20. However, we run a model that includes both priorities at the same time, and both remain significant and of similar magnitudes.

[15] We also run analyses that regress the 2008/2009 district value add to student achievement on reported DAIT-supported priorities and practices in the first year of the intervention. We find no significant relationships between districts' reported focus on specific priorities or practices and math outcomes, and we see that districts' reported focus on using data to support instruction and fostering high expectations for students are significantly and positively associated with ELA achievement outcomes. We interpret these results with caution as they may be biased by some relationship between 2008/2009 reported levels of foci and districts' unobserved strengths and weaknesses prior to the intervention.

We find that, at least in the immediate two years of the intervention, the provision of any TA is not associated with significant increases in student achievement. However, the requirement to work with a specific type of assistance provider—DAITs—does appear to increase student achievement in math. Moreover, these results are consistent in each of the two years of the intervention. However, we find no significant average two-year gains in ELA outcomes, although we do find suggestive evidence that ELA achievement may be increasing in DAIT districts in the second year of the intervention. These results make sense given that our qualitative findings show that districts with DAITs focused early on increasing math achievement, and switched their focus to ELA achievement only toward the end of the second year of reform.

Moreover, we find suggestive evidence that some particular DAIT-induced activities and shifts in reform priorities are associated with math and ELA achievement gains among districts that work with DAITs. Specifically, it appears that districts that work on changing their cultures to one that stresses high expectations of all students see increases in both math and ELA student achievement, and that districts that work with DAITs to improve all personnel's acceptance of accountability for student performance see increases in math achievement over the course of the intervention. In addition, two changes in practice encouraged by the DAIT providers are associated with increases in achievement. Districts that learn how to better use data to guide instruction see increases in student achievement in math over the two-year intervention, and districts that shift their focus to ELA instruction see increases in student achievement in ELA. Of course, we cannot definitively attribute increases in achievement levels to these changes in culture and practice, but we can at least begin to paint a picture that helps to flesh out our earlier quantitative findings.

Altogether, our qualitative data suggest that the DAIT providers are providing valuable assistance and support to districts, although the support is highly contextualized depending on district needs and capacity. This may speak to one of the strengths of the DAIT intervention: Districts have different needs and there is likely no one-size-fits-all solution. The California DAIT intervention enables the intermediaries to provide and facilitate targeted supports aimed at districts' specific needs.

The results of this study give cause for optimism regarding the efficacy of the DAIT intervention for student outcomes in math, and for changes in district capacity that may impact student achievement outcomes in later years, as district-level reform translates to schools and then to teachers and students. The strategy of melding state-level reforms with highly localized district-level capacity building and support may be worthy of future consideration. Given that the state government, in California's case the State Board of Education, is ultimately responsible for raising student achievement in California and helping schools and districts meet accountability mandates, states around the country are searching for ways to build district and school capacity to make instructional changes. The experience of DAITs, which capitalize on bringing highly context-specific assistance to districts to address their specific local needs may be a model worth building upon in future reform strategies.

However, there are some concerns with the DAIT model that are not readily apparent from the analyses above. First, our results indicate that the significant two-year average effect and the separate year effects in math, as well as the small suggestive second-year effect in ELA, may be driven in part by the performance of students at the basic and proficient levels of achievement. The big-picture significance of the differential treatment effects across the proficiency distribution depends on the specific policy questions of interest. If the goal of the DAIT intervention is to help all students at all levels of proficiency, then the differential effectiveness of the TA is problematic. Students scoring at the bottom and upper tails of the distribution do not see the same degree of improvement, on average, as do students around the center of the distribution. This may be particularly troublesome if the effectiveness of the reform for students at the basic and proficient levels is indicative of a strategy to improve the performance of students around the proficiency cut point in order to help districts exit Program Improvement status, regardless of real learning. However, if the goal of the intervention is to assist some of the students in the lowest performing districts, and those students happen to be those at the basic and proficient levels, then the outcomes are not concerning. Policymakers and practitioners can examine the activities of DAITs to understand why they are most impactful for specific students, and determine how to replicate the supports for these students in other districts and amend them to reach students elsewhere on the performance distribution.

Second, and especially pertinent in light of the current fiscal crisis facing districts and states across the country, this study cannot adequately assess the cost-effectiveness of the DAITs, an intervention that requires a substantial amount of time and money. However, we can perform a back-of-the-envelope calculation that sheds some light on this question. Specifically, California spent $44.25 million on the 43 districts that received the DAIT intervention over the 2008/2009 and 2009/2010 school years, or a little over $1 million per district that received the intervention (amounts per district ranged from $200,000 to $4.8 million). Given that, on average, DAIT districts in the 2008/2009 school year enrolled 12,715 students, the average cost of the DAIT intervention was $81 per pupil over the two years of the intervention. We found that DAITs had a positive effect on student math achievement of approximately 3 percent of a standard deviation. Although this may seem to be a somewhat expensive outlay for a small boost in math achievement, consider this intervention compared to another famous education reform: class size reduction. Estimates of the cost of reducing class sizes from 22 to 15 students in the Tennessee STAR experiment peg the cost at approximately 47 percent of per-pupil expenditures, which in 1998 equaled about $3,501 per pupil (Krueger, 2003). In today's inflation-adjusted dollars, this equates to $4,600 per student to obtain an achievement increase of approximately 0.16 of a standard deviation (Harris, 2009; US Bureau of Labor Statistics, 2012). The cost-effectiveness ratio of the DAIT intervention is approximately a 0.037 standard deviation achievement gain for an expenditure of $100, relative to the 0.004 standard deviation achievement gain for a $100 expenditure on class size reduction policies akin to the Tennessee STAR experiment. Although the DAIT effects appear small, the reform can actually be considered quite cost-effective given the intervention's relatively modest cost.[16]

In addition, it is important to consider how the intervention might fare if it were scaled up to provide assistance to more low-performing districts. As noted earlier in the paper, the reauthorized ESEA is likely to call for supports to be provided to failing districts to help increase student achievement. Given these trends, there will be more districts in need of assistance, and states will need to develop practical plans to meet this growing need. However, the very aspect of the DAIT reform that might have made it successful in the 43 California school districts discussed in this study may be its downfall if California or other states attempt to bring the intervention to greater scale. Specifically, it appears that one of the most helpful elements of the DAIT intervention is that it is a highly context-driven reform. Because of this, it is not clear that the DAIT intervention is particularly scalable. Multiple district respondents noted that there are not sufficient resources within the DAIT provider

[16] We note that this comparison is not entirely fair, given that we know districts spend more on the DAIT intervention than the funds they received from the state. However, even if we assume that districts spend twice the amount provided by the state, the DAIT intervention still appears efficient in comparison to other reforms.

teams to address all of districts' needs, and DAITs themselves report accessing TA providers from outside of their immediate teams to provide necessary assistance to districts. Given that the quantity of districts with DAITs in this early stage of the intervention is quite limited, it is concerning that resources already seem scarce. Although the DAITs clearly provide value to the districts with which they work, it is unclear whether or not there is enough capacity within the set of DAITs to provide assistance to more low-performing districts.

It is also important to note that this work captures only the relatively short-run impacts of DAITs on student achievement. More research is needed to study the longer-term impacts of DAITs on student achievement. It is possible that the effects of DAITs will fade over time as DAITs exit the district, if the DAITs were unable to transfer capacity to district leaders, or if further budget cuts or leadership turnover diminishes any capacity transfer that occurred during the intervention. We only draw conclusions about outcomes of districts while the DAIT providers are actively working with districts, and not about the increased and sustainable capacities of the districts. Future research will enable us to determine if DAITs have truly built districts' long-term capacity, and if the positive math achievement results and the hints of ELA achievement growth continue to exist in the longer term. The long-term sustainability of the reforms and achievement gains resulting from districts' work with their DAITs are particularly at risk given the ongoing budget crisis in California that continues to impact funding for education. Many district respondents to our surveys and in our interviews indicated they are concerned that staff reductions and general budget problems threaten their ability to continue to pursue the reform trajectories initiated by their work with the DAITs. In some cases, respondents expressed fears that all the ground they gained would be lost due to funding shortfalls and loss of key staff. This instability makes it markedly more difficult for districts to become and remain *learning organizations* with structures and organizational memory sufficient to the task of addressing the needs of challenging student populations.

It is also possible that our results are driven more by a latent accountability threat than a true capacity-building intervention. In the short two-year span of the intervention, it was clear to the DAITs that the initiative would be closely watched by the State Board of Education. In fact, the California legislature set aside funding in the state budget for an evaluation of the DAIT reform that was intended to help the SBE determine the efficacy of DAITs in building district capacity and improving student achievement (which funded this independent evaluation of the initiative). However, minimal accountability for DAITs is built into the system once this initial evaluation program is complete. Given this, DAITs may lose their incentive to work with districts to build their capacities for reform once they know that the state funding has run out.

There is much work left to be done on the analysis of the DAIT intervention. First, as stated above, we can draw conclusions about the effect of DAITs on student achievement only for the two years of implementation, and not for following years once the districts are no longer required to work with the DAITs. It is likely that the true impact of the intervention will become apparent over the longer term, as the capacity-building strategies implemented with DAIT assistance begin to bear fruit and as state and district policies and contexts impact and change the DAITs' work. In addition, this study does not follow later cohorts of DAITs and their matched districts. DAITs may have learned from their initial experiences with the first cohort of districts, and may have changed their strategies and activities with later cohorts. It is also possible that DAITs' capacity was reached, and they no longer have capacity to help more districts in the same context-specific ways. Further research is needed to determine if the short-term impact shown in this study remains in later years and later cohorts. In addition, more work is needed to better understand the specific factors that enable DAITs to be more or less successful in building district capacity for instructional reform.

KATHARINE O. STRUNK is Assistant Professor of Education and Policy at Rossier School of Education, University of Southern California, WPH 901E, 3470 Trousdale Parkway, Los Angeles, CA 90089.

ANDREW McEACHIN is IES Postdoctoral Fellow, University of Virginia, Charlottesville, VA 22904-4277.

THERESA N. WESTOVER is Director, Center for Education and Evaluation Services, University of California, Davis, CA 95616.

ACKNOWLEDGMENTS

The authors wish to thank Dr. Don Taylor and Dr. Laura Wagner at the CDE for their valuable assistance throughout this study. In addition, we acknowledge CDE funding for this study via Agreement/Contract CN088622. We owe a great deal to Amy Smith, Mary Stump, Shani Keller, and Stephanie Au for their excellent research assistance. Thanks also to Tatiana Melguizo, Morgan Polickoff, and the referees and editors of JPAM for their careful reading of the manuscript and their helpful suggestions that have greatly improved the study. We also appreciate the constructive comments made by session participants and discussants at the 2011 Annual Conference of the American Educational Research Association and the 2011 Annual Conference of the Association for Education Finance and Policy. All remaining errors are our own.

REFERENCES

Angrist, J. D., & Krueger, A. (1999). Empirical strategies in labor economics. In O. Ashenfelter & D. Card (Eds.), Handbook of labor economics (pp. 1277–1366). New York, NY: New Holland.

Angrist, J. D., & Pischke, J. S. (2009). Mostly harmless econometrics: An empiricist's companion. Princeton, NJ: Princeton University Press.

Ashenfelter, O., & Card, D. (1985). Using the longitudinal structure of earnings to estimate the effect of training programs. The Review of Economics and Statistics, 67, 648–660.

Ballou, D., & Springer, M. G. (2009). Achievement tradeoffs and No Child Left Behind. Unpublished manuscript, Vanderbilt University.

Berman, P., & McLaughlin, M. W. (1978). Federal programs supporting educational change, Vol. VIII: Implementing and sustaining innovations. Santa Monica, CA: The RAND Corporation.

Booher-Jennings, J. (2005). Below the bubble: "Educational triage" and the Texas accountability system. American Educational Research Journal, 42, 231–268.

California County Superintendents Educational Services Association. (2008). Building blocks of integrated academic district support. Unpublished manuscript.

Carnoy, M., & Loeb, S. (2002). Does external accountability affect student outcomes? A cross-state analysis. Educational Evaluation and Policy Analysis, 24, 305–331.

Coburn, C. E. (2001). Collective sensemaking about reading: How teachers mediate reading policy in their professional communities. Educational Evaluation and Policy Analysis, 23, 145–179.

Cohen, D. K., & Hill, H. C. (2001). Learning policy: When state education reform works. New Haven, CT: Yale University Press.

Dee, T. S., & Jacob, B. A. (2011). The impact of no Child Left Behind on student achievement. Journal of Policy Analysis and Management, 30(3), 418–446.

Elmore, R. F. (1993). The development and implementation of large-scale curriculum reforms. Background paper prepared for the American Association for the Advancement of Science, Project 2061, Washington, DC.

Elmore, R. F., & Burney, D. (1998). Continuous improvement in Community District #2, New York City. Pittsburgh, PA: High Performance Learning Communities Project, Learning Research and Development Center, University of Pittsburgh.

Elmore, R. F., & Fuhrman, S. H. (Eds.). (1994). The governance of curriculum. Alexandria, VA: The Association for Supervision and Curriculum Development.

Elmore, R. F., & McLaughlin, M. W. (1988). Stead work: Policy, practice, and the reform of American education. Santa Monica, CA: The RAND Corporation.

Firestone, W. A., Mangin, M., Martinez, M. C., & Polovsky, T. (2005). Leading coherent professional development: A comparison of three districts. Educational Administration Quarterly, 41, 413–448.

Gottfried, M. A., Stecher, B. M., Hoover, M., & Cross, A. B. (2011). Federal and state roles and capacity for improving schools. Santa Monica, CA: The RAND Corporation.

Hamilton, L. S., Berends, M., & Stecher, B. M. (2005). Teachers' responses to standards-based accountability. Santa Monica, CA: The RAND Corporation.

Hanushek, E. A., & Raymond, M. E. (2005). Does school accountability lead to improved student performance? Journal of Policy Analysis and Management, 24, 297–327.

Harris, D. N. (2009). Toward policy-relevant benchmarks for interpreting effect sizes: Combining effects with costs. Educational Evaluation and Policy Analysis, 31, 3–29.

Hill, P. T. (1995). Reinventing public education. Santa Monica, CA: The RAND Corporation.

Honig, M. I. (2004). The new middle management: Intermediary organizations in education policy implementation. Educational Evaluation and Policy Analysis, 26, 65–87.

Honig, M. I., & Ikemoto, G. S. (2008). Adaptive assistance for learning improvement efforts: The case of the Institute for Learning. Peabody Journal of Education, 83, 328–363.

Imbens, G. W., & Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. Journal of Economic Literature, 47, 5–86.

Krieg, J. M. (2008). Are students left behind? The distributional effects of the No Child Left Behind Act. Education Finance and Policy, 3, 250–281.

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

Ladd, H. F., & Lauen, D. L. (2010). Status versus growth: The distributional effects of school accountability policies. Journal of Policy Analysis and Management, 29, 426–450.

Marsh, J. A., & Robyn, A. E. (2006). School and district improvement efforts in response to the No Child Left Behind Act. Santa Monica, CA: The RAND Corporation.

Marsh, J. A., Kerr, K., Ikemoto, G. S., & Darilek, H. (2006). Developing district-intermediary partnerships to promote instructional improvement: Early experiences and lessons about the institute for learning. In K. Wong & S. Rutledge (Eds.), System-wide efforts to improve student achievement (pp. 241–270). Greenwich, CT: Information Age Publishing.

Mass Insight Education. (2010). School improvement grants: Take 2. Boston: Author. Retrieved February 1, 2011, from http://www.massinsight.org/publications/stg-resources/125/.

McLaughlin, M. W. (1987). Learning from experience: Lessons from policy implementation. Educational Evaluation and Policy Analysis, 9, 171–178.

Neal, D., & Schanzenbach, D. W. (2010). Left behind by design: Proficiency counts and test-based accountability. Review of Economics and Statistics, 92, 263–283.

O'Day, J. A., & Smith, M. S. (1993). Systemic reform and educational opportunity. In S. Furman (Ed.), Designing cohort educational policy: Improving the system (pp. 250–312). San Francisco: Jossey-Bass.

Opper, V. D., Henry, G. T., & Mashburn, A. J. (2008). The district effect: Systemic responses to high stakes accountability policies in six southern states. American Journal of Education, 114, 299–332.

Padilla, C., Tiffany-Morales, J., Bland, J., & Anderson, L. (2009). Evaluation of California's district intervention and capacity building initiative: Findings and lessons learned. Sacramento, CA: SRI International.

Reback, R. (2008). Teaching to the rating: School accountability and the distribution of student achievement. Journal of Public Economics, 92, 1394–1415.

Reback, R. (2010). Schools' mental health services and young children's emotions, behavior, and learning. Journal of Policy Analysis and Management, 29, 698–725.

Rockoff, J. E., & Turner, L. (2010). Short run impacts of accountability on school quality. American Economic Journal: Economic Policy, 2, 119–147.

Rowan, B., & Miskel, C. G. (1999). Institutional theory and the study of educational organizations. In J. Murphy & K. Seashore-Louis (Eds.), Handbook of research on educational administration (pp. 359–384). San Francisco: Jossey-Bass.

Slotnick, W. J. (2010). Levers for changes: Pathways for state-to-district assistance in underperforming school districts. Washington, DC: Center for American Progress.

Smith, M. S., & O'Day, J. A. (1991). Systemic school reform. In S. H. Fuhrman & B. Malen (Eds.), The politics of curriculum and testing (pp. 233–267). New York: Falmer Press.

Spillane, J. P., & Thompson, C. L. (1997). Reconstructing conceptions of local capacity: The local education agency's capacity for ambitious instructional reform. Educational Evaluation and Policy Analysis, 19, 185–203.

Stecher, B. M., Epstein, S., Hamilton, L. S., Marsh, J. A., Robyn, A. E., McCombs, J. S., Russell, J.L., & Naftel, S. (2008). Pain and gain: Implementing No Child Left Behind in three states, 2004–2006. Santa Monica, CA: The RAND Corporation.

Strunk, K. O., & McEachin, A. (2012, March). Differential impacts of intensive district-level technical assistance on student achievement: A study of California's District Assistance and Intervention Teams. Paper presented at the Association for Education Finance and Policy annual meeting, Boston, MA.

Sunderman, G. L., & Orfield, G. (2007). Do states have the capacity to meet the NCLB mandate? The Phi Delta Kappan, 89, 137–139.

The No Child Left Behind Act, PL 107–110, Title I, § 1116(c) (2001).

United States Department of Education. (2012a). Colorado ESEA flexibility request. Washington, DC: Author. Retrieved May 10, 2012, from http://www2.ed.gov/policy/eseaflex/approved-requests/co.pdf.

United States Department of Education. (2012b). ESEA flexibility request. Washington, DC: Author. Retrieved May 10, 2012, from http://www.ed.gov/esea/flexibility.

United States Department of Education. (2012c). Florida ESEA flexibility request. Washington, DC: Author. Retrieved May 10, 2012, from http://www2.ed.gov/policy/eseaflex/approved-requests/fl.pdf.

US Bureau of Labor Statistics. (2012). Consumer Price Index. Washington, DC: Author. Retrieved April 17, 2012, from http://stats.bls.gov/cpi/home.htm#data.

Weinstein, T. (2011, March). Interpreting No Child Left Behind corrective action and technical assistance programs: A review of state policy. Paper presented at the American Educational Research Association, New Orleans, LA.

APPENDIX

DAIT Assistance and Support

Data used in the paper to describe DAITs, their activities, and their relationships with district leaders come from a number of sources. We spoke frequently and at length with CDE administrators involved in the generation and operation of the DAIT intervention and analyzed CDE documentation about DAITs. We also collected information from all 43 districts required to work with DAITs in order to understand what capacity-building activities and resources DAITs provided for their paired districts in the two years of the intervention. First, we analyzed the capacity studies that outlined districts' strengths and weaknesses and that DAITs were required to complete at the onset of the intervention. Second, we interviewed district leaders from all districts working with DAITs, and separately interviewed their DAIT providers. In addition, we sent surveys to both of these groups, with a response rate of 100 percent for DAIT providers and 93 percent for district leaders. These interviews and surveys asked questions about DAITs' activities with their districts in each of the two years of the intervention, areas of specific focus for assistance, implementation of various practices and programs, barriers and facilitators encountered, and the outcomes of the intervention process from the respondents' perspectives.

Given that our results show that students in districts with DAITs improve their achievement on math CSTs, we are interested in learning about the assistance DAIT providers gave to their districts that may have helped to increase student performance in math. As discussed in the paper, DAITs were given a tall task, being asked to work closely with district leaders to assess districts' strengths and weaknesses and their capacity for reform and to help affect great changes in a short period of time. Reports from the majority of both district administrators and DAIT providers indicate that DAITs and districts were able to begin working together quickly, and that DAITs jumped in to begin assessing district needs, making recommendations for reforms, and providing support themselves or helping districts to access technical assistance to enact these reforms. To that end, Table A1 shows that initial surveys of both DAITs providers and district leaders suggest that the DAITs were provided with

Table A1. Reported early DAIT activities and access.

	District (% reporting "to a great extent" or "somewhat") $N = 36$	Provider (% reporting "to a great extent" or "somewhat") $N = 43$
The DAIT was provided with access and information necessary for an appropriate understanding of the district	97.5	97.6
The DAIT was able to effectively engage the District Leadership Team (DLT) in all EPCs	90.2	92.8
DAIT effectively diagnosed district needs and priorities	90.4	97.5
DAIT provided support in the revision of the LEA plan	90.4	92.8
DAIT convened and coordinated all external technical assistance providers in the district	64.1	76.7

Source:: District and DAIT implementation surveys.

Table A2. Reported ways DAITs provided support.

	District (% reporting DAIT provided support) $N = 39$	Provider (% reporting DAIT provided support) $N = 42$
General professional development	61.5	69.0
Assessments and use of data	28.2	57.1
Monitored implementation of recommendations	25.6	31.0
Presented to/assisted in prep for presentation to school board	23.1	23.8
Collected data	23.1	35.7
Curriculum/instruction: math (includes PD)	17.9	33.3
Curriculum/instruction: EL/ELD (includes PD)	17.9	35.7
"Hands on" assistance in developing policy/program	15.4	38.1
Curriculum/instruction: ELA (includes PD)	12.8	33.3
Governance (school board, DLT, policies)	12.8	14.3
Teacher collaboration	12.8	21.4
Fiscal	10.3	14.3
Curriculum/instruction SWD (includes PD)	2.6	9.5
HR (policies, practice)	2.6	4.8

Note:: Categories are not mutually exclusive; percentages do not sum to 100 percent. Activities reported in this table are in addition to DAIT: DLT meetings.

Source: District and DAIT follow-up interviews.

access and information necessary for an appropriate understanding of the district and effectively engaged the district leaders in nearly all treated districts, and that they effectively diagnosed district needs and priorities and then provided support in the revision of the LEA plans. In addition, DAITs in the majority of districts were reported to have convened and coordinated all necessary technical assistance providers that would help enact the suggested reforms incorporated into the revised LEA plans.

It is clear from the interview and survey results that there were as many variations as there were districts in the ways that the districts and their providers worked together, the activities and reforms they undertook, the barriers and facilitators they encountered, and their success in making significant changes in process/resources to address student achievement. Table A2 shows the specific activities in which DAITs provided support and assistance throughout the intervention. Generally, we found that most DAITs provided a variety of assistance to their districts—ranging from district staff development and assistance in developing district policies and systems to working with school site administrators and teachers. We see, perhaps unsurprisingly, that DAIT providers believed that they provided more support regarding specific areas than district leaders reported. This may be because district leaders were hesitant to admit support or did not recognize support, or because DAIT providers took credit for activities that district administrators did not believe they deserved, or simply because in interviews the respondents' willingness to provide detailed information (related to the time they were willing to provide to being interviewed) varied—this is particularly the case in instances in which key district informants were less likely than DAIT providers to have been present in the district for the entire duration of the DAIT's work.[17]

[17] We note that in 29 percent of the districts with DAITS, superintendents (comprising 63% of district level respondents) had been in place for less than two years.

Regardless, we see that approximately two-thirds of districts and providers reported that DAITs assisted districts with professional development activities, a category that includes coaching, providing information or resources for administrators and/or board members, training school site personnel on accurately collecting attendance data, training personnel on the use of data to inform instruction, and a variety of other types of support for both district and school staff, including insuring that teachers and principals attended the required training to implement newly adopted curriculum. Other reported activities include assisting administrators with implementing assessments of student progress and using those data, monitoring the district leaders' implementation of DAITs' recommendations, collecting data for district leaders to use in policymaking activities, assisting with implementing effective curriculum and instruction in math and ELA and for ELL students, and assisting districts with developing policies and programs and adjusting governance structures. Fewer DAITs and districts reported that DAITs assisted districts in updating fiscal and human resource policies. While the quantitative analysis shown in Table 2 of the paper suggests that DAIT districts may have targeted their instruction on the students close to the proficiency threshold in the two years after the implementation of the intervention, we do not have qualitative data to support the hypothesis that the districts engaged in "educational triage" (Booher-Jennings, 2005). However, our qualitative instruments did not specifically ask the districts working with DAITs or the DAIT providers if they altered their instruction to focus more on the "bubble students."

Data Management

We begin with approximately 24 million student-year observations from the CDE, consisting of all second- through 11th-grade students in the 2005/2006 through the 2009/2010 school years. Of these students, approximately 4.6 percent are dropped from our data set because they have missing or duplicate identifiers. Based on conversations with officials at the CDE, we do not believe that these students are any more likely to be transient students or the like. It appears that these missing identification numbers are simply due to entry error at the school level or some similar occurrence. We are forced to drop an additional 13 percent of students from the data set because they either (1) only appear in our data set for one year (5.8 percent) or (2) they showed abnormal patterns of grade progression between years (7.2 percent). In the first instance, it is quite possible that we are capturing students who are particularly mobile and leave the state. State identifiers should follow students between districts within the state, although it is possible that this does not occur as intended in some cases, in which case this group will also capture students who start in one district and then move to another. We cannot know the proportion of these 5.8 percent of students who truly leave the state, or who leave the district, or what proportion of these students are simply subject to entry error in a following year. The 7.2 percent of students who show patterns of unusual grade progression (usual grade progression is defined as students who either progress a single grade in each year, skip one grade between years, or are retained in a grade and repeat it in two consecutive years) are again more likely to be mislabeled rather than indicative of any specific mobility issue.

Once we have excluded students based on these reasons, we are left with approximately 20.3 million total student-year observations in our full five-year sample (from the 2005/2006 school year to the 2009/2010 school year). However, because our analyses rely on the use of a lagged test score indicator, we can only include observations from the 2006/2007 through 2009/2010 school years, resulting in approximately 15.95 million student-year observations in California. In addition, we can no longer include second graders in our analyses, as students are tested for the first time in the second grade (so they have no lagged test score), requiring us to drop approximately 1.7 million second-grade student-year observations from this group. Because we are interested in the technical assistance interventions required of districts due to NCLB, we restrict our sample to California districts who receive Title 1 funds, and are thus subject to NCLB requirements. This causes us to lose another 350,000 student-year observations from our analytic sample because they attend a district that does not receive Title 1 funds, and are therefore not subject to NCLB technical assistance requirements. We are left with approximately 13.9 million student-year observations in grades 3 through 11 between 2006/2007 and 2009/2010, which becomes our main analytic sample.

To insure that we are not systematically missing student data from specific districts, we examine missing observations across district types. Specifically of interest to our paper is whether or not districts with DAITs are missing more or fewer students than districts with non-DAIT technical assistance, or than districts that are in PI2, PI1, or non-PI status. We find that there do not appear to be wide discrepancies in the proportion of students who are missing across district PI status types. Specifically, we are missing 17 percent of students from intensive DAIT districts, 19 percent of students from moderate DAIT districts, 18 percent of students from non-DAIT TA districts, and 16 percent, 14 percent, and 13 percent of students from PI2, PI1, and non-PI districts, respectively. Since the comparison of most relevance to our study is between districts that receive either intensive or moderate DAIT assistance and non-DAIT TA, we do not find that these patterns lead to concerns about differences in the proportions of students missing from our main analyses.

Standardization of Outcome Variables

As noted in the text, we standardize our outcome variables to 12 different populations: (1) the entire universe of students with nonmissing or duplicate IDs who have valid test scores reported to the CDE; (2) the group of students with valid test scores who are able to include in our longitudinal data set (have more than one year of consistent and on-track assessment data); (3) the universe of students with nonmissing or duplicate IDs who have valid test scores who are also included in the state's API calculations; (4) the group of longitudinally tracked students who are also included in the state's API calculations; (5) the universe of students with valid test scores who are enrolled in PI districts; (6) the group of longitudinally tracked students who are also enrolled in PI districts; (7) the universe of students with nonmissing or duplicate IDs who have valid test scores who are also included in the state's AYP calculations; (8) the group of longitudinally tracked students who are also included in the state's API calculations; (9) the universe of students with valid test scores who are both in PI districts and included in the state's API calculations; (10) the group of longitudinally tracked students who are both in PI districts and included in the state's API calculations; (11) the universe of students with valid test scores who are both in PI districts and included in the state's AYP calculations; and (12) the group of longitudinally tracked students who are both in PI districts and included in the state's AYP calculations. Tables A3 and A4 show the results for the estimation of equations (2) and (3) using CST scores standardized to these different populations of test takers. They show that our results are consistent regardless of the standardization of the outcome variables, and that in fact we use the most conservative estimates in our main analyses.

Table A3. Difference-in-difference estimates for CST ELA scores standardized across multiple populations.

	Universe		Universe		Universe		Universe		Universe		Universe	
	All	All	PI	PI	API	API	AYP	AYP	API (PI)	API (PI)	AYP (PI)	AYP (PI)
Two-year treatment effect	0.005		0.007		0.008		0.009		0.009		0.012	
	(0.006)		(0.006)		(0.007)		(0.008)		(0.007)		(0.008)	
First-year effect		0.001		0.002		0.003		0.011		0.005		0.013
		(0.010)		(0.010)		(0.010)		(0.010)		(0.011)		(0.011)
Second-year effect		0.010*		0.011*		0.012*		0.007		0.013*		0.01
		(0.005)		(0.005)		(0.005)		(0.007)		(0.006)		(0.007)
Adj. R-squared	0.667	0.667	0.668	0.668	0.669	0.669	0.666	0.666	0.669	0.669	0.666	0.666
N (student/year)	3,866,090	3,866,090	3,866,090	3,866,090	3,551,521	3,551,521	2,471,999	2,471,999	3,550,138	3,550,138	2,471,999	2,471,999
N (districts)	95	95	95	95	95	95	95	95	95	95	95	95
	Longitudinal		Longitudinal		Longitudinal		Longitudinal		Longitudinal		Longitudinal	
	All	All	PI	PI	API	API	AYP	AYP	API (PI)	API (PI)	AYP (PI)	AYP (PI)
Two-year treatment effect	0.007		0.009		0.009		0.009		0.01		0.012	
	(0.006)		(0.006)		(0.007)		(0.008)		(0.007)		(0.008)	
First-year effect		0.003		0.005		0.004		0.011		0.006		0.014
		(0.010)		(0.010)		(0.010)		(0.010)		(0.010)		(0.011)
Second-year effect		0.010*		0.012*		0.013*		0.008		0.015*		0.011
		(0.005)		(0.005)		(0.005)		(0.007)		(0.006)		(0.007)
Adj. R-squared	0.668	0.668	0.668	0.668	0.669	0.669	0.666	0.666	0.67	0.67	0.666	0.666
N (student/year)	3,866,090	3,866,090	3,866,090	3,866,090	3,535,853	3,535,853	2,471,473	2,471,473	3,534,470	3,534,470	2,471,526	2,471,526
N (districts)	95	95	95	95	95	95	95	95	95	95	95	95

$^{+}p < 0.10$; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

Note: District-clustered standard errors in parentheses. All models include all covariates included in Table 1 as well as district and year fixed effects.

Table A4. Difference-in-difference estimates for CST math scores standardized across multiple populations.

	Universe		Universe		Universe		Universe		Universe		Universe	
	All	All	PI	PI	API	API	AYP	AYP	API (PI)	API (PI)	AYP (PI)	AYP (PI)
Two-year treatment effect	0.033*** (0.009)		0.038*** (0.010)		0.035*** (0.010)		0.042*** (0.012)		0.042*** (0.010)		0.046*** (0.012)	
First-year effect		0.033* (0.013)		0.040** (0.013)		0.035* (0.014)		0.045** (0.016)		0.042** (0.014)		0.050** (0.017)
Second-year effect		0.033*** (0.010)		0.037*** (0.010)		0.036*** (0.010)		0.038** (0.013)		0.041*** (0.011)		0.042** (0.013)
Adj. R-squared	0.547	0.547	0.548	0.548	0.551	0.551	0.552	0.552	0.552	0.552	0.557	0.557
N (student/year)	3,734,900	3,734,900	3,734,900	3,734,900	3,436,137	3,436,137	2,477,005	2,477,005	3,434,877	3,434,877	2,473,848	2,473,848
N (districts)	95	95	95	95	95	95	95	95	95	95	95	95
	Longitudinal		Longitudinal		Longitudinal		Longitudinal		Longitudinal		Longitudinal	
	All	All	PI	PI	API	API	AYP	AYP	API (PI)	API (PI)	AYP (PI)	AYP (PI)
Two-year treatment effect	0.032*** (0.009)		0.038*** (0.010)		0.035*** (0.010)		0.041*** (0.012)		0.041*** (0.010)		0.046*** (0.012)	
First-year effect		0.033* (0.013)		0.040** (0.013)		0.035* (0.014)		0.045** (0.016)		0.041** (0.014)		0.050** (0.017)
Second-year effect		0.032** (0.009)		0.037*** (0.010)		0.036*** (0.010)		0.038** (0.013)		0.041*** (0.011)		0.042** (0.013)
Adj. R-squared	0.547	0.547	0.548	0.548	0.552	0.552	0.554	0.554	0.553	0.553	0.557	0.557
N (student/year)	3,734,900	3,734,900	3,734,900	3,734,900	3,424,745	3,424,745	2,473,352	2,473,352	3,423,485	3,423,485	2,473,351	2,473,351
N (districts)	95	95	95	95	95	95	95	95	95	95	95	95

$^{+}p < 0.10$; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

Note: District-clustered standard errors in parentheses. All models include all covariates included in Table 1 as well as district and year fixed effects.

Table A5. The percent of students in each proficiency level for DAIT and TA districts in 2007/2008.

	Non-DAIT TA			DAIT		
	Mean	*SD*	*N*	Mean	*SD*	*N*
% ELA Far Below Basic	**0.104**	0.037	52	**0.140**	0.038	43
% ELA Below Basic	**0.169**	0.036	52	**0.209**	0.032	43
% ELA Basic	**0.319**	0.044	52	**0.345**	0.027	43
% ELA Proficient	**0.258**	0.037	52	**0.217**	0.036	43
% ELA Advanced	**0.150**	0.066	52	**0.089**	0.041	43
% Math Far Below Basic	**0.089**	0.043	52	**0.110**	0.042	43
% Math Below Basic	**0.251**	0.06	52	**0.290**	0.047	43
% Math Basic	**0.281**	0.033	52	**0.276**	0.026	43
% Math Proficient	**0.246**	0.05	52	**0.220**	0.043	43
% Math Advanced	**0.133**	0.055	52	**0.104**	0.035	43

Table A6. Summary statistics for factor analysis-generated and summative index measures used in descriptive analyses.

	Cronbach α*		Mean (*SD*)		
Focus on . . .	2008 to 2009	2009 to 2010	2008 to 2009	2009 to 2010	*D*
ELA	.73	.71	0.000 (1.000)	0.000 (1.000)	0.000 (0.6156)
Math	.79	.62	0.000 (1.000)	0.000 (1.000)	0.000 (0.7830)
Data management	.90	.84	0.000 (1.000)	0.000 (1.000)	0.000 (0.9917)
Use of data in instruction	.72	.76	0.000 (1.000)	0.000 (1.000)	0.000 (0.6467)
Open communication w/in district	.79	.75	0.000 (1.000)	0.000 (1.000)	0.000 (0.9674)
Within-district accountability**	.91	.82	4.140 (1.641)	5.348 (1.343)	1.209 (1.389)
High expectations of students**	.50	.60	4.023 (1.551)	5.047 (1.343)	1.203 (0.9383)
Fiscal responsibility	.87	.88	0.000 (1.000)	0.000 (1.000)	0.000 (0.6834)

Notes:: The survey questions used to generate these variables used Likert scale ratings. Likert scale ratings are commonly used scales in survey design that allow respondents to choose between different intensities of response to a given item. For instance, in our district leader survey, we ask respondents to rate the degree to which the DAIT providers helped them implement a given priority. Respondents were given the response options "not at all," "minimally," "somewhat," and "to a great extent."

*Cronbach alphas measure the internal consistency, or reliability, of a measure that is generated from a set of items taken from a survey or other instrument.

**These measures are generated by summing two Likert scale survey measures. In each year, they range from 2 (lowest) to 8 (highest).