

June 2017

This research report was prepared by American Institutes for Research (AIR) by Martyna Citkowicz, PhD; Melissa Brown-Sims, MA; Ryan Williams, PhD; and Dean Gerdeman, PhD.

AIR is one of the world's largest behavioral and social science research and evaluation organizations. AIR's mission is to conduct and apply the best behavioral and social science research and evaluation toward improving people's lives, with a special emphasis on the disadvantaged.

## Iowa's Teacher Leadership and Compensation Program

### Evaluation of the Teacher Leadership and Compensation Program

The Iowa Teacher Leadership and Compensation (TLC) program was launched in the 2014–15 academic year with the following five goals:<sup>1</sup>

- (1) Attract able and promising new teachers by offering competitive starting salaries and offering short-term and long-term professional development and leadership opportunities.
- (2) Retain effective teachers by providing enhanced career opportunities.
- (3) Promote collaboration by developing and supporting opportunities for teachers in schools and school districts statewide to learn from each other.
- (4) Reward professional growth and effective teaching by providing pathways for career opportunities that come with increased leadership responsibilities and involve increased compensation.
- (5) Improve student achievement by strengthening instruction.

The TLC program was rolled out in three successive district cohorts, each covering approximately one third of Iowa's students:

- Cohort 1 in 2014–15 ( $n = 39$  districts)
- Cohort 2 in 2015–16 ( $n = 76$  districts)
- Cohort 3 in 2016–17 ( $n = 221$  districts)

The Iowa Department of Education (DE) contracted American Institutes of Research (AIR) to evaluate TLC in June 2015. The evaluation was designed to inform the DE about TLC's progress related to implementation and intended goals.

The report presents findings on implementation progress; these are mainly based on teacher and administrator surveys, although additional information from interviews and focus groups is occasionally presented. The findings also examine student achievement in TLC-implementing districts, comparing student achievement in Cohorts 1 and 2 with student achievement in Cohort 3. These findings are considered interim because the evaluation is in progress, and additional data on TLC implementation and student achievement will be collected and analyzed over the next 2 years.

<sup>1</sup>These are available on the Iowa Department of Education's website:  
<https://www.educateiowa.gov/teacher-leadership-and-compensation-system>

# Evaluation of TLC Implementation

To examine TLC implementation in 2015–16, we administered surveys and conducted focus groups and interviews in spring 2016. The interim findings about implementation in this report are based on the perspectives of three different sets of respondents: teachers, school administrators, and district administrators.

**Surveys.** We administered statewide online surveys to Iowa teachers and school and district administrators to obtain feedback on changes that accompanied the TLC program implementation. The survey included items related to four areas of change in the TLC program: teacher leadership roles; professional development and supports for teachers; opportunities for teacher collaboration; perceived outcomes of TLC implementation. All Iowa districts were included in the target survey sample, including districts in TLC Cohorts 1 and 2, which were expected to be implementing the program, and districts in Cohort 3, which had not yet started implementation.

Overall, 30% of teachers ( $n = 10,746$ ), 50% of school administrators ( $n = 738$ ), and 22% of district administrators ( $n = 316$ ) in the DE's Basic Educational Data Survey database completed the survey.<sup>2</sup> To reduce a large number of survey items to a smaller set of key constructs, we constructed scale scores by combining related survey items and calculated response percentages for the scale scores.<sup>3</sup> (See Box 1.)

**Box 1. Survey Scale Scores.** Scale scores indicate the degree to which a measured construct is present (for example, the degree to which respondents agree to a set of statements about the utility of supports provided). Higher scores indicate a more positive perception, whereas lower scores indicate a more negative perception, generally speaking. We categorized these scores along the original response options for each construct (e.g., *disagree strongly*, *disagree somewhat*, *agree somewhat*, or *agree strongly*), where the lowest scale scores were categorized in the lowest response categories and the highest scale scores were in the highest response categories. We then calculated percentages of respondents in each category in the scale to highlight the typical responses from surveyed teachers and administrators. For example, when we asked a series of questions about the presence of opportunities to advance into leadership roles, the 75% of surveyed teachers fell in either the *agree strongly* or *agree somewhat* categories, indicating opportunities to advance were available and attainable in their school or district. We calculated overall percentages, as well as percentages by cohort and other subgroupings of respondents. For findings presented in this report, we tested for cohort differences in the extreme category percentages, such as *agree strongly*. The survey methodological approach is presented in more detail in Appendix A.

Because TLC has three cohorts, and each cohort begins implementing the program in a different year, we were able to compare survey responses based on the presence of the program (comparing Cohorts 1 and 2 with Cohort 3), and based on the maturity of the program (comparing Cohorts 1 and 2).

<sup>2</sup> We examined the representativeness of our sample along years of experience and degree earned to examine sample representativeness using a raking technique (Battaglia, Hoaglin, & Frankel, 2009).

<sup>3</sup> We used the Rasch rating scale model (Wright & Masters, 1982), using Winsteps® (Linacre, 2015) for these analyses.

For TLC-implementing districts (Cohorts 1 and 2), we also examined whether survey responses varied across districts based on district size tier (an approach for categorizing districts based on the number of students they serve),<sup>4</sup> Area Education Agency (AEA) that serves the districts, the grade band the teacher and school administrators serve (Grades K–5, 6–8, and 9–12),<sup>5</sup> years of teaching experience (early career versus veteran teachers),<sup>6</sup> and teacher role (whether the respondent is a teacher leader or a regular classroom teacher).<sup>7,8</sup>

**Focus groups and interviews.** We invited a randomly selected group of Cohort 1 and Cohort 2 teacher leaders across 10 randomly selected districts to participate in 90-minute focus groups on TLC program implementation. Ninety-three teacher leaders (39 from Cohort 1 and 54 from Cohort 2) participated, including those in lead teacher, mentor teacher, model teacher, instructional coach, and curriculum or professional development leader roles. We also interviewed 10 superintendents or assistant superintendents from the selected districts. Focus groups utilized iClicker software, which allowed for quick and anonymous polling of the respondents. We analyzed iClicker response data, transcripts for patterns, themes, and categories to determine the most important findings and key similarities and differences across the focus group and interview responses.

## Findings on TLC Program Implementation

The following sections provide findings related to the early implementation of TLC that focus on four main areas: teacher leaders, supports for teachers, teacher collaboration, and perceived outcomes. To examine differences in perspectives related to length of implementation, we highlight differences across the three TLC cohorts, with Cohort 1 in its second year of implementation, Cohort 2 in its first year of implementation, and Cohort 3 not yet implementing TLC at the time the survey was conducted. Within districts actively implementing TLC at the time of the survey (i.e., Cohorts 1 and 2), we examine contrasts between teacher leaders and regular classroom teachers<sup>9</sup> and contrasts between early career and veteran teachers.<sup>10</sup> We further present contrasts between teachers in different grade bands, contrasts between respondents from different AEAs, and contrasts from respondents from different district size tiers in Appendix B.

<sup>4</sup> District size tiers, as defined by the DE, include: 9,000 or more students (Tier 1), 2,500 students to 8,999 students (Tier 2), 1,000 students to 2,499 students (Tier 3), 600 students to 999 students (Tier 4), 300 students to 599 students (Tier 5), and fewer than 300 students (Tier 6).

<sup>5</sup> Only teacher and school administrators' survey responses were examined by grade band, as district administrators were not asked about the grade bands with which they work (most likely all grade bands). Respondents could select multiple grade bands if they worked across the three populations; thus, a respondent could be included in multiple grade bands in the survey analysis.

<sup>6</sup> Early career teachers were defined as teachers who had been teaching for three years or less, and veteran teachers were defined as teachers who had been teaching at least four years.

<sup>7</sup> Teachers and teacher leaders completed the same survey.

<sup>8</sup> Significant differences in survey responses for the subgroups are reported only when the percentage difference is at least 5%.

<sup>9</sup> Contrasts between teacher leaders and regular classroom teachers were examined for Cohorts 1 and 2 combined, as well as separately for each cohort. The pattern of results was similar for the combined Cohort 1 and 2 teacher leader versus regular classroom teacher contrasts and the disaggregated by cohort contrasts; thus we only present the combined cohort contrasts here (see Appendix B for contrast results disaggregated by cohort).

<sup>10</sup> Survey response rates were higher for teacher leaders and veteran teachers compared to other teachers (see Table A1 in Appendix A).

The findings are correlational and descriptive in nature, and do not provide evidence about the effects of TLC in a causal framework. Differences in responses could be due to preexisting differences among districts and respondents.

## Teacher Leadership

### **Respondents in TLC districts were more familiar with teacher leadership roles compared to respondents in non-TLC districts.**

*“My understanding is that we are working to help teachers be reflective practitioners and that a variety of different positions within the TLC grant are there to help facilitate the way that we implemented that.”*

—Cohort 2 Teacher Leader

Survey respondents were asked a series of questions about how familiar they were with teacher leadership roles in their districts. Respondents in TLC districts were more familiar with teacher leadership roles. Survey respondents in Cohort 1 (57% of teachers and 85% of administrators) and Cohort 2 (52% of teachers and 80% of administrators) were significantly more likely to indicate that they are *very familiar* with teacher leadership roles than respondents in Cohort 3 (22% of teachers and 50% of administrators).

Cohort 1 teachers were significantly more likely to be *very familiar* with teacher leadership roles than respondents in Cohort 2. However, more than 40% of teacher respondents in TLC districts still were not *very familiar* with teacher leadership roles, and teachers reported lower familiarity than administrators.

### **Respondents in TLC districts reported greater availability of teacher leadership roles compared to respondents in non-TLC districts, but teachers in TLC and non-TLC districts had similar perceptions about opportunities to assume teacher leadership roles.**

More teachers in TLC districts reported holding teacher leadership roles compared to teachers in the non-TLC districts, but the differences were small. Overall, 27% of surveyed teachers indicated that they held a teacher leadership role; 26% in Cohort 1 indicated this, with 29% in Cohort 2 and 23% in Cohort 3.

Survey respondents were asked to indicate what teacher leadership roles are available in their school or district. Respondents from TLC districts reported greater availability of several teacher leadership roles compared to non-TLC districts, particularly lead teachers, model teachers, and instructional coaches (Figures 1 and 2). Administrators were more likely than teachers to report the presence of teacher leader roles. All focus group and interview respondents from TLC districts reported that instructional coaches were available in their districts. Moreover, respondents from 5 out of the 10 focus group districts also reported the availability of mentor teacher and mentor coach roles.

Surveyed teachers also were asked a series of questions about the presence of opportunities to advance into leadership roles. Scale scores for 75% of surveyed teachers fell in the *agree strongly* or *agree somewhat* categories; no significant differences were found across teachers in TLC and non-TLC districts.

“One of the biggest things is that there’s a barrier between the coaches and the administration. So we are not evaluative. And we are simply there to help teachers. It’s had a positive impact because teachers can come with us. They come to us and ask us about anything that has to do with teaching.”

—Cohort 1 Teacher Leader

### **More respondents in TLC districts perceived teacher leadership roles as effective compared to respondents in non-TLC districts.**

Based on a series of questions about the effectiveness of teacher leadership roles, scores from a majority of respondents (87% of teachers and 96% of administrators) fell in the *agree somewhat* or *agree strongly* categories, indicating that these roles were perceived as effective in increasing support for teachers and improving student achievement through improved instruction. Scores from respondents in Cohort 1 (42% of teachers and 69% of administrators) and

Cohort 2 (42% of teachers and 55% of administrators) were significantly more likely to fall in the *agree strongly* category compared to those in Cohort 3 (27% of teachers and 10% of administrators), indicating that early adopting TLC cohorts were more likely to view the teacher leadership roles as effective.

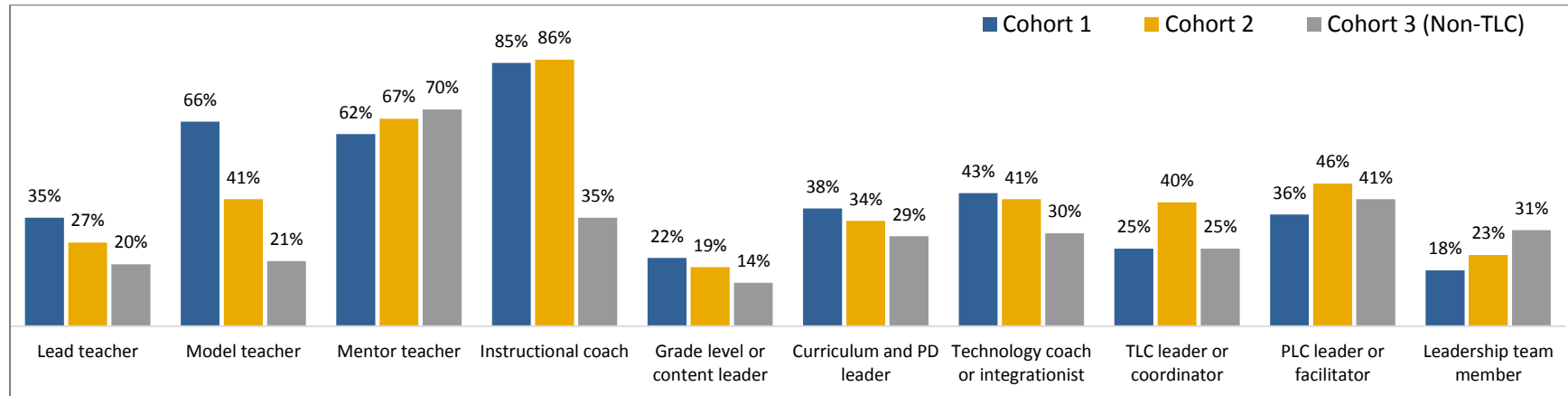
### **Teacher leaders were more aware of and had more positive perceptions about teacher leadership roles compared to other teachers.**

Respondents who self-identified as teacher leaders reported significantly more positive perceptions about teacher leadership compared to other teachers. Within TLC-implementing districts, compared to other teachers, teacher leaders were more likely to indicate that they were *very familiar* with teacher leadership roles (86% to 44%), more likely to *agree strongly* about the presence of opportunities to advance into leadership roles (48% to 12%), and more likely to *agree strongly* about the perceived effectiveness of teacher leadership roles (53% to 38%). Among the teacher leaders who participated in focus groups, all of whom worked in TLC districts, the large majority (81 out of 93) *agreed or agreed strongly* that they personally had participated in targeted professional development opportunities to build their skills as teacher leaders.

### **Early career teachers were less familiar with teacher leadership roles compared to veteran teachers but perceived these roles as more effective.**

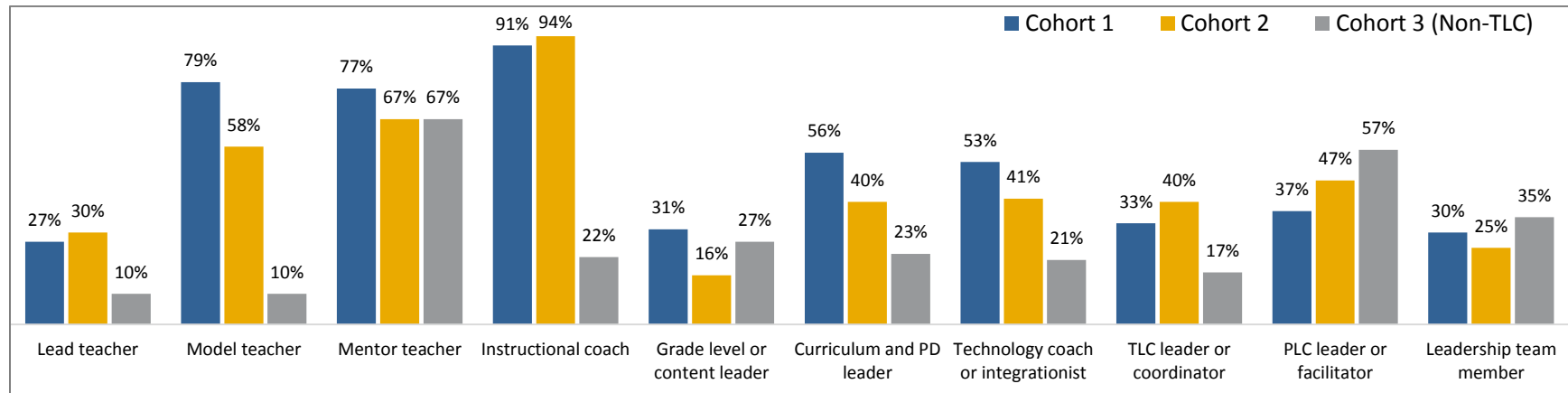
Within TLC-implementing districts, early career teacher respondents (those with three years of experience or less), were less likely to indicate they were *very familiar* with teacher leadership roles compared to veteran teachers (40% to 58%). However, in TLC-implementing districts, early career teachers also were more likely to fall into the *agree strongly* category on the perceived effectiveness of teacher leadership roles scale compared to veteran teachers (53% to 39%).

**Figure 1. Percentage of Surveyed Teachers Who Reported Available Teacher Leadership Roles in Their School or District**



Note: Sample sizes for surveyed teachers and teacher leaders:  $n_{c1} = 3,493$ ,  $n_{c2} = 3,210$ ,  $n_{c3} = 1,118$ . Response options selected by less than 10% of respondents, on average, and response options *Other* and *Do not know* were omitted.

**Figure 2. Percentage of Surveyed Administrators Who Reported Available Teacher Leadership Roles in Their School or District**



Note: Sample sizes for surveyed administrators:  $n_{c1} = 287$ ,  $n_{c2} = 299$ ,  $n_{c3} = 118$ . Response options selected by less than 10% of respondents, on average, and response options *Other* and *Do not know* were omitted.

## Supports for Teachers

***More teachers in TLC districts indicated that professional development supports were available compared to teachers in non-TLC districts.***

Survey respondents were asked about the professional development supports available in their school or district. Across all three TLC cohorts, 99% of surveyed teachers indicated that they participated in some kind of professional development in the 2015–16 school year. Teachers in Cohort 1 (78% of teachers) were significantly more likely to indicate that professional development support was offered at the school *and* district levels (as opposed to just the school or district level) than teachers in Cohort 2 (71% of teachers) or Cohort 3 (70% of teachers).

Teachers and administrators across all three cohorts reported professional development supports that covered a variety of topics (Figures 3 and 4). Teachers in TLC districts more frequently indicated that most of the professional development supports are offered. Differences between administrators in TLC and non-TLC districts were not as consistent as they were among teachers. In nearly all cases, a higher percentage of administrators indicated that professional development supports are offered compared to teachers.

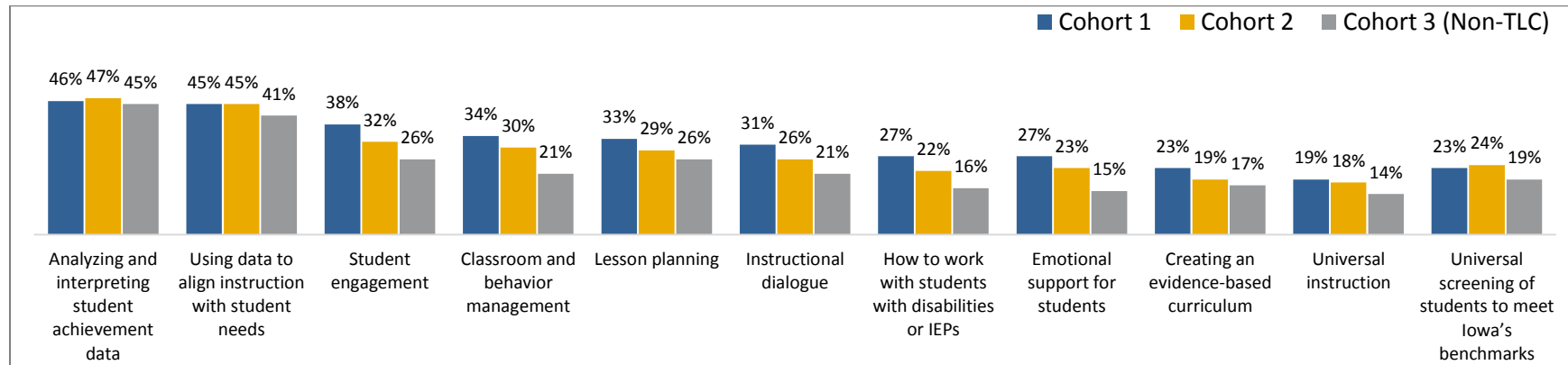
***More teachers and teacher leaders in TLC districts indicated that the professional development supports provided were useful and of high quality compared to other teachers.***

Surveyed teachers responded to two sets of items: one about the utility of available supports and the other about the quality of available supports for improving instruction. Most respondents across cohorts fell in the *agree somewhat* or *agree strongly* categories on survey scales for their perceptions about the utility (80%) and quality (79%) of the supports provided. Teachers in Cohort 1 (25% of teachers) and Cohort 2 (26% of teachers) were significantly more likely to fall in the *agree strongly* category, indicating the provided supports were perceived as useful, compared to teachers in Cohort 3 (17% of teachers). Similarly, teachers in Cohort 1 (26% of teachers) and Cohort 2 (27% of teachers) were significantly more likely to respond *agree strongly* that the supports offered were perceived to be of high quality, compared to teachers in Cohort 3 (18% of teachers).

In TLC-implementing districts, teacher leaders were more likely than other teachers to *agree strongly* that supports were useful (36% to 22%) and of high quality (36% to 23%).

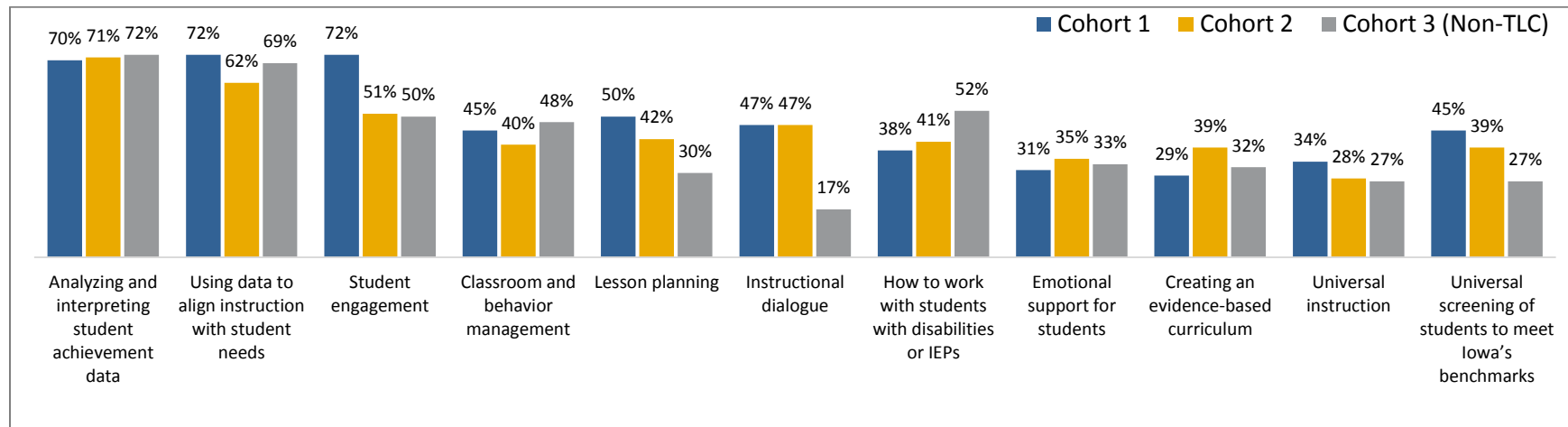


**Figure 3. Percentage of Surveyed Teachers Who Reported\* the Following Teacher Supports Offered in Their School or District**



Note: Sample sizes for surveyed teachers:  $n_{c1} = 3,495$ ,  $n_{c2} = 3,204$ ,  $n_{c3} = 3,622$ ; \*Omitted response options include *Other* and *Do not know*.

**Figure 4. Percentage of Surveyed Administrators Who Reported\* the Following Teacher Supports Offered in Their School or District**



Note: Sample sizes for surveyed administrators:  $n_{c1} = 286$ ,  $n_{c2} = 299$ ,  $n_{c3} = 464$ ; \*Omitted response options include *Other* and *Do not know*.



## Teacher Collaboration

### ***More respondents in TLC districts reported opportunities for teacher collaboration compared to respondents in non-TLC districts.***

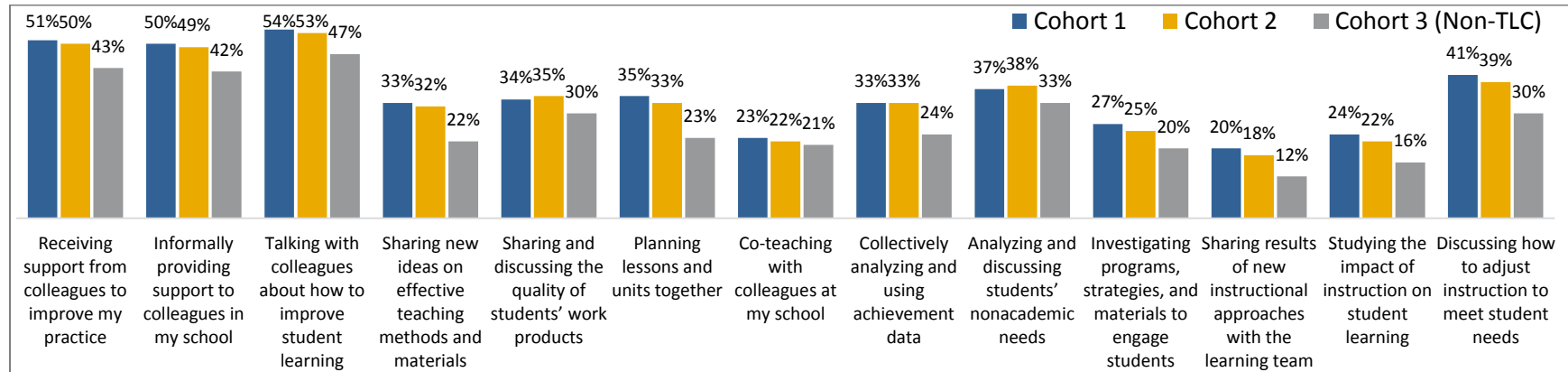
Survey respondents were asked about the frequency with which teachers participate in various collaboration activities. Compared to teachers and administrators in non-TLC districts, surveyed teachers and administrators in TLC districts more frequently reported that teachers participate in weekly collaboration activities (Figures 5 and 6). In all cohorts, teachers and administrators generally had similar perceptions about the frequency of collaboration.

Nearly all of the teacher leaders participating in focus groups (98%) *agreed* or *agreed strongly* that the TLC program provided teachers with opportunities to engage in high-quality collaboration with their peers. The most frequently cited mechanism was professional learning communities.

### ***More teachers and teacher leaders in TLC districts reported greater satisfaction with teacher collaboration compared to other teachers.***

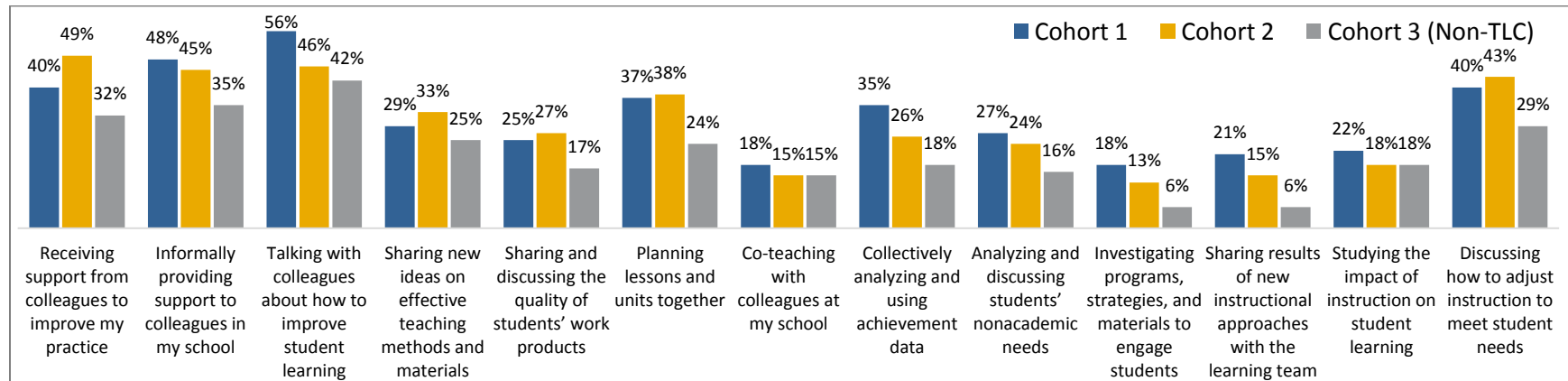
Teachers were asked a series of questions about their satisfaction with the productivity and facilitation of teacher collaboration. Survey scale scores for 82% of all surveyed teachers fell in either the *agree strongly* or the *agree somewhat* categories, indicating that most teacher respondents were satisfied with these aspects of teacher collaboration. Teachers in TLC districts were more likely to be satisfied with these aspects of teacher collaboration, with larger proportions of teachers in Cohort 1 (30% of teachers) and Cohort 2 (29% of teachers) in the *agree strongly* range on the perceived satisfaction with teacher collaboration scale compared to teachers in Cohort 3 (21% of teachers). Within TLC-implementing districts, teacher leaders (34%) were more likely to fall into the *agree strongly* category on the perceived satisfaction with teacher collaboration scale compared to other teachers (29%).

**Figure 5. Percentage of Surveyed Teachers Who Reported That Teachers Participate in the Following Collaboration Activities at Their School at Least Once a Week**



Note: Sample sizes for surveyed teachers:  $n_{c1} = 3,379$ ,  $n_{c2} = 3,115$ ,  $n_{c3} = 3,637$ . Responses for *observing colleagues teaching practice* and *having colleagues observe my teaching practice* were endorsed by less than 10% of respondents, on average, and were not included in this figure.

**Figure 6. Percentage of Surveyed Administrators Who Reported That Teachers Participate in the Following Collaboration Activities in Their School at Least Once a Week**



Note: Sample sizes for surveyed administrators:  $n_{c1} = 276$ ,  $n_{c2} = 286$ ,  $n_{c3} = 448$ . Responses for *observing colleagues teaching practice* and *having colleagues observe my teaching practice* were endorsed by less than 10% of respondents, on average, and were not included in this figure.

## Perceived Outcomes

### **Most respondents in TLC districts reported that TLC is effective in improving instruction.**

Survey respondents in TLC-implementing districts (Cohorts 1 and 2) were asked a series of questions about the effectiveness of TLC in improving instruction. Surveyed teachers and administrators in TLC districts perceived that TLC is improving instruction; responses were similar for both Cohort 1 and Cohort 2 survey respondents. When asked a series of questions about the effectiveness of the TLC program, scale scores for 77% of teachers and 87% of administrators fell in either the *agree somewhat* or the *agree strongly* categories.

*“We don’t know yet. . . . It takes four to five years . . . from what research says to see impact. I can tell you looking at the high school test scores we didn’t have many kids make growth. We had a lot of proficiency, but we didn’t have a lot of kids make growth in certain areas . . . I think it’s having an impact . . . it’s going to trickle down to the kids. Are we seeing fantastic numbers yet? No. Will we? I’m hoping so.”*

—Cohort 1 Teacher Leader

Among focus group participants, nearly all (97%) of the teacher leaders *agreed* or *agreed strongly* that TLC supported effective instruction. The focus group participants indicated that the TLC program supports instruction through its instructional coaches and the increased collaboration among teachers. In interviews, 6 out of 10 district administrators reported that the TLC program had improved teacher instruction in their school as a result of the use of instructional coaches. Interviewed AEA staff reported that TLC improved instruction by improving the instructional strategies used in the classroom (4 out of 12 AEA staff) and by embedding instructional coaches (4 out of 12 AEA staff).

### **Most respondents in TLC districts reported that TLC had a positive impact on their professional work climate.**

Survey respondents in TLC-implementing districts (Cohorts 1 and 2) were asked a series of questions about the perceived TLC-related changes in professional climate. Most teachers and administrators in TLC districts indicated that TLC is positively affecting their professional work climate, although larger percentages of administrators reported positive perceptions; scores for 72% of teachers and 90% of administrators fell in either the *agree somewhat* or the *agree strongly* categories.

### **Teacher leaders were more likely to report positive impacts about TLC outcomes compared to other teachers.**

Within TLC-implementing districts, compared to other teachers, teacher leaders were more likely to fall into the *agree strongly* category on the perceived effectiveness of TLC scale (37% to 15%) and more likely to fall into the *agree strongly* category on the perceived positive changes in professional climate scale (36% to 11%).

***Early career teachers were more likely to report that TLC is effective in improving instruction compared to veteran teachers.***

Within TLC-implementing districts, early career teachers (27%) were more likely to fall into the *agree strongly* category on the perceived effectiveness of TLC scale when compared to veteran teachers (20%).



# Evaluation of Impacts on Student Achievement

To estimate the impact of TLC on student achievement in the first two years of program implementation, we used a comparative interrupted time series (CITS) design.<sup>11</sup> The CITS design used the historical (pre-TLC) performance of students in TLC-implementing districts and students in districts that had not yet implemented the program to predict student achievement outcomes. Because TLC was implemented in three successive

The CITS design uses the historical performance of students in TLC districts and students in districts that had not yet implemented the program to predict post-TLC implementation outcomes. The design does not require that comparison districts be identical to TLC districts.

cohorts, this CITS design had multiple baseline or preintervention periods that allowed us to examine the effects of TLC across one and two years of implementation. To produce the overall Year 1 effect, we compared Cohort 1 districts to districts in Cohorts 2 and 3 in 2014–15 and we compared Cohort 2 districts to Cohort 3 districts in 2015–16. To produce the overall Year 2 effect, we compared Cohort 1 districts to Cohort 3 districts in 2015–16. We also examined the impact of TLC across district, school, and student subgroups, including district size tiers,<sup>12</sup> grade bands (Grades 3–5, 6–8, and 10–11), students eligible for free or reduced-price lunch (FRPL), English language learners (ELLs), students with an individualized education plan (IEP), and TLC schools' participation in the New York City Leadership Academy (NYCLA). The analyses used Iowa Assessment scores in reading and mathematics from 2005–06 to 2015–16, standardized so that scores from different assessments were on the same scale over time.<sup>13</sup>

We included data from all Iowa districts in the primary analysis. The CITS design does not require that comparison districts be identical to TLC districts, and assumes that events that coincide with implementing TLC (e.g., policies implemented, assessment changes) equally affect TLC and non-TLC comparison districts. The CITS design is especially useful for evaluating enacted programs or policies in a statewide rollout over time.

To test the robustness of the analytic methods, we conducted two sensitivity analyses using alternative analytic approaches. First, we used a propensity score matching approach to identify comparison groups of districts that were most similar to TLC districts (as opposed to using data from all available districts). Second, we conducted a different analytic method, a regression discontinuity design, to determine whether a different method would provide similar results to the CITS. These sensitivity analyses are described in detail in Appendix C, with findings included in Appendix D.

<sup>11</sup> CITS designs are among the strongest quasi-experimental designs for causal inference (Shadish, Cook, & Campbell, 2002; St. Clair, Cook, & Hallberg, 2014), and recent methodological studies have demonstrated their comparability to randomized controlled trials—the gold standard for measuring an intervention's effect—using within-study comparisons (e.g., Hallberg, Williams, & Swanlund, 2015; Jacob, Somers, Zhu, & Bloom, 2016).

<sup>12</sup> District size tiers, as defined by the DE, include: 9,000 or more students (Tier 1), 2,500 students to 8,999 students (Tier 2), 1,000 students to 2,499 students (Tier 3), 600 students to 999 students (Tier 4), 300 students to 599 students (Tier 5), and fewer than 300 students (Tier 6).

<sup>13</sup> AIR standardized assessment scores by grade, subject, and year. Since Iowa has multiple testing windows (fall, midyear, and spring) for data provided to AIR, the DE calibrated student scale scores to match the spring testing window.

## Findings on TLC Program Impact

This section describes findings about the impact of TLC on student achievement in the first and second years of program implementation, including overall impact estimates based on data from nearly all students in Iowa followed by findings related to subgroups. We present our results in effect sizes, which convey the direction and magnitude of a relationship.<sup>14</sup> Positive effect sizes indicate that the TLC districts outperformed non-TLC districts, whereas negative effect sizes indicate that non-TLC districts outperformed TLC districts. We also provide context on how observed effect sizes translate into score differences on the Iowa Assessments.

We also conducted two sensitivity analyses to evaluate the robustness of our findings to compositional differences between the TLC and non-TLC districts. For the first sensitivity analysis, we restricted the comparison groups to those students and districts that were most similar to Cohort 1 and Cohort 2 districts in their preintervention academic achievement and demographic compositions. For Cohort 1 in Year 1, we created a matched comparison group from Cohorts 2 and 3. For Cohort 1 in Year 2, we created a matched comparison group from Cohort 3 only. And for Cohort 2 in Year 1, we created a matched comparison group from Cohort 3. For the second sensitivity analysis, we conducted a regression discontinuity analysis, using district TLC application scores and the state-defined cutoffs for Cohorts 1 and 2. Additional details about the sensitivity analyses are provided in Appendix C and the findings are presented in Appendix D.

### ***In the first year of implementation, TLC districts improved slightly less in achievement compared to non-TLC districts.***

In the first year of implementation (i.e., 2014–15 for Cohort 1 and 2015–16 for Cohort 2), student achievement in the TLC districts improved slightly less, by about 0.02 standard deviations in both reading and mathematics, compared to students in non-TLC districts (see Table D1 in Appendix D). This first-year TLC difference is small, but statistically significant.<sup>15</sup> The difference translates into less than a point on the Iowa Assessments that span approximately 200 points.<sup>16</sup>

Historically, TLC Cohort 1 districts performed marginally below the state averages in reading and mathematics, whereas TLC Cohort 2 districts performed marginally above (see Figure 7). Our results show that in the first year of implementation, Cohort 1 districts improved slightly less, by about 0.04 standard deviations (approximately 1–2 points on the Iowa Assessment) in reading and by about 0.05 standard deviations (approximately 1–2 points on the Iowa Assessment) in mathematics, compared to non-TLC districts (Cohorts 2 and 3), whereas Cohort 2 districts remained unchanged compared to non-TLC comparison districts (Cohort 3) in the first year of implementation (see Figure 8).

<sup>14</sup> Effect sizes here are presented in standard deviation units. The standard deviation is a measure of variation or dispersion around the mean, with larger values indicating greater variation and smaller values closer to zero indicating less.

<sup>15</sup> With 10 years of Iowa's historical student achievement data for a large statewide sample, the analytic approach can detect such minor changes in academic performance.

<sup>16</sup> We calculated the standard deviation for the raw Iowa Assessments scale scores in reading and in mathematics, pooling across grades and years, and then multiplied impact estimates by the scale score standard deviation values for an approximation of change in terms of scale score points on the Iowa Assessments.

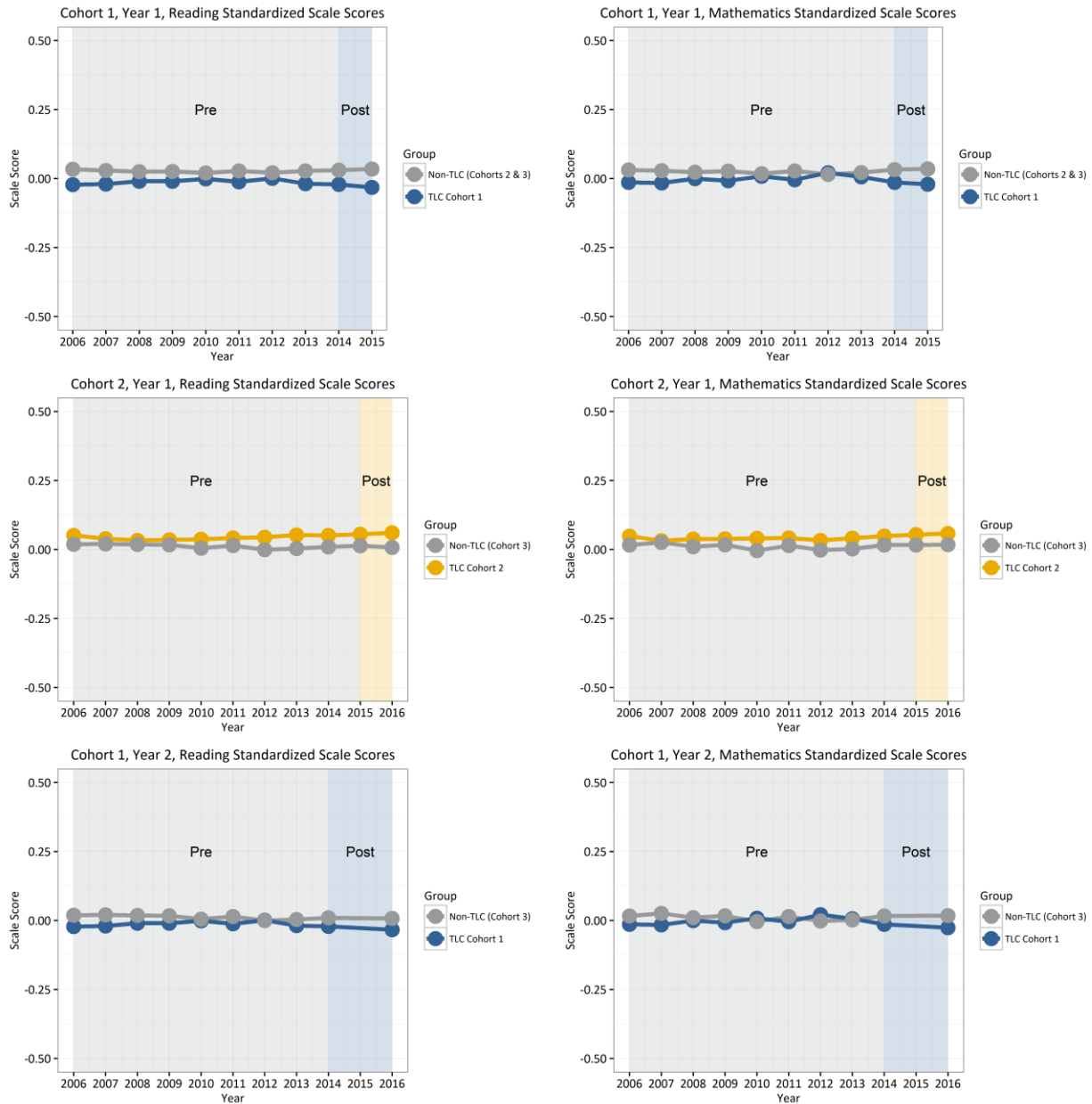
Results from both sensitivity analyses were similar to these results (Table D2 in Appendix D).

***In the second year of implementation, TLC Cohort 1 districts improved slightly less in achievement compared to non-TLC districts.***

Student achievement in the TLC Cohort 1 districts continued to improve less, by about 0.03 standard deviations (or approximately 1 point on the Iowa Assessment) in reading and by about 0.04 standard deviations in mathematics (approximately 1–2 points on the Iowa Assessment), compared non-TLC districts (Cohort 3) in the second year of implementation (see Table D1 in Appendix D). Sensitivity analyses using matched comparison districts from Cohort 3 confirm these results (Table D2 in Appendix D).

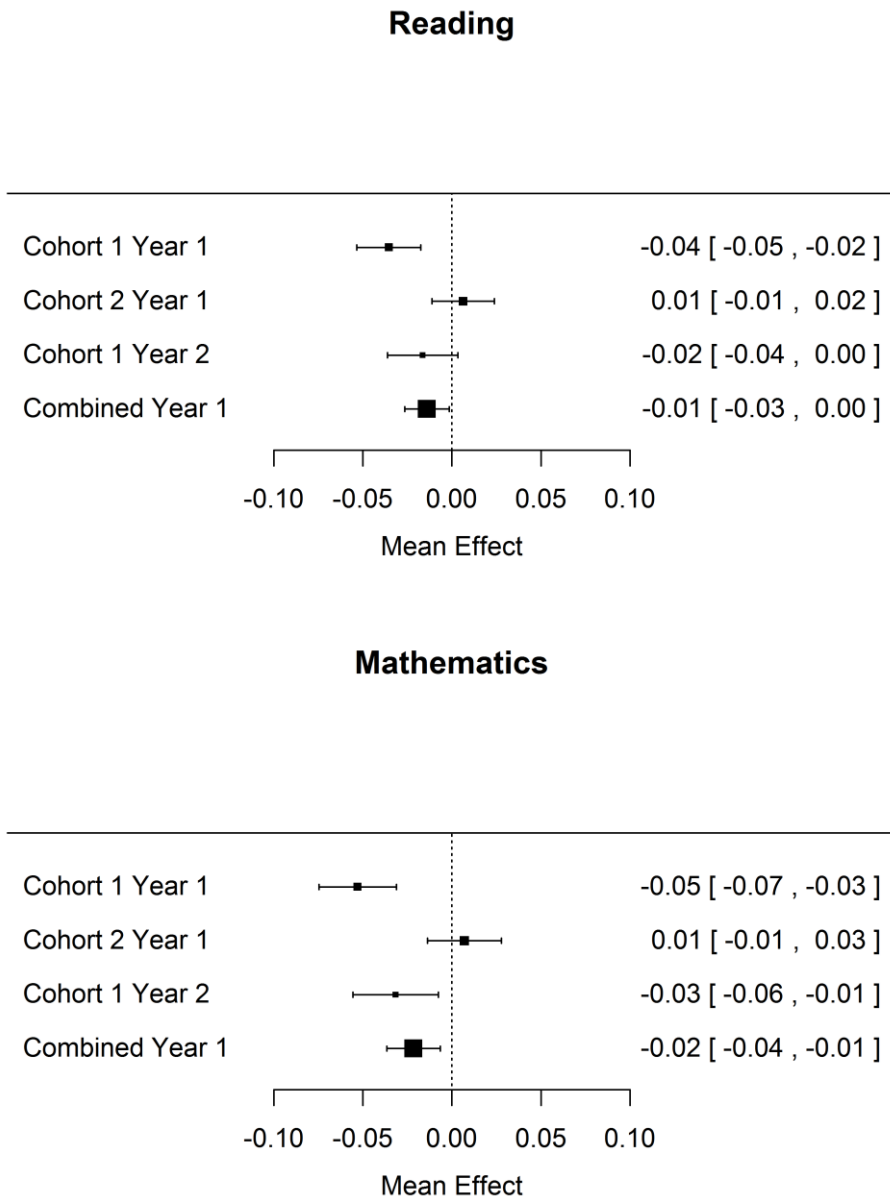


**Figure 7. Average Standardized Reading and Mathematics Scale Scores by School Year and TLC Cohort**



Note. Figure 7 presents the average standardized scores for reading and mathematics for the various TLC cohort comparisons by year of implementation. Unlike the estimates from the impact models, the scores are simple averages; thus, they do not adjust for student- or district-level differences between the TLC and non-TLC districts. The overall mean in each year of data is zero. Values above zero indicate that group’s average was above the state average. Values below zero indicate that group’s average was below the state average.

**Figure 8. Estimated TLC Impact on Student Achievement by Cohort and Year**



Note. Figure 8 presents forest plots of individual TLC cohort impact estimates on TLC-implementing districts (Cohorts 1 and 2) and non-TLC districts (Cohort 3). Combined Year 1 effects then were pooled using meta-analytic weighting techniques that allow the larger sample (in this case, Cohort 2) to contribute more information to calculating the overall Year 1 effect. The black squares represent the impact estimates and the error bars represent 95% confidence intervals. Confidence intervals that include zero indicate non-significant effects (i.e., no change in achievement).

### ***Differences in achievement between TLC districts and non-TLC districts varied by district size.***

We examined whether the impact of TLC in Cohorts 1 and 2 on student achievement differed by district size by estimating the impact *within* each district size tier and examining the differences in estimates between tiers (see Figure 9 and Table D1 in Appendix D).<sup>17</sup> This section highlights significant findings observed.

In the first year of implementation, in districts with 2,500 to 8,999 students, student achievement in Cohort 1 and 2 districts improved less relative to their preintervention average, by about 0.03 standard deviations in reading and by about 0.08 standard deviations in mathematics, compared to the pre-post difference for students in non-TLC districts.

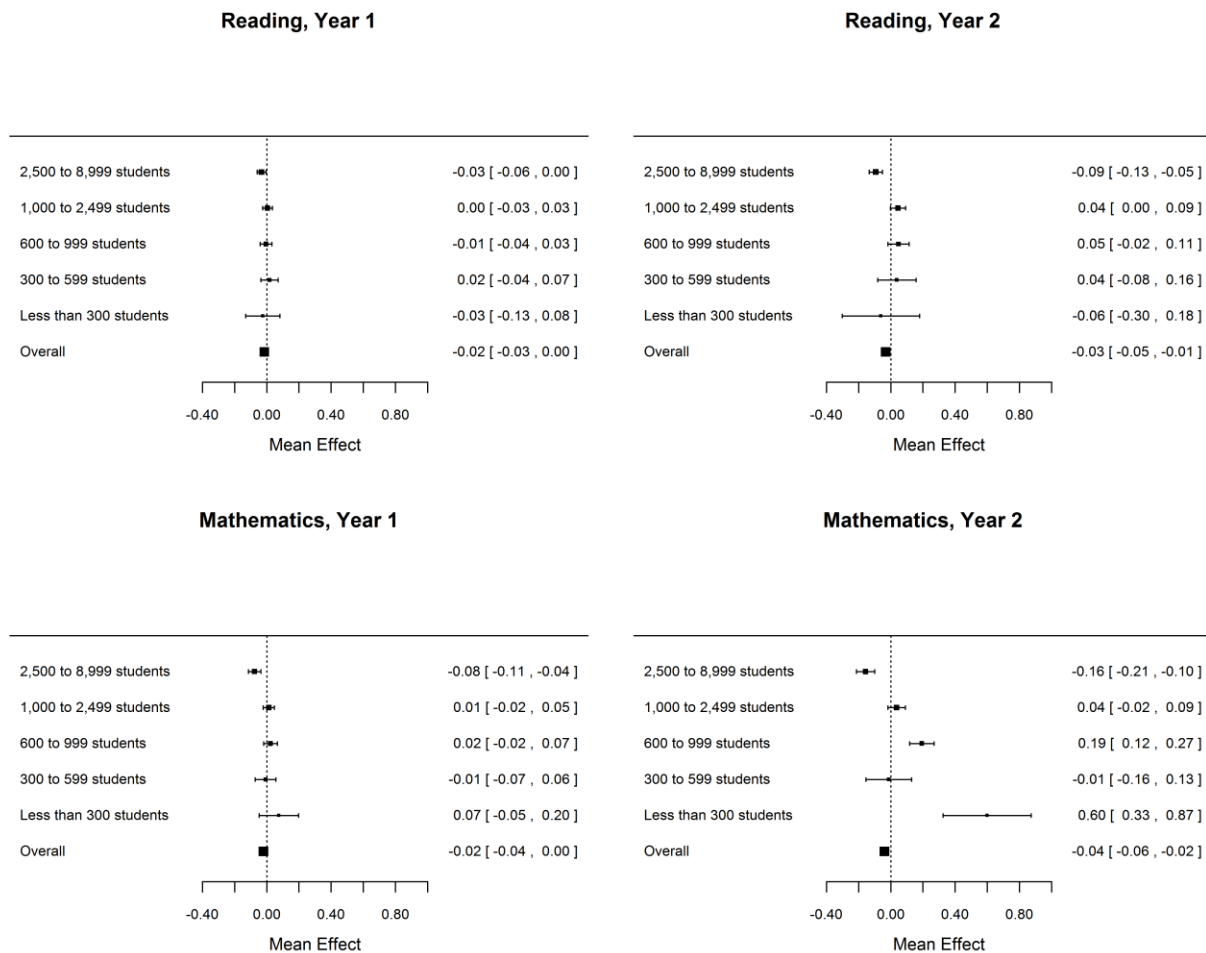
In the second year of implementation, TLC Cohort 1 districts in several size tiers showed significant differences in pre-post intervention achievement compared to the change in achievement in non-TLC districts.

- In districts with 2,500 to 8,999 students, student achievement improved less by about 0.09 standard deviations in reading and by about 0.16 standard deviations in mathematics;
- In districts with 600 to 999 students, student achievement improved by about 0.19 standard deviations in mathematics;
- In districts with fewer than 300 students, student achievement improved by about 0.60 standard deviations in mathematics.

No significant impacts were found for districts with 1,000 to 2,499 students and with 300 to 599 students in either the first or second year of implementation.

<sup>17</sup> District size tiers examined include: 9,000 or more students (Tier 1), 2,500 students to 8,999 students (Tier 2), 1,000 students to 2,499 students (Tier 3), 600 students to 999 students (Tier 4), 300 students to 599 students (Tier 5), and fewer than 300 students (Tier 6). However, impact estimates for students in Tier 1 were not estimated because all Tier 1 districts are included in Cohorts 1 and 2 (i.e., the treatment districts); thus, there are no Tier 1 comparison districts against which we could compare achievement scores.

**Figure 9. Estimated TLC Impact on Student Achievement by District Size Tier and Year**



Note. Figure 9 presents forest plots of TLC impact estimates on TLC-implementing districts (Cohorts 1 and 2) and non-TLC districts (Cohort 3) within each district size tier and TLC implementation year. Effects for Tier 1 size district (districts with 9,000 or more students) were not estimated because all Tier 1 districts were implementing TLC by Year 2 (i.e., all Tier 1 districts are members of Cohorts 1 and 2, thus Tier 1 does not have a comparison group). The overall impact estimates represent the effects estimated from the main model (i.e., the overall, not pooled, impact estimates). The black squares represent the impact estimates and the error bars represent 95% confidence intervals. Confidence intervals that include zero indicate nonsignificant effects (i.e., no change in achievement).

**Differences in achievement between TLC districts and non-TLC districts varied slightly by grade band.**

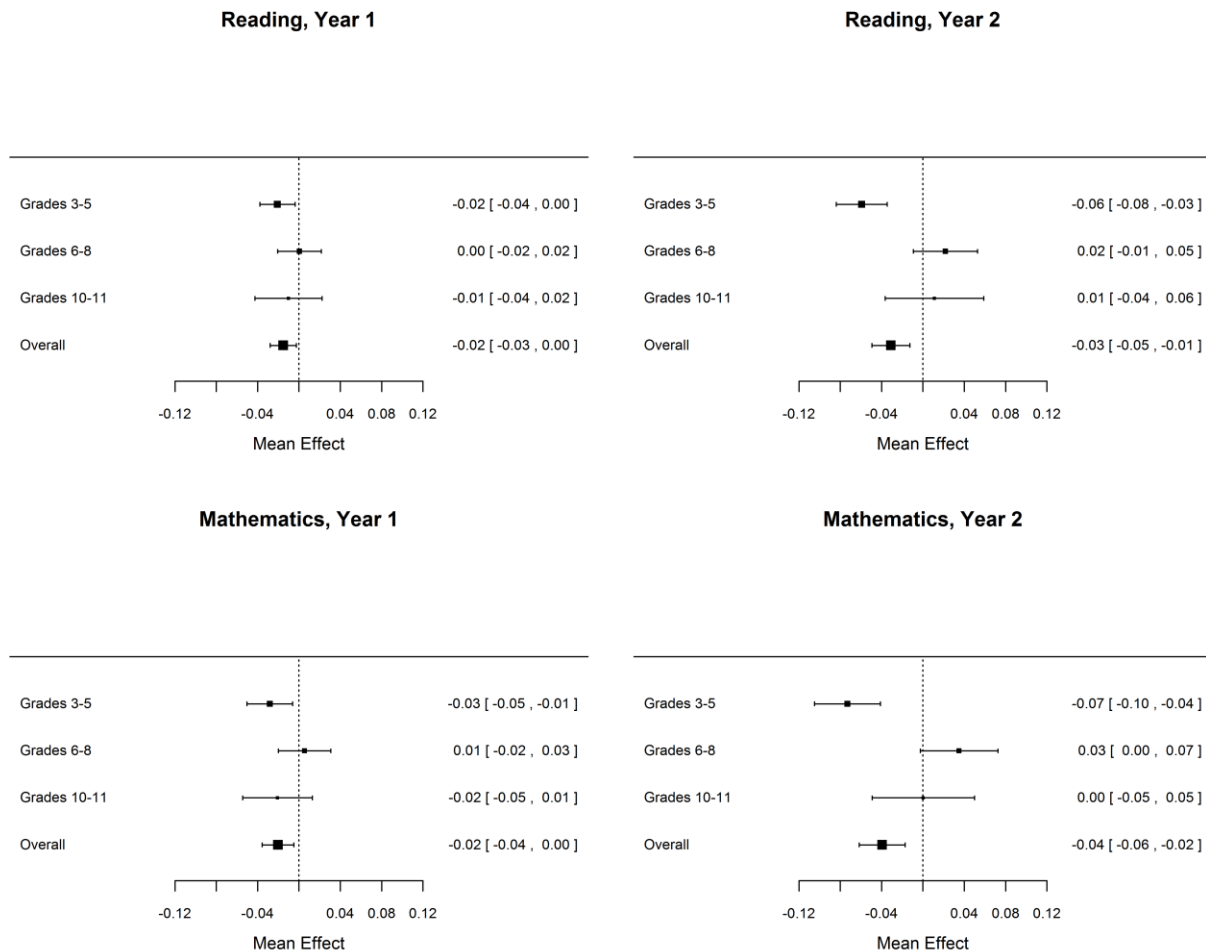
In a similar fashion, we examined whether the impact of TLC in Cohorts 1 and 2 on student achievement differed by grade band (Grades 3–5, 6–8, and 10–11) by estimating the impact *within* each grade band and examining the differences in estimates between grade bands (see Figure 10 and Table D1 in Appendix D). This section highlights significant findings observed.

For students in Grades 3–5, student achievement in Cohort 1 and 2 districts improved slightly less relative to their preintervention average, by about 0.02 standard deviations in reading and by about 0.03 standard deviations in mathematics, compared to the pre-post difference for students in non-

TLC districts in the first year of implementation. In the second year of implementation, for students in Grades 3–5, student achievement in Cohort 1 districts continued to improve less, by about 0.06 standard deviations in reading and by about 0.07 standard deviations in mathematics.

No significant impacts were found for students in Grades 6–8 and 10–11 in either the first or second year of implementation.

**Figure 10. Estimated TLC Impact on Student Achievement by Grade Band and Year**



Note. Figure 10 presents forest plots of TLC impact estimates on TLC-implementing districts (Cohorts 1 and 2) and non-TLC districts (Cohort 3) within each grade band and TLC implementation year. The overall impact estimates represent the effects estimated from the main model (i.e., the overall, not pooled, impact estimates). The black squares represent the impact estimates and the error bars represent 95% confidence intervals. Confidence intervals that include zero indicate non-significant effects (i.e., no change in achievement).

***Achievement gaps decreased slightly for most student subgroups between TLC districts and non-TLC districts.***

We also examined variation in TLC’s impact across different student subgroups, including ELLs, students eligible for FRPL, and students with IEP. Differences in achievement between TLC and non-TLC districts among students defined by a subgroup (e.g., ELL students versus non-ELL students) can

be translated as changes in the achievement gaps, such that a more positive estimated TLC impact among subgroup members (e.g., ELLs) indicates a decrease in the achievement gap with other students (e.g., non-ELLs). This section highlights significant findings observed (see Table D1 in Appendix D).

- The achievement gaps between ELL students and non-ELL students decreased by about 0.07 standard deviations in both reading and mathematics in the first year of implementation and similarly in the second year of implementation.
- The achievement gap between students eligible for FRPL and non-FRPL students decreased by about 0.02 standard deviations in mathematics in the first year of implementation. In the second year of implementation, the achievement gap decreased by about 0.02 standard deviations in reading and 0.04 standard deviations in mathematics.
- The achievement gap in mathematics between students with an IEP and students without an IEP decreased by about 0.02 standard deviations in the first year of implementation and 0.05 standard deviations in the second year of implementation.

***Differences in achievement between TLC districts and non-TLC districts did not vary by NYCLA participation.***

Last, we examined whether TLC impacts on student achievement differed for students at 40 Cohort 1 TLC schools that participated in the first wave of NYCLA in summer 2015, compared to students in non-TLC schools, using postintervention achievement data from the 2015–16 academic year. No significant impacts were found for students in TLC schools that participated in NYCLA (see Table D1 in Appendix D).

## Conclusions and Implications

Teacher and administrator input on TLC implementation was generally consistent with expected progress in implementing of the program's services. Compared to non-TLC districts, respondents from TLC districts were more likely to perceive greater availability, frequency, or quality of key aspects of the TLC program, including teacher leadership roles, professional development supports for teachers, and teacher collaboration. Among districts that had implemented TLC, respondents in Cohort 1 districts, where implementation had occurred for a longer period of time, often had more positive responses, on average, than respondents in Cohort 2 districts. Most teachers and administrators in TLC districts also indicated that the program was effective for improving instruction, teacher satisfaction, and professional climate.

The survey data indicates areas for improvement in program implementation related to teacher awareness and buy-in. Teachers in TLC districts who did not have teacher leadership roles were less familiar with the roles and supports provided by the program and tended to have less positive perceptions, compared to teacher leaders and administrators. One aspect for particular attention is that teacher in TLC districts did not perceive more opportunities for teacher leadership roles. As the program implementation continues, clearly defining and communicating about the roles and supports available for teachers at the local level may help expand teachers' understanding of the professional opportunities that TLC is intended to provide.

Initial analyses of student achievement from the first two years of program implementation—two years for Cohort 1 and one year for Cohort 2—indicate that the program has not resulted in substantial change at this point. The methods used, with a large amount of data, allow detection of very small changes. The small negative effect observed for TLC districts in Cohort 1 approximates to a 1- to 2-point change on the Iowa Assessments, and no change was detected for Cohort 2. The effect of the additional TLC-related supports that districts put in place may not yet be apparent and it is too early to draw clear conclusions about impacts based on student achievement data. As the 2016–17 achievement data become available, further analyses will be conducted focusing on statewide impacts.



## References

- Battaglia, M. P., Hoaglin, D. C., & Frankel, M. R. (2009). Practical considerations in raking survey data. *Survey Practice*, 2(5). Retrieved from <http://www.surveypractice.org/index.php/SurveyPractice/article/view/176/html>
- Hallberg, K., Williams, R. T., & Swanlund, A. P. (2015). *Assessing the validity of comparative interrupted time series designs in practice: Lessons learned from two within-study comparisons*. Paper presented at the annual meeting of the Association for Public Policy Analysis and Management, Miami, FL.
- Hansen, B. B. (2004). Full matching in an observational study of coaching for the SAT. *Journal of the American Statistical Association*, 99(467), 609–618. doi: 10.1198/016214504000000647
- Jacob, R., Somers, M.-A., Zhu, P., & Bloom, H. (2016). The validity of the comparative interrupted time series design for evaluating the effect of school-level interventions. *Evaluation Review*, 0193841X16663414. doi: 10.1177/0193841X16663414
- Linacre, J. M. (2015). *Winsteps® (Version 3.91.0) [Computer software]*. Beaverton, OR: Winsteps.com.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference* (2nd ed.). Boston, MA: Houghton Mifflin.
- St. Clair, T. S., Cook, T. D., & Hallberg, K. (2014). Examining the internal validity and statistical precision of the comparative interrupted time series design by comparison with a randomized experiment. *American Journal of Evaluation*, 35(3), 311–327.
- Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. Chicago, IL: Mesa Press.

# Appendix A. Evaluation of the Teacher Leadership and Compensation Program Survey Methods

American Institutes for Research (AIR) administered surveys to teachers and school and district administrators statewide to obtain feedback on changes that accompanied the Teacher Leadership and Compensation (TLC) program implementation. The survey included topics related to changes in school supports, opportunities for teacher professional development and career advancement, teacher collaboration, and perceived quality and effectiveness of the TLC program. This appendix describes the survey administration process, sample, and analytic approaches taken to examine the survey responses.

## Survey Administration

Prior to administering the surveys statewide, AIR first piloted the teacher and school and district administrator surveys in one TLC district to determine the length of time it took to complete the surveys, to examine the appropriateness of the survey items, and to confirm that the survey scales had a high level of reliability.<sup>18</sup> After reviewing the survey results and scales, adjustments were made to the piloted survey in consultation with the Iowa Department of Education (DE).

In the statewide survey data collection effort, all teachers and school and district administrators (specifically, principals, assistant principals, superintendents, assistant superintendents, directors, department heads, curriculum officers, central office assessment leaders, and any other district administrators in charge of teaching and learning) in 333 Iowa school districts were invited to participate in the Evaluation of the Teacher Leadership and Compensation Program and Teacher Support Survey.<sup>19</sup> We developed three versions of the survey: one for teachers, one for school administrators, and one for district administrators. We administered the surveys online during a four-week period, from April to May 2016.

We sent the appropriate Area Education Agency (AEA) survey links<sup>20</sup> and accompanying survey information to all principals and superintendents, who were asked to distribute the survey link to eligible respondents and follow up with those who had not completed the survey. Superintendents were asked to distribute the district administrator survey invitation, and principals were asked to send an invitation with both the teacher<sup>21</sup> and school administrator survey links. In follow-up e-mails, we shared survey response rates with AEA directors and superintendents, thus allowing AEA directors and superintendents to reach out to individuals to encourage survey participation. In addition, the DE contacted district administrators and AEA directors to encourage survey participation. Our survey team included a survey administrator who was available through e-mail and

<sup>18</sup> As an incentive, at the end of the pilot survey administration, all teachers entered a raffle to win one of three \$50 Amazon gift cards and all district and school administrators entered a raffle to win one of two \$50 Amazon gift cards.

<sup>19</sup> There were a total of 336 districts in Iowa in the 2015–16 school year; however, one district was in the process of dissolution and two districts opted out of AIR's survey due to conflicts with other surveys.

<sup>20</sup> Unique survey links were created for the teacher, school administrator, and district administrator surveys for all nine AEAs, resulting in 27 survey links.

<sup>21</sup> Teachers and teacher leaders completed the teacher survey.

a toll-free telephone number to help individuals who had trouble opening the survey or had any concerns regarding the survey or use of survey results.

All survey respondents were eligible for a raffle as an incentive. Each week, for all four weeks of survey administration, teachers and school and district administrators who completed the survey were entered into a raffle for a chance to win one of three iPad Mini 2s or one of 10 \$50 Amazon gift cards. In addition, we produced a customized district-level report that summarized teacher survey responses for each district that had a 50% or higher teacher survey response rate.

## Sample

We obtained large samples of the statewide target populations. In the 2015–16 school year, Iowa staff included 36,303 teachers, 1,489 school leaders, and 1,412 district leaders. Of these, 10,746 teachers, 738 school leaders, and 316 district leaders completed the TLC survey, resulting in statewide response rates of 30% for teachers, 50% for school leaders, and 22% for district leaders.<sup>22</sup> Table A1 presents the survey response rates and sample sizes by various district-level and respondent characteristics.<sup>23</sup> The table indicates that the majority of respondents were teacher leaders, veteran teachers, teachers from smaller size tier districts, those in AEA 5, and those who work in either elementary or high schools. TLC cohorts are approximately equally represented in the teacher survey, with some variation in the school and administrator surveys.

<sup>22</sup> Completers were defined as respondents who completed at least 50% of their survey items. In addition, the survey data were systematically examined and cleaned according to the following criteria:

1. Respondents who started the survey on multiple occasions and had duplicate records were removed such that the less complete set of responses was removed from the final dataset. If all records were completed, the records with the later dates were removed. As a result, 471 teacher surveys and 189 leader surveys were removed.
2. At the beginning of the survey, there was some confusion about whether teacher leaders should complete the teacher or school leader survey. When it was clear that a respondent completed the wrong survey, the set of responses was removed. This did not affect the teacher survey, but 43 leader surveys were removed.
3. Surveys were examined for patterns that suggested respondents simply “clicked through” and responded to items without reading them. Based on these patterns, respondents who completed the teacher survey in less than five minutes or who completed the leader survey in less than three minutes were removed from the final dataset. As a result, 1,730 teacher surveys and 2,153 leader surveys were removed.
4. Although it was required for respondents to select the district in which they work, some choose to write in a district that we were unable to match to the existing set of districts (e.g., some respondents wrote, “Other”). These records were removed from the final dataset. Eighteen teacher surveys and nine leader surveys were removed.
5. Last, 154 teacher surveys and 101 leader surveys were removed due to responding to less than 50% of their survey. The total count of how many items a respondent was required to answer was adjusted by the way in which they answered a few key questions. For example, if a teacher indicated that professional development was not offered in the current school year, the teacher did not receive the six questions pertaining to professional development. As such, the total number of questions the teacher received was 23 instead of the full 29 questions.

<sup>23</sup> Completion rates were calculated as the total number of respondents who completed the survey in the sample divided by the total number of staff in the population. Population counts were obtained using the Basic Educational Data Survey (BEDS) data provided by the DE.

**Table A1. Survey Response Rates and Sample Sizes by District-Level and Respondent Characteristics**

Characteristic	Teacher Survey		School Administrator Survey		District Administrator Survey	
	<i>N Respondents</i>	Response Rate	<i>N Respondents</i>	Response Rate	<i>N Respondents</i>	Response Rate
<b>TLC Cohorts</b>						
Cohort 1	3,610	27.5%	207	38.8%	82	26.3%
Cohort 2	3,323	24.9%	214	42.3%	87	18.1%
Cohort 3	3,813	28.0%	317	50.9%	147	24.7%
<b>Teacher Leaders Versus Regular Classroom Teachers</b>						
Teacher leaders	2,327	37.3%	-	-	-	-
Regular classroom teachers	5,468	16.2%	-	-	-	-
<b>Veteran Versus Early Career Teacher</b>						
Veteran teachers	9,187	28.5%	-	-	-	-
Early career teachers	1,348	17.3%	-	-	-	-
<b>District Size Tiers</b>						
9,000 or more students	1,867	15.7%	117	25.9%	17	4.7%
2,500 to 8,999 students	2,360	28.3%	164	51.7%	66	36.1%
1,000 to 2,499 students	3,303	32.1%	213	48.1%	106	32.6%
600 to 999 students	1,913	33.9%	129	52.4%	68	25.8%
300 to 599 students	1,112	32.9%	90	55.2%	49	22.4%
Fewer than 300 students	191	27.3%	25	58.1%	10	13.5%
<b>Area Education Agencies</b>						
AEA 1	528	19.5%	41	35.0%	16	14.2%
AEA 5	1,058	41.0%	68	56.7%	32	30.5%
AEA 7	1,576	30.4%	99	42.1%	32	18.2%
AEA 9	1,137	28.9%	72	46.8%	39	47.0%
AEA 10	1,461	26.6%	97	43.1%	37	22.0%
AEA 11	2,320	21.6%	173	44.1%	77	19.2%
AEA 12	899	26.0%	68	43.9%	26	26.3%
AEA 13	915	29.3%	59	42.4%	23	19.3%
AEA 15	852	29.9%	61	48.8%	34	27.0%
<b>School Level</b>						
Elementary school (0-5)	5,384	25.5%	391	42.7%	-	-
Middle school (6-8)	3,476	25.8%	226	29.0%	-	-
High school (9-12)	3,524	25.8%	245	41.6%	-	-

Note. Early career teachers were defined as teachers who had been teaching for three years or less, and veteran teachers were defined as teachers who had been teaching at least four years.

## Differences Between Samples and Populations

Teachers in the sample were similar to teachers in the population on most characteristics, although the sample appears more experienced on average and the percentage of subjects the sample of teachers taught differed from the population. In addition, teachers in the sample earned higher degrees than those in the population. Table A2 presents the demographic characteristics for the sample of teachers who completed the survey and the population of teachers in Iowa.

School and district administrators in the sample were less experienced, held higher level roles in their school or district, and earned higher degrees when compared to the population. In addition, similar percentages of school administrators worked in elementary and high schools when compared to the population, but a smaller percentage of the sample worked in middle schools. Tables A3 and A4 present the demographic characteristics for school and district administrators, respectively.

Some of these differences in our sample of respondents and the population of teachers and school and district leader in Iowa are due to differences in data definitions (e.g., years of experience: the number of years an administrator has been a school/district leader *overall*, in the survey, versus the total number of years of experience completed in K–12 education, in the DE’s administrative data), but some are also likely due to actual differences in the two groups. We recommend caution in interpreting these differences because we are unable to verify what might have caused the differences.

**Table A2. Teacher Demographic Characteristics, Sample Versus Population**

Characteristic	Sample		Population	
	N	Percentage	N	Percentage
<b>Years of Experience</b>				
0–3 years	1,348	12.5%	7,804	19.5%
4–9 years	2,201	20.5%	9,058	22.6%
10–19 years	3,473	32.3%	12,204	30.5%
20 or more years	3,513	32.7%	10,983	27.4%
<b>School Level</b>				
Elementary school (0–5)	5,384	50.1%	21,120	52.7%
Middle school (6–8)	3,476	32.3%	13,468	33.6%
High school (9–12)	3,524	32.8%	13,684	34.2%
<b>Subject Taught</b>				
English language arts	3,098	28.8%	4,958	12.4%
Mathematics	2,387	22.2%	2,879	7.2%
Science	1,669	15.5%	2,473	6.2%
Social studies	1,702	15.8%	2,536	6.3%
Elementary (multiple subjects)	3,551	33.0%	10,604	26.5%
<b>Degree Earned</b>				
Bachelor’s degree	5,560	51.7%	26,313	65.7%

Characteristic	Sample		Population	
	N	Percentage	N	Percentage
Master's degree	4,598	42.8%	12,834	32.0%
Certificate above Master's degree	519	4.8%	35	0.1%
Doctorate or professional degree	45	0.4%	76	0.2%

Note. The background questions on the TLC teacher survey were not required; thus, table cells do not always add up to  $n = 10,746$  (or 100%) due to missing data. Similarly, not all demographic characteristics may be available for the population of Iowa's teachers. The table cells may also add up to more than 100% because some teachers hold multiple positions and teach multiple subjects in multiple schools (and school levels). They were allowed to indicate this on the survey by checking all available response options. Due to teachers holding multiple positions, the population of teachers in this table is a bit larger than the population size noted in the text ( $N = 40,049$ ).

**Table A3. School Leader Demographic Characteristics, Sample Versus Population**

Characteristic	Sample		Population	
	N	Percentage	N	Percentage
<b>Years of Experience</b>				
0–3 years	138	18.7%	98	5.9%
4–9 years	238	32.2%	174	10.5%
10–19 years	280	37.9%	590	35.6%
20 or more years	73	9.9%	795	48.0%
<b>Role</b>				
Principal	629	85.2%	1,301	78.5%
Assistant principal	85	11.5%	357	21.5%
<b>School Level</b>				
Elementary school	391	53.0%	915	55.2%
Middle school	226	30.6%	780	47.1%
High school	245	33.2%	589	35.5%
<b>Degree Earned</b>				
Bachelor's degree	19	2.6%	233	14.1%
Master's degree	506	68.6%	1,243	75.0%
Certificate above Master's degree	188	25.5%	96	5.8%
Doctorate or professional degree	23	3.1%	48	2.9%

Note. The background questions on the TLC school leader survey were not required; thus, table cells do not always add up to  $n = 738$  (or 100%) due to missing data. Similarly, not all demographic characteristics may be available for the population of Iowa's school leaders. The table cells may also add up to more than 100% because some school leaders hold multiple positions in multiple schools (and school levels). They were allowed to indicate this on the survey by checking all available response options. Due to school leaders holding multiple positions, the population of school leaders in this table is a bit larger than the population size noted in the text ( $N = 1,657$ ).

**Table A4. District Leader Demographic Characteristics, Sample Versus Population**

Characteristic	Sample		Population	
	N	Percentage	N	Percentage
<b>Years of Experience</b>				
0–3 years	72	22.8%	98	7.2%
4–9 years	92	29.1%	191	14.0%
10–19 years	105	33.2%	439	32.1%
20 or more years	47	14.9%	640	46.8%
<b>Role</b>				
Superintendent	173	54.7%	300	21.9%
Assistant superintendent	15	4.7%	27	2.0%
Director/coordinator/department head	111	35.1%	746	54.5%
Other administrator	17	5.4%	309	22.6%
<b>Degree Earned</b>				
Bachelor’s degree	14	4.4%	385	28.1%
Master’s degree	76	24.1%	729	53.3%
Certificate above Master’s degree	167	52.8%	146	10.7%
Doctorate or professional degree	58	18.4%	87	6.4%

Note. Most background questions on the TLC district leader survey were not required (indicating one’s role was required); thus, table cells do not always add up to  $n = 316$  due to missing data. Similarly, not all demographic characteristics may be available for the population of Iowa’s district leaders. Some district leaders hold multiple positions in multiple districts. However, they were *not* allowed to indicate this on the survey as they could only choose one response option. The population of district leaders is slightly smaller ( $N = 1,368$ ) than the population size noted in the text. The reason for this is unclear.

## Analytic Approach

The survey included four topics (or domains) related to changes that accompanied the TLC program implementation:

1. **Teacher leadership**, including teacher leadership responsibilities and activities, perceived effectiveness, knowledge of the teacher leadership roles, and opportunities for career advancement
2. **Supports for teachers**, including supports and professional development opportunities, perceived utility of the supports, and perceived quality of the supports
3. **Teacher collaboration**, including collaboration activities and perceived satisfaction with teacher collaboration
4. **Perceived outcomes**, including perceived effectiveness of the TLC program and perceived change in professional structure as a result of TLC program implementation

Items (or questions) in the survey closely align to these four domains. Within each domain, we chose constructs (or concepts) that are important for understanding each domain (see Table A5). For



example, to obtain a comprehensive view of teacher leadership in TLC schools, we asked respondents questions related to four constructs, including (1) teacher leadership responsibilities and activities, (2) perceived effectiveness of these roles, (3) knowledge and understanding of the teacher leadership roles, and (4) opportunities to advance into teacher leadership roles.

For most constructs, we included several items on the surveys to obtain a comprehensive view on the given topic. For other constructs (e.g., quality of supports), only one survey item was necessary to obtain a respondents' perception. For constructs that included multiple survey items, we combined those items into one scale score (discussed in the next section). For individual survey items, the results were reported directly (i.e., without scaling).

### **Scaling**

Using the question-construct links presented in Table A5, we first conducted a psychometric analysis (separately for teachers and school and district administrators) to ensure that appropriate survey items were combined to represent a particular construct (or concept). We combined items to reduce a large set of items to a small number of summary scores that represent each construct. As a result, one or two scale scores, rather than (say) five or 10 individual survey items, may summarize a construct. After we combined the items, we created Rasch scale scores for each construct using Winsteps (Linacre, 2015), a Rasch analysis software program.<sup>24</sup> The scales were examined for item fit and internal consistency. Scale reliability ranged from 0.70 to 0.89 on the teacher survey and from 0.65 to 0.76 on the administrator survey (see Table A5).<sup>25</sup> The scale scores then were converted back into their original metric (i.e., the Likert scale) for ease of interpretation and merged with the DE's administrative data.

### **Weighting**

As previously described, our sample of respondents differs from the population of teachers and school and district administrators in Iowa on various respondent-level characteristics, including years of experience, degree earned, leadership role, and subjects taught. To address these differences, we adjusted, or weighted, the survey sample responses to ensure the responses are representative of the full population of teachers and administrators in Iowa. We weighted the survey results using the raking method, which ensures that the sample sizes of the specified sample characteristics match the corresponding sample sizes for the population (Battaglia, Hoaglin, & Frankel, 2009). The characteristics by which we weighted the survey results include years of experience and degree earned.

<sup>24</sup> Items that were not combined into a single construct were analyzed individually (i.e., item-level frequencies).

<sup>25</sup> A principal component analysis also was conducted to examine multidimensionality. All scale scores were unidimensional (i.e., measure one construct).

## Descriptive Analyses

We conducted descriptive analyses on the converted scale scores as well as on individual survey items.<sup>26</sup> Specifically, we calculated percentages for both converted scale scores and individual survey items to determine the dominant response patterns for each item. The individual item percentages represent the percentage of respondents who selected a specific response option. The converted scale score percentages, on the other hand, identify the percentage of respondents who were most likely to indicate a specific response option to the set of survey items included in the scale score. For example, in Figure A1, we compare teachers in terms of their perceptions about teacher leadership role effectiveness by TLC cohort. Notice that a greater percentage of Cohort 1 and 2 teachers' scores fall in the *agree strongly* category compared to those in Cohort 3. This pattern suggests that teachers in TLC-implementing districts were more likely than teachers in non-TLC districts to view the teacher leadership roles as effective.

We also conducted descriptive analyses on subgroups of individuals in order to examine how survey responses differ by various district-level and respondent characteristics. First, for both the teacher and school and district administrator surveys, we examined how survey responses differ by TLC cohort. Second, for teachers in TLC-implementing districts (i.e., TLC Cohorts 1 and 2), we examined how teacher survey responses differ by the following characteristics: years of teaching experience (early career versus veteran teachers),<sup>27</sup> teacher role (teacher leader versus regular classroom teacher),<sup>28</sup> district size tier,<sup>29</sup> AEA, and grade band (Grades K–5, 6–8, and 9–12).<sup>30</sup> Third, for school and district administrators in TLC-implementing districts, we examined how administrator survey responses differ by district size tier, AEA, and grade band (school administrators only). For subgroups with more than two possible categories of respondents (e.g., AEA), we conducted pairwise comparisons of respondents from each category to respondents in each of the other categories.

We examined whether differences in extreme category percentages (e.g., *agree strongly* or *at least once a week*) are statistically significant (i.e., if they vary by more than chance) between *each* of these subgroups by conducting postestimation Wald tests. Wald tests are used to determine whether two variables are associated, allowing us to test whether the responses to a given research question are associated with subgroup membership (i.e., whether the results differ significantly for various

<sup>26</sup> Scale scores were not produced for research questions and constructs that included fewer than three survey items. For these questions, item-level frequencies were calculated.

<sup>27</sup> Early career teachers were defined as teachers who had been teaching for three years or less, and veteran teachers were defined as teachers who had been teaching at least four years.

<sup>28</sup> Contrasts between teacher leaders and regular classroom teachers were examined for Cohorts 1 and 2 combined, as well as separately for each cohort.

<sup>29</sup> District size tiers, as defined by the DE, include: 9,000 or more students (Tier 1), 2,500 students to 8,999 students (Tier 2), 1,000 students to 2,499 students (Tier 3), 600 students to 999 students (Tier 4), 300 students to 599 students (Tier 5), and fewer than 300 students (Tier 6). The 2014–15 school-level enrollment data was used to define the tiers.

<sup>30</sup> Respondents could select multiple grade bands if they worked across the three populations; thus, a respondent could be included in multiple grade bands in the survey analysis. Of the 6,933 teachers and 421 school administrators that indicated the grade span(s) they work with, 6% of teachers ( $n = 443$ ) and 2% of school administrators ( $n = 9$ ) indicated working in both grade spans K–5 and 6–8, 7% of teachers ( $n = 472$ ) and 5% of school administrators ( $n = 22$ ) indicated working in both grade spans 6–8 and 9–12, and 2% of teachers ( $n = 134$ ) and 1% of school administrators ( $n = 5$ ) indicated working in both grade spans K–5 and 9–12.

subgroups). We reported significant differences in survey responses only when the extreme category percentage difference between the two subgroups was at least 5%.



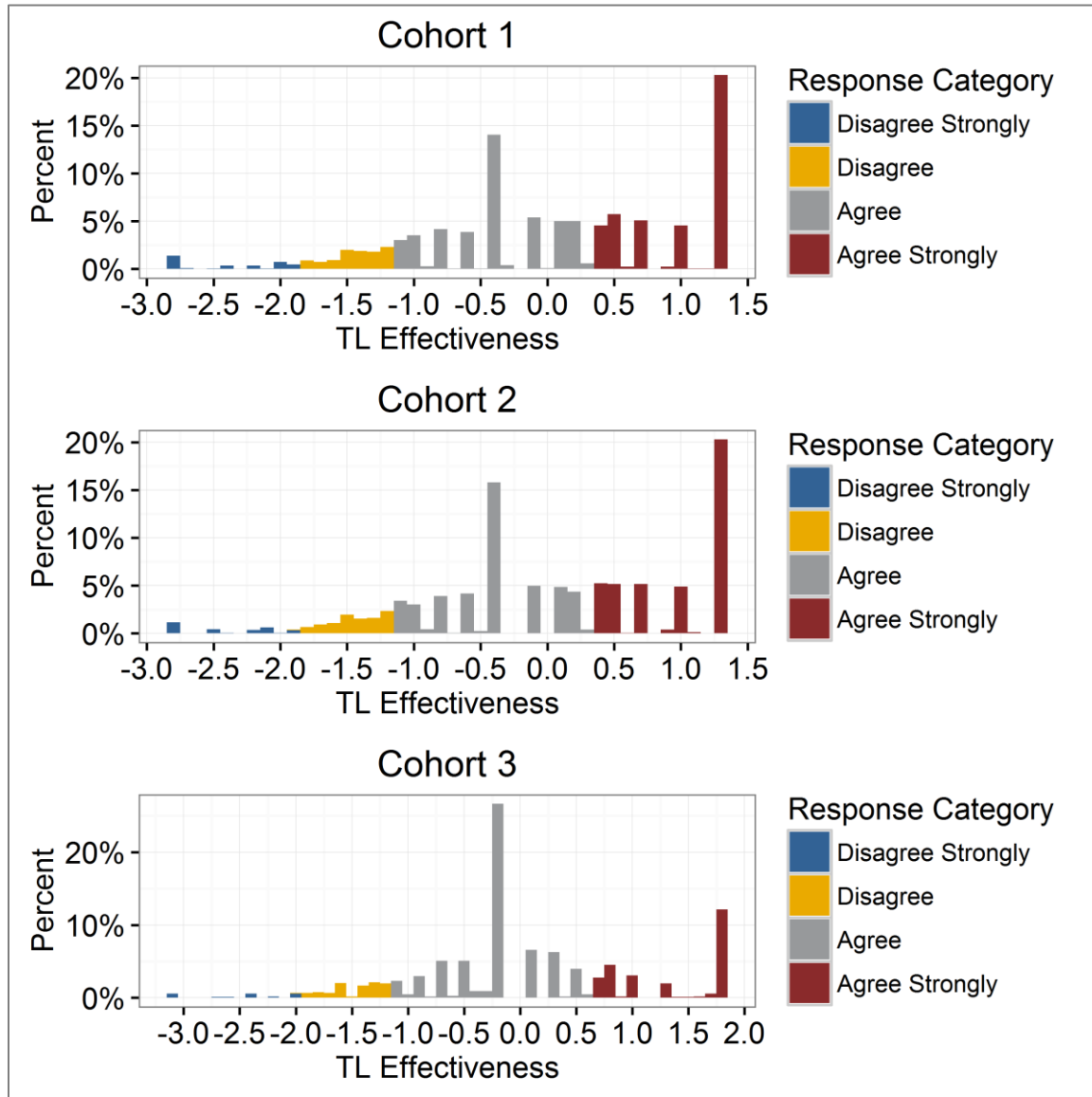
**Table A5. TLC Program and Teacher Support Survey Research Question to Survey Item Analysis Crosswalk**

Research Questions	TLC Goals	Respondent Group	Domain	Construct	Teacher Survey Scale Score Reliability	School/District Administrator Survey Scale Score Reliability
Has the development of TLC created multiple new leadership roles for teachers? What are the responsibilities of teacher leaders?	Goals 1, 2, 4	TLC and non-TLC	Teacher leadership	Responsibilities /Activities	NA	NA
To what extent do teachers and school and district administrators report that teacher leaders are effective in their roles?	Goals 1, 2, 4	TLC and non-TLC	Teacher leadership	Perceived effectiveness	0.89	0.73
To what extent do teachers and school and district administrators report that teachers have a clear understanding of the teacher leadership roles?	NA	TLC and non-TLC	Teacher leadership	Knowledge	NA	NA
To what extent do teachers and school and district administrators report that opportunities are available to advance into teacher leadership roles?	Goals 1-2, 4-5	TLC and non-TLC	Teacher leadership	Opportunity for career advancement	0.70	NA
To what extent do teachers and school and district leaders report that supports are being provided to new and senior teachers? What supports are being provided?	Goals 1-2	TLC and non-TLC	Supports for teachers	Supports and PD provided	NA	NA
To what extent do teachers and school and district administrators report that the supports provided are associated with teachers' impact on instructional practice, satisfaction, and efficacy?	Goals 1-2, 4-5	TLC and non-TLC	Supports for teachers	Perceived utility	0.89	NA
To what extent do teachers and school and district administrators report that the supports provided are associated with teachers' impact on instructional practice, satisfaction, and efficacy?	Goals 1-2, 4-5	TLC and non-TLC	Supports for teachers	Perceived quality	NA	NA

Research Questions	TLC Goals	Respondent Group	Domain	Construct	Teacher Survey Scale Score Reliability	School/District Administrator Survey Scale Score Reliability
To what extent do teachers and school and district administrators report that there is time for teacher collaboration? How is this time being used?	Goal 3	TLC and non-TLC	Teacher collaboration	Collaboration activities	NA	NA
To what extent do teachers and school and district administrators report that teacher collaboration is associated with teacher productivity and satisfaction?	Goal 3	TLC and non-TLC	Teacher collaboration	Perceived satisfaction	0.82	NA
To what extent do teachers and school and district administrators report that TLC is associated with teachers' impact on instructional practice, satisfaction, and efficacy?	Goals 4–5	TLC only	Perceived outcomes	Perceived effectiveness of TLC	0.85	0.76
To what extent do teachers and school and district administrators perceive a positive change in the professional structure after TLC implementation?	Goals 4–5	TLC only	Perceived outcomes	Perceived change in professional structure	0.86	0.68
To what extent do school and district leaders report that there is a shared level of responsibility in their school and districts?	NA	TLC and non-TLC	School climate	Distributive leadership	NA	0.65

Note. The TLC goals refer to the following five goals: (1) attract able and promising new teachers by offering competitive starting salaries and offering short-term and long-term professional development and leadership opportunities, (2) retain effective teachers by providing enhanced career opportunities, (3) promote collaboration by developing and supporting opportunities for teachers in schools and school districts statewide to learn from each other, (4) reward professional growth and effective teaching by providing pathways for career opportunities that come with increased leadership responsibilities and involve increased compensation, and (5) improve student achievement by strengthening instruction.

**Figure A1. Example Percentage of Teachers that Fall Into the Four Response Categories Based on Their Teacher Leader Effectiveness Scale Scores**



## Appendix B. Teacher Leadership and Compensation Program Survey Subgroup Results

This appendix presents the survey subgroup analyses for the key constructs included in Table A5 in Appendix A, including items related to teacher leadership, supports for teachers, teacher collaboration, and perceived outcomes of TLC. Teacher and administrator survey responses were first examined by TLC cohort. For TLC implementing districts (i.e., TLC Cohorts 1 and 2), survey responses were also examined by years of teaching experience (early career versus veteran teachers),<sup>31</sup> teacher role (teacher leader versus regular classroom teacher; for TLC Cohorts 1 and 2 combined and separately), district size tier,<sup>32</sup> area education agency (AEA), and grade band (Grades K–5, 6–8, and 9–12). Specifically, we examined whether differences in extreme category percentages (e.g., *agree strongly* or *very familiar*) were statistically significant between *each* of these subgroups. We reported significant differences in survey responses in the “Differs from” column only when the extreme category percentage difference between the two subgroups was at least 5%.

**Table B1. Percentage of Teachers Who Responded *Very Familiar* to Familiarity with Teacher Leadership Roles**

Group	Percentage Very Familiar	N	Differs From
<b>TLC Cohorts</b>			
Cohort 1	57%	2,146	Cohort 2, Cohort 3
Cohort 2	52%	1,788	Cohort 1, Cohort 3
Cohort 3	22%	927	Cohort 1, Cohort 2
<b>Teacher Leaders Versus Regular Classroom Teachers</b>			
Teacher leaders	86%	1,753	Regular classroom teachers
Regular classroom teachers	44%	2,126	Teacher leaders
<b>Teacher Leaders Versus Regular Classroom Teachers (Cohort 1)</b>			
Teacher leaders	88%	898	Regular classroom teachers
Regular classroom teachers	47%	1,217	Teacher leaders
<b>Teacher Leaders Versus Regular Classroom Teachers (Cohort 2)</b>			
Teacher leaders	84%	855	Regular classroom teachers
Regular classroom teachers	40%	909	Teacher leaders
<b>Veteran Versus Early Career Teachers</b>			
Veteran teachers	58%	3,513	Early career teachers
Early career teachers	40%	344	Veteran teachers

<sup>31</sup> Early career teachers were defined as teachers who had been teaching for three years or less, and veteran teachers were defined as teachers who had been teaching at least four years.

<sup>32</sup> District size tiers, as defined by the DE, include: 9,000 or more students (Tier 1), 2,500 students to 8,999 students (Tier 2), 1,000 students to 2,499 students (Tier 3), 600 students to 999 students (Tier 4), 300 students to 599 students (Tier 5), and less than 300 students (Tier 6).



Group	Percentage Very Familiar	N	Differs From
<b>District Size Tiers</b>			
9,000 or more students	54%	1,066	
2,500 to 8,999 students	53%	1,165	
1,000 to 2,499 students	56%	1,085	600 to 999 students
600 to 999 students	51%	415	1,000 to 2,499 students
300 to 599 students	53%	161	
Fewer than 300 students	63%	42	
<b>Area Education Agencies</b>			
AEA 1	55%	222	AEA 12
AEA 5	50%	257	AEA 10, AEA 11
AEA 7	52%	288	
AEA 9	56%	569	AEA 12
AEA 10	56%	685	AEA 5, AEA 12
AEA 11	57%	1,109	AEA 5, AEA 12, AEA 13
AEA 12	45%	249	AEA 1, AEA 9, AEA 10, AEA 11, AEA 15
AEA 13	51%	236	AEA 11
AEA 15	54%	319	AEA 12
<b>Grade Bands</b>			
Grades K-5	54%	1,946	Grades 6-8
Grades 6-8	49%	1,108	Grades K-5
Grades 9-12	51%	1,075	

**Table B2. Percentage of Administrators Who Responded *Very Familiar* to Familiarity with Teacher Leadership Roles**

Group	Percentage Very Familiar	N	Differs From
<b>TLC Cohorts</b>			
Cohort 1	85%	258	Cohort 3
Cohort 2	80%	277	Cohort 3
Cohort 3	50%	294	Cohort 1, Cohort 2
<b>District Size Tiers</b>			
9,000 or more students	88%	121	Fewer than 300 students
2,500 to 8,999 students	69%	171	1,000 to 2,499 students, 300 to 599 students, fewer than 300 students

Group	Percentage Very Familiar	N	Differs From
1,000 to 2,499 students	98%	155	2,500 to 8,999 students
600 to 999 students	80%	53	
300 to 599 students	96%	23	2,500 to 8,999 students
Fewer than 300 students	100%	12	9,000 or more students, 2,500 to 8,999 students
<b>Area Education Agencies</b>			
AEA 1	69%	29	AEA 13
AEA 5	94%	28	
AEA 7	78%	36	
AEA 9	81%	71	AEA 13
AEA 10	92%	93	
AEA 11	73%	164	AEA 13
AEA 12	92%	34	
AEA 13	100%	30	AEA 1, AEA 9, AEA 11
AEA 15	89%	50	
<b>Grade Bands</b>			
Grades K-5	91%	214	
Grades 6-8	85%	94	
Grades 9-12	87%	101	

**Table B3. Percentage of Teachers Who Fell into the Agree Strongly Category for the Perceived Opportunities to Assume Teacher Leadership Roles Scale**

Group	Percentage Agree Strongly	N	Differs From
<b>TLC Cohorts</b>			
Cohort 1	20%	707	
Cohort 2	22%	714	
Cohort 3	19%	729	
<b>Teacher Leaders Versus Regular Classroom Teachers</b>			
Teacher leaders	48%	927	Regular classroom teachers
Regular classroom teachers	12%	475	Teacher leaders
<b>Teacher Leaders Versus Regular Classroom Teachers (Cohort 1)</b>			
Teacher leaders	48%	454	Regular classroom teachers
Regular classroom teachers	12%	243	Teacher leaders
<b>Teacher Leaders Versus Regular Classroom Teachers (Cohort 2)</b>			
Teacher leaders	48%	473	Regular classroom teachers
Regular classroom teachers	11%	232	Teacher leaders

Group	Percentage Agree Strongly	N	Differs From
<b>Veteran Versus Early Career Teachers</b>			
Veteran teachers	21%	1,233	
Early career teachers	20%	167	
<b>District Size Tiers</b>			
9,000 or more students	20%	360	Fewer than 300 students
2,500 to 8,999 students	20%	407	Fewer than 300 students
1,000 to 2,499 students	21%	383	Fewer than 300 students
600 to 999 students	22%	175	Fewer than 300 students
300 to 599 students	23%	69	Fewer than 300 students
Fewer than 300 students	40%	27	9,000 or more students, 2,500 to 8,999 students, 1,000 to 2,499 students, 600 to 999 students, 300 to 599 students
<b>Area Education Agencies</b>			
AEA 1	18%	69	AEA 10, AEA 11
AEA 5	17%	83	AEA 10, AEA 11
AEA 7	17%	91	AEA 10, AEA 11
AEA 9	19%	187	AEA 10
AEA 10	24%	263	AEA 1, AEA 5, AEA 7, AEA 9, AEA 12
AEA 11	24%	438	AEA 1, AEA 5, AEA 7, AEA 12
AEA 12	16%	81	AEA 10, AEA 11, AEA 13
AEA 13	22%	97	AEA 12
AEA 15	20%	112	
<b>Grade Bands</b>			
Grades K-5	21%	706	
Grades 6-8	19%	394	
Grades 9-12	17%	352	

**Table B4. Percentage of Teachers Who Fell Into the Agree Strongly Category for the Perceived Effectiveness of Teacher Leadership Roles Scale**

Group	Percentage Agree Strongly	N	Differs From
<b>TLC Cohorts</b>			
Cohort 1	42%	1,419	Cohort 3
Cohort 2	42%	1,320	Cohort 3
Cohort 3	27%	276	Cohort 1, Cohort 2

Group	Percentage Agree Strongly	N	Differs From
<b>Teacher Leaders Versus Regular Classroom Teachers</b>			
Teacher leaders	53%	1,070	Regular classroom teachers
Regular classroom teachers	38%	1,661	Teacher leaders
<b>Teacher Leaders Versus Regular Classroom Teachers (Cohort 1)</b>			
Teacher leaders	54%	544	Regular classroom teachers
Regular classroom teachers	38%	871	Teacher leaders
<b>Teacher Leaders Versus Regular Classroom Teachers (Cohort 2)</b>			
Teacher leaders	52%	526	Regular classroom teachers
Regular classroom teachers	38%	790	Teacher leaders
<b>Veteran Versus Early Career Teachers</b>			
Veteran teachers	39%	2,257	Early career teachers
Early career teachers	53%	431	Veteran teachers
<b>District Size Tiers</b>			
9,000 or more students	42%	719	
2,500 to 8,999 students	38%	757	1,000 to 2,499 students, 300 to 599 students
1,000 to 2,499 students	45%	780	2,500 to 8,999 students
600 to 999 students	42%	320	
300 to 599 students	47%	136	2,500 to 8,999 students
Fewer than 300 students	47%	27	
<b>Area Education Agencies</b>			
AEA 1	46%	165	AEA 5, AEA 11
AEA 5	37%	165	AEA 1, AEA 10
AEA 7	41%	198	AEA 10
AEA 9	41%	392	AEA 10
AEA 10	49%	513	AEA 5, AEA 7, AEA 9, AEA 11, AEA 13, AEA 15
AEA 11	39%	709	AEA 1, AEA 10
AEA 12	44%	194	
AEA 13	43%	176	AEA 10
AEA 15	42%	227	AEA 10
<b>Grade Bands</b>			
Grades K-5	47%	1,482	Grades 6-8, Grades 9-12
Grades 6-8	37%	736	Grades K-5
Grades 9-12	35%	659	Grades K-5

**Table B5. Percentage of Administrators Who Fell Into the Agree Strongly Category for the Perceived Effectiveness of Teacher Leadership Roles Scale**

Group	Percentage Agree Strongly	N	Differs From
<b>TLC Cohorts</b>			
Cohort 1	69%	206	Cohort 3
Cohort 2	55%	191	Cohort 3
Cohort 3	10%	33	Cohort 1, Cohort 2
<b>District Size Tiers</b>			
9,000 or more students	49%	81	Fewer than 300 students
2,500 to 8,999 students	65%	152	Fewer than 300 students
1,000 to 2,499 students	64%	104	Fewer than 300 students
600 to 999 students	53%	36	Fewer than 300 students
300 to 599 students	34%	14	Fewer than 300 students
Fewer than 300 students	99%	10	9,000 or more students, 2,500 to 8,999 students, 1,000 to 2,499 students, 600 to 999 students, 300 to 599 students
<b>Area Education Agencies</b>			
AEA 1	76%	24	AEA 10, AEA 15
AEA 5	89%	26	AEA 10, AEA 11, AEA 15
AEA 7	64%	25	
AEA 9	72%	60	AEA 10, AEA 15
AEA 10	45%	66	AEA 1, AEA 5, AEA 9
AEA 11	58%	127	AEA 5
AEA 12	74%	24	AEA 15
AEA 13	61%	19	
AEA 15	41%	26	AEA 1, AEA 5, AEA 9, AEA 12
<b>Grade Bands</b>			
Grades K-5	69%	160	Grades 9-12
Grades 6-8	72%	67	Grades 9-12
Grades 9-12	43%	65	Grades K-5, Grades 6-8

**Table B6. Percentage of Teachers Who Indicated That Professional Development Support Was Offered at the School and District Levels**

Group	Percentage Yes	N	Differs From
<b>TLC Cohorts</b>			
Cohort 1	78%	2,789	Cohort 2, Cohort 3
Cohort 2	71%	2,372	Cohort 1
Cohort 3	70%	2,671	Cohort 1

Group	Percentage Yes	N	Differs From
<b>Teacher Leaders Versus Regular Classroom Teachers</b>			
Teacher leaders	75%	1,522	
Regular classroom teachers	74%	3,456	
<b>Teacher Leaders Versus Regular Classroom Teachers (Cohort 1)</b>			
Teacher leaders	78%	772	
Regular classroom teachers	78%	1,924	
<b>Teacher Leaders Versus Regular Classroom Teachers (Cohort 2)</b>			
Teacher leaders	73%	750	
Regular classroom teachers	70%	1,532	
<b>Veteran Versus Early Career Teachers</b>			
Veteran teachers	74%	4,391	
Early career teachers	76%	662	
<b>District Size Tiers</b>			
9,000 or more students	77%	1,434	600 to 999 students
2,500 to 8,999 students	75%	1,557	600 to 999 students
1,000 to 2,499 students	74%	1,371	
600 to 999 students	70%	544	9,000 or more students, 2,500 to 8,999 students
300 to 599 students	73%	210	
Fewer than 300 students	76%	45	
<b>Area Education Agencies</b>			
AEA 1	78%	301	AEA 7, AEA 9, AEA 13, AEA 15
AEA 5	75%	362	
AEA 7	72%	380	AEA 1, AEA 10
AEA 9	72%	712	AEA 1, AEA 10
AEA 10	77%	888	AEA 7, AEA 9, AEA 13, AEA 15
AEA 11	76%	1,420	AEA 13, AEA 15
AEA 12	76%	383	
AEA 13	70%	318	AEA 1, AEA 10, AEA 11
AEA 15	70%	397	AEA 1, AEA 10, AEA 11
<b>Grade Bands</b>			
Grades K-5	75%	2,562	
Grades 6-8	75%	1,547	
Grades 9-12	73%	1,480	

**Table B7. Percentage of Teachers Who Fell Into the Agree Strongly Category for the Perceived Utility of Teacher Supports Provided Scale**

Group	Percentage Agree Strongly	N	Differs From
<b>TLC Cohorts</b>			
Cohort 1	25%	835	Cohort 3
Cohort 2	26%	795	Cohort 3
Cohort 3	17%	600	Cohort 1, Cohort 2
<b>Teacher Leaders Versus Regular Classroom Teachers</b>			
Teacher leaders	36%	687	Regular classroom teachers
Regular classroom teachers	22%	896	Teacher leaders
<b>Teacher Leaders Versus Regular Classroom Teachers (Cohort 1)</b>			
Teacher leaders	38%	358	Regular classroom teachers
Regular classroom teachers	21%	452	Teacher leaders
<b>Teacher Leaders Versus Regular Classroom Teachers (Cohort 1)</b>			
Teacher leaders	34%	329	Regular classroom teachers
Regular classroom teachers	22%	444	Teacher leaders
<b>Veteran Versus Early Career Teachers</b>			
Veteran teachers	24%	1,364	
Early career teachers	28%	232	
<b>District Size Tiers</b>			
9,000 or more students	25%	435	300 to 599 students
2,500 to 8,999 students	26%	497	300 to 599 students
1,000 to 2,499 students	25%	424	300 to 599 students
600 to 999 students	22%	166	300 to 599 students
300 to 599 students	32%	91	9,000 or more students, 2,500 to 8,999 students, 1,000 to 2,499 students, 600 to 999 students
Fewer than 300 students	27%	17	
<b>Area Education Agencies</b>			
AEA 1	25%	86	AEA 10
AEA 5	21%	101	AEA 10
AEA 7	25%	120	AEA 10
AEA 9	24%	221	AEA 10
AEA 10	31%	320	AEA 1, AEA 5, AEA 7, AEA 9, AEA 11, AEA 12, AEA 15
AEA 11	24%	435	AEA 10
AEA 12	23%	108	AEA 10
AEA 13	26%	103	
AEA 15	25%	136	AEA 10

Group	Percentage Agree Strongly	N	Differs From
<b>Grade Bands</b>			
Grades K-5	29%	901	Grades 6-8, Grades 9-12
Grades 6-8	22%	428	Grades K-5
Grades 9-12	19%	357	Grades K-5

**Table B8. Percentage of Teachers Who Responded Agree Strongly That the Teacher Supports Provided are of High Quality**

Group	Percentage Agree Strongly	N	Differs From
<b>TLC Cohorts</b>			
Cohort 1	26%	878	Cohort 3
Cohort 2	27%	832	Cohort 3
Cohort 3	18%	646	Cohort 1, Cohort 2
<b>Teacher Leaders Versus Regular Classroom Teachers</b>			
Teacher leaders	36%	680	Regular classroom teachers
Regular classroom teachers	23%	979	Teacher leaders
<b>Teacher Leaders Versus Regular Classroom Teachers (Cohort 1)</b>			
Teacher leaders	37%	352	Regular classroom teachers
Regular classroom teachers	23%	500	Teacher leaders
<b>Teacher Leaders Versus Regular Classroom Teachers (Cohort 2)</b>			
Teacher leaders	35%	328	Regular classroom teachers
Regular classroom teachers	24%	479	Teacher leaders
<b>Veteran Versus Early Career Teachers</b>			
Veteran teachers	26%	1,432	
Early career teachers	29%	242	
<b>District Size Tiers</b>			
9,000 or more students	25%	423	300 to 599 students
2,500 to 8,999 students	28%	534	300 to 599 students
1,000 to 2,499 students	26%	449	300 to 599 students
600 to 999 students	25%	182	300 to 599 students
300 to 599 students	36%	100	9,000 or more students, 2,500 to 8,999 students, 1,000 to 2,499 students, 600 to 999 students
Fewer than 300 students	37%	22	
<b>Area Education Agencies</b>			
AEA 1	32%	112	AEA 5, AEA 7, AEA 9, AEA 15
AEA 5	24%	110	AEA 1, AEA 10
AEA 7	25%	119	AEA 1, AEA 10



Group	Percentage Agree Strongly	N	Differs From
AEA 9	25%	232	AEA 1, AEA 10
AEA 10	31%	324	AEA 5, AEA 7, AEA 9, AEA 11, AEA 12, AEA 13, AEA 15
AEA 11	26%	460	AEA 10
AEA 12	25%	117	AEA 10
AEA 13	26%	102	AEA 10
AEA 15	23%	134	AEA 1, AEA 10
<b>Grade Bands</b>			
Grades K-5	30%	932	Grades 6-8, Grades 9-12
Grades 6-8	23%	459	Grades K-5
Grades 9-12	20%	383	Grades K-5

**Table B9. Percentage of Teachers Who Fell Into the Agree Strongly Category for the Perceived Satisfaction With Teacher Collaboration Scale**

Group	Percentage Agree Strongly	N	Differs From
<b>TLC Cohorts</b>			
Cohort 1	30%	1,022	Cohort 3
Cohort 2	29%	947	Cohort 3
Cohort 3	21%	755	Cohort 1, Cohort 2
<b>Teacher Leaders Versus Regular Classroom Teachers</b>			
Teacher leaders	34%	651	Regular classroom teachers
Regular classroom teachers	29%	1,253	Teacher leaders
<b>Teacher Leaders Versus Regular Classroom Teachers (Cohort 1)</b>			
Teacher leaders	35%	331	Regular classroom teachers
Regular classroom teachers	29%	659	Teacher leaders
<b>Teacher Leaders Versus Regular Classroom Teachers (Cohort 2)</b>			
Teacher leaders	32%	320	
Regular classroom teachers	28%	594	
<b>Veteran Versus Early Career Teachers</b>			
Veteran teachers	30%	1,673	
Early career teachers	30%	247	
<b>District Size Tiers</b>			
9,000 or more students	29%	498	
2,500 to 8,999 students	30%	598	
1,000 to 2,499 students	32%	561	600 to 999 students
600 to 999 students	26%	203	1,000 to 2,499 students, fewer than 300 students

Group	Percentage Agree Strongly	N	Differs From
300 to 599 students	28%	85	
Fewer than 300 students	41%	24	600 to 999 students
<b>Area Education Agencies</b>			
AEA 1	41%	145	AEA 5, AEA 7, AEA 9, AEA 11, AEA 12, AEA 13, AEA 15
AEA 5	26%	124	AEA 1, AEA 9, AEA 10
AEA 7	26%	131	AEA 1, AEA 10
AEA 9	32%	288	AEA 1, AEA 5, AEA 11
AEA 10	36%	380	AEA 5, AEA 7, AEA 11, AEA 12, AEA 13, AEA 15
AEA 11	27%	485	AEA 1, AEA 9, AEA 10
AEA 12	29%	141	AEA 1, AEA 10
AEA 13	29%	117	AEA 1, AEA 10
AEA 15	28%	158	AEA 1, AEA 10
<b>Grade Bands</b>			
Grades K-5	35%	1,131	Grades 6-8, Grades 9-12
Grades 6-8	27%	545	Grades K-5, Grades 9-12
Grades 9-12	21%	412	Grades K-5, Grades 6-8

**Table B10. Percentage of Teachers Who Fell Into the Agree Strongly Category for the Perceived Effectiveness of TLC Scale**

Group	Percentage Agree Strongly	N	Differs From
<b>TLC Cohorts</b>			
Cohort 1	23%	660	
Cohort 2	20%	586	
<b>Teacher Leaders Versus Regular Classroom Teachers</b>			
Teacher leaders	37%	676	Regular classroom teachers
Regular classroom teachers	15%	550	Teacher leaders
<b>Teacher Leaders Versus Regular Classroom Teachers (Cohort 1)</b>			
Teacher leaders	41%	371	Regular classroom teachers
Regular classroom teachers	16%	280	Teacher leaders
<b>Teacher Leaders Versus Regular Classroom Teachers (Cohort 2)</b>			
Teacher leaders	34%	305	Regular classroom teachers
Regular classroom teachers	15%	270	Teacher leaders
<b>Veteran Versus Early Career Teachers</b>			
Veteran teachers	20%	1,040	Early career teachers
Early career teachers	27%	177	Veteran teachers

Group	Percentage Agree Strongly	N	Differs From
<b>District Size Tiers</b>			
9,000 or more students	20%	293	
2,500 to 8,999 students	23%	385	600 to 999 students
1,000 to 2,499 students	22%	362	
600 to 999 students	18%	126	2,500 to 8,999 students, fewer than 300 students
300 to 599 students	23%	62	
Fewer than 300 students	34%	18	600 to 999 students
<b>Area Education Agencies</b>			
AEA 1	23%	76	
AEA 5	19%	80	
AEA 7	22%	87	
AEA 9	19%	164	AEA 11
AEA 10	23%	217	
AEA 11	24%	383	AEA 9, AEA 12, AEA 13, AEA 15
AEA 12	18%	69	AEA 11
AEA 13	18%	71	AEA 11
AEA 15	19%	99	AEA 11
<b>Grade Bands</b>			
Grades K-5	23%	659	Grades 6-8, Grades 9-12
Grades 6-8	18%	322	Grades K-5
Grades 9-12	14%	242	Grades K-5

Note. Only teachers in TLC-implementing districts (Cohorts 1 and 2) were surveyed about the perceived effectiveness of TLC.

**Table B11. Percentage of Administrators Who Fell Into the Agree Strongly Category for the Perceived Effectiveness of TLC Scale**

Group	Percentage Agree Strongly	N	Differs From
<b>TLC Cohorts</b>			
Cohort 1	38%	123	
Cohort 2	35%	97	
<b>District Size Tiers</b>			
9,000 or more students	22%	41	2,500 to 8,999 students, 300 to 599 students
2,500 to 8,999 students	49%	93	9,000 or more students, 600 to 999 students, 300 to 599 students
1,000 to 2,499 students	34%	58	300 to 599 students
600 to 999 students	18%	18	2,500 to 8,999 students

Group	Percentage Agree Strongly	N	Differs From
300 to 599 students	6%	5	9,000 or more students, 2,500 to 8,999 students, 1,000 to 2,499 students
Fewer than 300 students	48%	5	
<b>Area Education Agencies</b>			
AEA 1	32%	12	
AEA 5	32%	8	
AEA 7	34%	15	
AEA 9	40%	33	
AEA 10	27%	35	
AEA 11	44%	80	
AEA 12	42%	13	
AEA 13	30%	9	
AEA 15	18%	15	
<b>Grade Bands</b>			
Grades K-5	42%	93	
Grades 6-8	37%	31	
Grades 9-12	26%	37	

Note. Only administrators in TLC-implementing districts (Cohorts 1 and 2) were surveyed about the perceived effectiveness of TLC.

**Table B12. Percentage of Teachers Who Fell Into the Agree Strongly Category for the Perceived Positive Changes in Professional Climate Scale**

Group	Percentage Agree Strongly	N	Differs From
<b>TLC Cohorts</b>			
Cohort 1	18%	528	
Cohort 2	19%	550	
<b>Teacher Leaders Versus Regular Classroom Teachers</b>			
Teacher leaders	36%	648	Regular classroom teachers
Regular classroom teachers	11%	408	Teacher leaders
<b>Teacher Leaders Versus Regular Classroom Teachers (Cohort 1)</b>			
Teacher leaders	35%	324	Regular classroom teachers
Regular classroom teachers	11%	194	Teacher leaders
<b>Teacher Leaders Versus Regular Classroom Teachers (Cohort 2)</b>			
Teacher leaders	36%	324	Regular classroom teachers
Regular classroom teachers	12%	214	Teacher leaders
<b>Veteran Versus Early Career Teachers</b>			
Veteran teachers	18%	919	
Early career teachers	21%	136	

Group	Percentage Agree Strongly	N	Differs From
<b>District Size Tiers</b>			
9,000 or more students	17%	251	
2,500 to 8,999 students	18%	310	
1,000 to 2,499 students	20%	327	
600 to 999 students	17%	123	
300 to 599 students	18%	52	
Fewer than 300 students	27%	15	
<b>Area Education Agencies</b>			
AEA 1	18%	62	
AEA 5	17%	75	
AEA 7	15%	65	AEA 10, AEA 11
AEA 9	18%	154	
AEA 10	20%	179	AEA 7
AEA 11	19%	309	AEA 7
AEA 12	19%	78	
AEA 13	17%	65	
AEA 15	18%	91	
<b>Grade Bands</b>			
Grades K-5	19%	543	Grades 9-12
Grades 6-8	15%	286	
Grades 9-12	13%	241	Grades K-5

Note. Only teachers in TLC-implementing districts (Cohorts 1 and 2) were surveyed about the perceived positive changes in professional climate.

**Table B13. Percentage of Administrators Who Fell Into the Agree Strongly Category for the Perceived Positive Changes in Professional Climate Scale**

Group	Percentage Agree Strongly	N	Differs From
<b>TLC Cohorts</b>			
Cohort 1	49%	159	
Cohort 2	55%	149	
<b>District Size Tiers</b>			
9,000 or more students	32%	53	2,500 to 8,999 students
2,500 to 8,999 students	69%	115	9,000 or more students, 1,000 to 2,499 students, 300 to 599 students, fewer than 300 students
1,000 to 2,499 students	47%	89	2,500 to 8,999 students, 300 to 599 students
600 to 999 students	52%	31	

Group	Percentage Agree Strongly	N	Differs From
300 to 599 students	21%	12	2,500 to 8,999 students, 1,000 to 2,499 students
Fewer than 300 students	27%	8	2,500 to 8,999 students
<b>Area Education Agencies</b>			
AEA 1	38%	15	
AEA 5	44%	18	
AEA 7	45%	21	
AEA 9	68%	50	AEA 10
AEA 10	35%	45	AEA 9, AEA 11, AEA 12
AEA 11	63%	94	AEA 10
AEA 12	69%	23	AEA 10
AEA 13	38%	14	
AEA 15	41%	28	
<b>Grade Bands</b>			
Grades K-5	49%	116	
Grades 6-8	58%	51	
Grades 9-12	37%	53	

Note. Only administrators in TLC-implementing districts (Cohorts 1 and 2) were surveyed about the perceived positive changes in professional climate.

## Appendix C. Teacher Leadership Compensation Impacts Analytical Approach

American Institutes for Research (AIR) conducted one set of confirmatory analyses using the comparative interrupted time series (CITS) design and two sets of sensitivity analyses, including CITS analyses with different sets of comparison groups for each cohort and a regression discontinuity (RD) analysis, to estimate the impact of TLC on student achievement. This appendix describes these analyses.

### Comparative Interrupted Time Series Confirmatory Analysis

To estimate student achievement outcomes in the first and second years of TLC program implementation, we compared trends in outcomes from students in 39 TLC Cohort 1 districts and 76 TLC Cohort 2 districts to students in the remaining 221 districts using a multiple baseline multilevel CITS model that accounts for nesting by means of random and fixed effects. The analysis examines changes in the outcomes of students in TLC-implementing districts in the first (2014–15) and second year of program implementation (2015–16) compared with changes in outcomes of students in all other districts that had not yet started implementing the TLC program. Because the CITS design uses the historical, or preintervention, performance of students in TLC districts and students in districts that had not yet implemented the program to predict post-TLC implementation outcomes, the design does not require that comparison districts be identical to TLC districts. In addition, with all years and districts included in the analyses, we have a large amount of statistical power to detect small changes in performance.

The CITS model is represented by the following equation:

$$Y_{itjk} = \beta_0 + \beta_1 TLC_k + \beta_2 Post1_t + \beta_3 Post2_t + \beta_4 PY1_{tk} + \beta_5 PY2_{tk} + \beta_6 Time_t + \beta_7 TLC_k Time_t + Grade_{itjk} + AEA_k + X_{itjk} + z_k + v_{jk} + u_{tjk} + e_{itjk}$$

where  $Y_{itjk}$  is the outcome measure (i.e., standardized scaled achievement score for reading or mathematics) for a student  $i$  at time  $t$  in school  $j$  in district  $k$ ;  $TLC_k$  is an indicator of whether a district  $k$  is a member of a district that started implementing TLC (Cohort 1 started implementing TLC in the 2014–15 school year and Cohort 2 started implementing TLC in the 2015–16 school year);  $Time_t$  is the linear outcome trend across time (school years 2005–06 to 2015–16 are coded –9 through 1, respectively);  $Post1_t$  and  $Post2_t$  are indicators for whether the TLC program was in its first (2014–15) and second (2015–16) years of implementation, respectively; and  $PY1_{tk}$  and  $PY2_{tk}$  are interaction indicators for whether, at time  $t$ , a district started implementing TLC in the first (2014–15) and second (2015–16) years of implementation, respectively (i.e., the post  $x$  intervention interaction terms of interest that represent the first and second year treatment effects, respectively). In this model, each indicator for a student is coded as 1 if it applies to a student and 0 otherwise. For example, a student who has an outcome observed in a TLC Cohort 1 district in 2014–15 would be coded 1 for  $TLC_k$ , 1 for  $Post1_t$ , and 1 for  $PY1_{tk}$  (because 2014–15 is the first year in the post-program implementation time period for districts in Cohort 1). Because  $TLC_k Time_t$  is an interaction

between  $TLC_k$  and  $Time_t$ , the student also would be coded as 1 multiplied by Time for  $TLC_kTime_t$ , which allows for different pretreatment trends for TLC-implementing districts and non-TLC districts. Area Education Agency (AEA) fixed effects ( $AEA_k$ ) and grade fixed effects ( $Grade_{itjk}$ ) are included to allow for students in TLC districts to be compared only with students in non-TLC districts within the same AEA and grade. The model also includes a set of student-level characteristics  $X_{itjk}$  (i.e., gender, race, English language learner [ELL] status, individualized education program [IEP] status, and free or reduced-price lunch [FRPL] status) to account for differences in student characteristics. Random effects were included to account for the residual effects of each district ( $z_k$ ), school ( $v_{jk}$ ), time ( $u_{itjk}$ ), and student ( $e_{itjk}$ ).

### Cohort Analyses

To examine the varying achievement outcome trajectories of the TLC cohorts, we estimated TLC impacts for Cohorts 1 and 2 separately. Specifically, instead of the comprehensive model defined above, we estimated the following CITS model for each cohort and year separately:

$$Y_{itjk} = \beta_0 + \beta_1 TLC_k + \beta_2 Post_t + \beta_3 TLC_k Post_t + \beta_4 Time_t + \beta_5 TLC_k Time_t + Grade_{itjk} + AEA_k + X_{itjk} + z_k + v_{jk} + u_{itjk} + e_{itjk}$$

where  $TLC_k Post_t$  represents the posttreatment effect of interest. The model was run separately for the following groups:

1. Cohort 1 in Year 1: 2014–15 effect for Cohort 1;
2. Cohort 2 in Year 1: 2015–16 effect for Cohort 2; and
3. Cohort 1 in Year 2: 2015–16 effect for Cohort 1.

We also estimated a pooled Year 1 effect, where we combined the postintervention treatment effects ( $TLC_k Post_t$ ) for Cohort 1 Year 1 and Cohort 2 Year 1 using meta-analysis. Meta-analysis is a statistical technique that combines results from multiple effects by weighting the contribution of each estimate of the effect based on the statistical precision with which that effect was estimated. Thus, effects that are estimated from a larger sample (in this case, from Cohort 2) are weighted more heavily in the pooled effect, allowing the larger sample to contribute more information to calculating the overall pooled Year 1 effect. The individual and pooled postintervention treatment effects then were plotted on a forest plot to show how much treatment effects varied between the cohorts and years.



## District Size Tier and Grade Band Subgroup Analyses

We examined TLC impacts for students within Tier 1 through Tier 6 size districts<sup>33</sup> and students within different grade bands (Grades 3–5, 6–8, and 10–11).<sup>34</sup> For each subgroup analysis, the data were limited to the given subgroup and the main CITS model defined above was conducted to allow for the estimation of treatment effects *within* subgroups. The postintervention treatment effects ( $PY1_{tk}$  and  $PY2_{tk}$ ) then were plotted on forest plots—separately for tier size and grade band subgroups—and heterogeneity statistics were calculated to examine how much treatment effects varied between the subgroups.

## Special Populations Subgroup Analyses

We examined how TLC impacts differ by various subgroups, including ELLs, those eligible for FRPL, and those who have an IEP. For each subgroup analysis, additional terms were added to determine whether participating in a TLC district had an impact on increasing or decreasing the achievement gap between students in special populations and other students after program implementation. The general form of the subgroup model is as follows:

$$Y_{itjk} = \beta_0 + \beta_1TLC_k + \beta_2Post1_t + \beta_3Post2_t + \beta_4PY1_{tk} + \beta_5PY2_{tk} + \beta_6TLC_kSubgroup_{itjk} + \beta_7Post1_tSubgroup_{itjk} + \beta_8Post2_tSubgroup_{itjk} + \beta_9PY1_{tk}Subgroup_{itjk} + \beta_{10}PY2_{tk}Subgroup_{itjk} + \beta_{11}Time_t + \beta_{12}TLC_kTime_t + \beta_{13}Time_tSubgroup_{itjk} + \beta_{14}TLC_kTime_tSubgroup_{itjk} + Grade_{itjk} + AEA_k + X_{itjk} + z_k + v_{jk} + u_{tjk} + e_{itjk}$$

where  $PY1_{tk}Subgroup_{itjk}$  and  $PY2_{tk}Subgroup_{itjk}$  were added to indicate whether there was a differential effect of TLC on students in a subgroup versus students not in that subgroup. Specifically, the added interaction terms allow for the estimation of the change in the gap between students in a subgroup or not in TLC districts versus the gap of students in non-TLC districts in the first and second years of program implementation.

## New York City Leadership Academy Analysis

The first cohort of TLC principals participating in the New York City Leadership Academy (NYCLA) started in the summer of 2015. We examined whether achievement outcomes for students in 40 TLC schools participating in the first wave of NYCLA differed from other students in comparison districts, using postintervention data from 2015–16. The CITS model is represented by the following equation:

<sup>33</sup> Iowa Department of Education defined district tier sizes as follows:

- Tier 1: 9,000 or more students (10 districts)
- Tier 2: 2,500 students to 8,999 students (24 districts)
- Tier 3: 1,000 students to 2,499 students (85 districts)
- Tier 4: 600 students to 999 students (94 districts)
- Tier 5: 300 students to 599 students (86 districts)
- Tier 6: fewer than 300 students (37 districts)

<sup>34</sup> Grade fixed effects are excluded in the grade-level subgroup analyses.

$$Y_{itjk} = \beta_0 + \beta_1 NYCLA\_TLC_j + \beta_2 Post_t + \beta_3 NYCLA\_TLC_j Post_t + \beta_4 Time_t + \beta_5 NYCLA\_TLC_j Time_t + Grade_{itjk} + AEA_k + X_{itjk} + z_k + v_{jk} + u_{tjk} + e_{itjk}$$

where  $NYCLA\_TLC_j$  is dummy coded in such a way as to allow for two separate comparisons (i.e., one for TLC versus comparison and one for TLC plus NYCLA versus comparison).  $NYCLA\_TLC_j Post_t$  includes two posttreatment effects of interest: a 2015–16 treatment effect comparing students in TLC-implementing districts to those in non-TLC districts and a 2015–16 treatment effect comparing students in TLC NYCLA schools to those in non-TLC non-NYCLA schools.

## Comparative Interrupted Time Series Sensitivity Analysis

Because TLC Cohort 1 and 2 districts differ somewhat from other districts (particularly in size), as a sensitivity analysis, we compared trends in outcomes from students in the 39 TLC Cohort 1 districts to students in 39 matched comparison districts and from students in the 76 TLC Cohort 2 districts to students in 76 matched comparison districts using the CITS design.<sup>35</sup> The CITS sensitivity analysis has less statistical power to detect changes in performance than the CITS confirmatory analysis due to the smaller number of comparison districts included in the sensitivity analysis.

To estimate individual cohort and year effects, we estimated the following CITS model:

$$Y_{itjk} = \beta_0 + \beta_1 TLC_k + \beta_2 Post_t + \beta_3 TLC_k Post_t + \beta_4 Time_t + \beta_5 TLC_k Time_t + Grade_{itjk} + AEA_k + X_{itjk} + z_k + v_{jk} + u_{tjk} + e_{itjk}$$

where  $TLC_k Post_t$  represents the posttreatment effect of interest. The model was run separately for the following comparisons:

1. Cohort 1 in Year 1: 2014–15 effect for Cohort 1 using matches from Cohorts 2 and 3;
2. Cohort 2 in Year 1: 2015–16 effect for Cohort 2 using matches from Cohort 3 only; and
3. Cohort 1 in Year 2: 2015–16 effect for Cohort 1 using matches from Cohort 3 only.

Similar to the CITS confirmatory individual cohort analyses, we estimated a pooled Year 1 effect by combining the postintervention treatment effects ( $TLC_k Post_t$ ) for Cohort 1 Year 1 and Cohort 2 Year 1 using meta-analysis. The individual and pooled postintervention treatment effects then were plotted on a forest plot to show how much treatment effects varied between the cohorts and years.

### Constructing Matched Comparison Groups

Propensity score matching was used to identify each of these comparison groups. Propensity score matching is a statistical technique that estimates the predicted probability of group membership (treatment versus control) based on observed characteristics, and then uses that predicted probability to create a control group similar to the treatment group on these observed

<sup>35</sup> Only overall CITS models were estimated using the matched comparison groups. The matching process included matching on numerous student- and district-level characteristics; thus, it may not be appropriate to estimate subgroup treatment effects on the matched comparison groups, as they may no longer be matched within the subgroups.

characteristics. We included average enrollment, demographic composition, and student achievement data (aggregated to the district level) in the years prior to TLC implementation in the matching process. Specifically, we used the following variables: total student enrollment in the district, percentage of White students, percentage of students eligible for FRPL, percentage of ELLs, percentage of students with an IEP, and average student achievement on the mathematics and reading state assessments.<sup>36</sup>

## **Matching Results**

The results of the matching procedures are presented in Tables C1 through C3. Each table presents the mean enrollment, demographic, and achievement characteristics for TLC districts for all available comparisons and each cohort's matched comparison for each year of program implementation. The matched samples of TLC districts were similar to non-TLC districts in their demographic composition, proficiency, and academic achievement in the years prior to implementing TLC. Matching helped reduce differences in demographic composition and achievement. The matched samples differed in average enrollment, where TLC districts were larger on average, but matching helped reduce this difference in average enrollment between TLC and non-TLC districts. The matching methods were most useful for reducing differences between the groups when there is a larger pool of comparison districts. The usefulness of the matching process diminished as the pool of available comparisons reduced from Cohort 2 to Cohort 3.

The resulting matched sample districts are listed in Tables C4, C5, and C6 for Cohort 1 Year 1, Cohort 1 Year 2, and Cohort 2 Year 1, respectively, with their corresponding propensity score. Propensity scores are the predicted probabilities of group membership. Thus, the values in the tables represent the probability that each district is included in the corresponding group (TLC cohort or comparison group) based on the observed characteristics included in the model defined above. The districts are presented such that the districts with higher propensity scores, and thus higher probabilities of group membership, are included on the top of the table.

<sup>36</sup> To implement propensity score matching, we ran the following logistic regression model:

$$\text{logit}(P(TLC)) = \eta + \sum \lambda Z + A,$$

where  $TLC$  is an indicator of whether a district is a member of the TLC cohort in a given year (for Cohort 1 Year 1,  $TLC = 1$  if a district started implementing TLC in 2014–15,  $TLC = 0$  for Cohorts 2 and 3; for Cohort 1 Year 2,  $TLC = 1$  if a district was implementing TLC for a second year in 2015–16,  $TLC = 0$  for Cohort 3; for Cohort 2 Year 1,  $TLC = 1$  if a district started implementing TLC in 2015–16,  $TLC = 0$  for Cohort 3),  $P(TLC)$  is the propensity of a district being a member of the given TLC cohort,  $\eta$  is an intercept,  $Z$  is a set of district-level characteristics for all nine or ten school years (i.e., 2005–06 to 2013–14 for Cohort 1 and 2005–06 to 2014–15 for Cohort 2; number of students enrolled in the district, percentage of White students, percentage of students who are eligible for FRPL, percentage of students who are ELLs, percentage of students with IEPs, and average student achievement on the mathematics and reading state assessments),  $\lambda$  is a set of coefficients that represents the association between each district characteristic and the logit of the propensity score, and  $A$  represents the AEA fixed effects which allow for matching within AEAs. We used an optimal matching algorithm (Hansen, 2004) for each match. This approach minimizes the distances in TLC and non-TLC propensity score distributions and retains all TLC districts.

**Table C1. Matched Comparisons Results by School Year, TLC Cohort 1, Year 1, 2014–15**

Year	Characteristic	Full Sample			Matched Sample		
		TLC C1 Mean	All Other Mean	Mean Diff	TLC C1 Mean	All Other Mean	Mean Diff
2006	Enrollment	3996	1103	2893	3996	2486	1510
	White	89%	94%	-5%	89%	93%	-4%
	FRPL	27%	29%	-2%	27%	25%	2%
	ELL	3%	2%	1%	3%	2%	1%
	IEP	12%	12%	0%	12%	12%	0%
	Mathematics scale score	248	248	0	248	249	-1
	Reading scale score	245	245	0	245	246	-1
2007	Enrollment	4020	1100	2920	4020	2509	1511
	White	89%	94%	-5%	89%	93%	-4%
	FRPL	28%	30%	-2%	28%	26%	3%
	ELL	3%	2%	1%	3%	2%	1%
	IEP	12%	13%	0%	12%	12%	0%
	Mathematics scale score	248	249	-2	248	250	-2
	Reading scale score	245	246	-1	245	248	-3
2008	Enrollment	4012	1095	2917	4012	2504	1507
	White	88%	93%	-5%	88%	92%	-4%
	FRPL	29%	30%	-1%	29%	27%	2%
	ELL	3%	2%	1%	3%	2%	1%
	IEP	12%	13%	-1%	12%	12%	0%
	Mathematics scale score	248	248	-1	248	249	-2
	Reading scale score	245	245	0	245	247	-2
2009	Enrollment	4079	1108	2971	4079	2544	1535
	White	87%	93%	-6%	87%	91%	-4%
	FRPL	31%	32%	-1%	31%	27%	3%
	ELL	3%	2%	1%	3%	2%	1%
	IEP	12%	12%	-1%	12%	11%	0%
	Mathematics scale score	248	249	-1	248	249	-1
	Reading scale score	246	246	0	246	247	-1
2010	Enrollment	4143	1110	3033	4143	2596	1547
	White	85%	92%	-6%	85%	90%	-4%
	FRPL	33%	34%	-1%	33%	30%	3%
	ELL	4%	2%	2%	4%	2%	1%
	IEP	12%	12%	0%	12%	11%	0%
	Mathematics scale score	249	248	1	249	250	-1
	Reading scale score	246	245	1	246	247	-1
2011	Enrollment	4176	1126	3050	4176	2627	1549
	White	85%	91%	-7%	85%	89%	-4%
	FRPL	36%	36%	0%	36%	32%	3%
	ELL	4%	2%	2%	4%	2%	1%
	IEP	12%	12%	0%	12%	11%	0%
	Mathematics scale score	248	248	0	248	250	-1
	Reading scale score	246	246	0	246	248	-2

Year	Characteristic	Full Sample			Matched Sample		
		TLC C1 Mean	All Other Mean	Mean Diff	TLC C1 Mean	All Other Mean	Mean Diff
2012	Enrollment	4166	1121	3046	4166	2635	1531
	White	84%	91%	-7%	84%	89%	-5%
	FRPL	36%	37%	-1%	36%	33%	4%
	ELL	4%	2%	2%	4%	2%	2%
	IEP	12%	12%	0%	12%	11%	1%
	Mathematics scale score	240	240	0	240	243	-3
	Reading scale score	239	239	0	239	242	-3
2013	Enrollment	4191	1133	3058	4191	2684	1507
	White	83%	90%	-7%	83%	88%	-5%
	FRPL	37%	38%	0%	37%	34%	3%
	ELL	4%	2%	2%	4%	2%	1%
	IEP	12%	12%	0%	12%	11%	1%
	Mathematics scale score	242	241	1	242	243	-1
	Reading scale score	241	240	1	241	243	-2
2014	Enrollment	4483	1158	3325	4483	2767	1715
	White	82%	90%	-8%	82%	87%	-5%
	FRPL	37%	38%	-1%	37%	34%	3%
	ELL	4%	2%	2%	4%	2%	2%
	IEP	11%	12%	-1%	11%	11%	0%
	Mathematics scale score	247	245	2	247	248	-1
	Reading scale score	247	246	1	247	249	-2
2015	Enrollment	4500	1164	3335	4500	2792	1708
	White	82%	90%	-8%	82%	87%	-5%
	FRPL	36%	38%	-2%	36%	34%	2%
	ELL	5%	2%	2%	5%	3%	2%
	IEP	11%	11%	-1%	11%	11%	0%
	Mathematics scale score	246	245	1	246	248	-2
	Reading scale score	248	248	0	248	251	-3

Note. The full sample includes 39 TLC Cohort 1 districts and 297 non-TLC districts. The optimal matching sample includes 39 TLC Cohort 1 districts and 39 non-TLC districts.

**Table C2. Matched Comparisons Results by School Year, TLC Cohort 1, Year 2, 2015–16**

Year	Characteristic	Full Sample			Matched Sample		
		TLC C1 Mean	TLC C3 Mean	Mean Diff	TLC C1 Mean	TLC C3 Mean	Mean Diff
2006	Enrollment	3996	793	3202	3996	988	3007
	White	89%	94%	-5%	89%	93%	-3%
	FRPL	27%	30%	-3%	27%	29%	-2%
	ELL	3%	2%	1%	3%	3%	0%
	IEP	12%	12%	0%	12%	13%	-1%
	Mathematics scale score	248	248	0	248	250	-2
	Reading scale score	245	245	0	245	247	-2

Year	Characteristic	Full Sample			Matched Sample		
		TLC C1 Mean	TLC C3 Mean	Mean Diff	TLC C1 Mean	TLC C3 Mean	Mean Diff
2007	Enrollment	4020	784	3235	4020	984	3036
	White	89%	94%	-5%	89%	92%	-4%
	FRPL	28%	31%	-2%	28%	30%	-2%
	ELL	3%	2%	1%	3%	3%	0%
	IEP	12%	13%	0%	12%	13%	-1%
	Mathematics scale score	248	249	-1	248	250	-3
	Reading scale score	245	246	0	245	248	-2
2008	Enrollment	4012	777	3235	4012	989	3023
	White	88%	94%	-5%	88%	92%	-4%
	FRPL	29%	31%	-2%	29%	31%	-1%
	ELL	3%	2%	1%	3%	3%	0%
	IEP	12%	13%	-1%	12%	13%	-1%
	Mathematics scale score	248	248	0	248	249	-2
	Reading scale score	245	245	1	245	247	-2
2009	Enrollment	4079	782	3298	4079	1000	3079
	White	87%	93%	-6%	87%	92%	-4%
	FRPL	31%	32%	-2%	31%	32%	-1%
	ELL	3%	2%	1%	3%	3%	0%
	IEP	12%	12%	-1%	12%	12%	-1%
	Mathematics scale score	248	248	0	248	250	-2
	Reading scale score	246	245	0	246	247	-1
2010	Enrollment	4143	772	3370	4143	1004	3139
	White	85%	92%	-6%	85%	90%	-5%
	FRPL	33%	35%	-2%	33%	35%	-2%
	ELL	4%	2%	2%	4%	3%	1%
	IEP	12%	12%	-1%	12%	12%	-1%
	Mathematics scale score	249	247	2	249	250	-1
	Reading scale score	246	244	2	246	247	-1
2011	Enrollment	4176	781	3394	4176	1006	3170
	White	85%	91%	-7%	85%	90%	-5%
	FRPL	36%	37%	-1%	36%	35%	0%
	ELL	4%	2%	2%	4%	3%	1%
	IEP	12%	12%	0%	12%	12%	0%
	Mathematics scale score	248	248	0	248	249	-1
	Reading scale score	246	245	1	246	247	-1
2012	Enrollment	4166	772	3395	4166	1005	3162
	White	84%	91%	-7%	84%	89%	-6%
	FRPL	36%	38%	-2%	36%	37%	0%
	ELL	4%	2%	2%	4%	3%	1%
	IEP	12%	12%	0%	12%	12%	0%
	Mathematics scale score	240	240	0	240	241	-1
	Reading scale score	239	239	0	239	240	-1

Year	Characteristic	Full Sample			Matched Sample		
		TLC C1 Mean	TLC C3 Mean	Mean Diff	TLC C1 Mean	TLC C3 Mean	Mean Diff
2013	Enrollment	4191	774	3417	4191	1007	3184
	White	83%	91%	-8%	83%	89%	-6%
	FRPL	37%	38%	-1%	37%	37%	0%
	ELL	4%	2%	2%	4%	3%	1%
	IEP	12%	12%	0%	12%	12%	0%
	Mathematics scale score	242	241	1	242	242	0
	Reading scale score	241	240	1	241	241	-1
2014	Enrollment	4483	789	3693	4483	1034	3448
	White	82%	90%	-8%	82%	88%	-6%
	FRPL	37%	39%	-2%	37%	36%	1%
	ELL	4%	2%	2%	4%	3%	1%
	IEP	11%	12%	-1%	11%	11%	0%
	Mathematics scale score	247	244	2	247	248	-1
	Reading scale score	247	246	2	247	248	-1
2015	Enrollment	4500	789	3711	4500	1042	3458
	White	82%	90%	-8%	82%	88%	-6%
	FRPL	36%	39%	-2%	36%	37%	-1%
	ELL	5%	2%	2%	5%	3%	1%
	IEP	11%	11%	-1%	11%	11%	0%
	Mathematics scale score	246	245	2	246	248	-1
	Reading scale score	248	247	1	248	250	-2

Note. The full sample includes 39 TLC Cohort 1 districts and 221 Cohort 3 districts. The optimal matching sample includes 39 TLC Cohort 1 districts and 39 Cohort 3 districts.

**Table C3. Matched Comparisons Results by School Year, TLC Cohort 2, Year 1, 2015–16**

Year	Characteristic	Full Sample			Matched Sample		
		TLC C2 Mean	TLC C3 Mean	Mean Diff	TLC C2 Mean	TLC C3 Mean	Mean Diff
2006	Enrollment	2002	793	1208	2002	943	1058
	White	94%	94%	0%	94%	95%	-1%
	FRPL	27%	30%	-2%	27%	29%	-1%
	ELL	1%	2%	0%	1%	1%	0%
	IEP	12%	12%	-1%	12%	12%	-1%
	Mathematics scale score	249	248	1	249	248	1
	Reading scale score	246	245	1	246	245	1
2007	Enrollment	2016	784	1232	2016	938	1079
	White	94%	94%	0%	94%	94%	-1%
	FRPL	28%	31%	-2%	28%	30%	-2%
	ELL	1%	2%	-1%	1%	1%	0%
	IEP	13%	13%	0%	13%	13%	0%
	Mathematics scale score	250	249	1	250	249	1
	Reading scale score	247	246	2	247	247	1

Year	Characteristic	Full Sample			Matched Sample		
		TLC C2 Mean	TLC C3 Mean	Mean Diff	TLC C2 Mean	TLC C3 Mean	Mean Diff
2008	Enrollment	2019	777	1242	2019	934	1085
	White	93%	94%	0%	93%	94%	-1%
	FRPL	29%	31%	-2%	29%	30%	-1%
	ELL	2%	2%	0%	2%	1%	0%
	IEP	12%	13%	0%	12%	12%	0%
	Mathematics scale score	250	248	2	250	248	1
	Reading scale score	246	245	1	246	245	1
2009	Enrollment	2059	782	1277	2059	944	1115
	White	93%	93%	0%	93%	94%	-1%
	FRPL	30%	32%	-3%	30%	31%	-2%
	ELL	2%	2%	0%	2%	2%	0%
	IEP	12%	12%	0%	12%	12%	0%
	Mathematics scale score	250	248	2	250	249	1
	Reading scale score	247	245	2	247	246	1
2010	Enrollment	2092	772	1320	2092	942	1151
	White	91%	92%	-1%	91%	92%	-1%
	FRPL	32%	35%	-3%	32%	35%	-2%
	ELL	2%	2%	0%	2%	2%	0%
	IEP	12%	12%	0%	12%	12%	0%
	Mathematics scale score	250	247	2	250	248	2
	Reading scale score	247	244	2	247	245	1
2011	Enrollment	2127	781	1345	2127	957	1170
	White	91%	91%	-1%	91%	92%	-1%
	FRPL	34%	37%	-3%	34%	36%	-2%
	ELL	2%	2%	0%	2%	2%	0%
	IEP	12%	12%	0%	12%	12%	0%
	Mathematics scale score	249	248	1	249	249	1
	Reading scale score	247	245	1	247	246	1
2012	Enrollment	2136	772	1365	2136	948	1189
	White	90%	91%	-1%	90%	91%	-1%
	FRPL	35%	38%	-3%	35%	37%	-2%
	ELL	2%	2%	0%	2%	2%	0%
	IEP	12%	12%	0%	12%	12%	0%
	Mathematics scale score	241	240	1	241	240	1
	Reading scale score	240	239	1	240	239	1
2013	Enrollment	2176	774	1402	2176	959	1217
	White	90%	91%	-1%	90%	91%	-1%
	FRPL	36%	38%	-3%	36%	37%	-2%
	ELL	2%	2%	0%	2%	2%	0%
	IEP	12%	12%	0%	12%	12%	0%
	Mathematics scale score	241	241	1	241	240	1
	Reading scale score	241	240	2	241	240	2



Year	Characteristic	Full Sample			Matched Sample		
		TLC C2 Mean	TLC C3 Mean	Mean Diff	TLC C2 Mean	TLC C3 Mean	Mean Diff
2014	Enrollment	2229	789	1440	2229	977	1252
	White	89%	90%	-1%	89%	91%	-1%
	FRPL	36%	39%	-3%	36%	38%	-2%
	ELL	2%	2%	0%	2%	2%	0%
	IEP	12%	12%	0%	12%	12%	0%
	Mathematics scale score	245	244	1	245	244	1
	Reading scale score	248	246	2	248	246	1
2015	Enrollment	2256	789	1467	2256	984	1272
	White	89%	90%	-1%	89%	90%	-1%
	FRPL	36%	39%	-3%	36%	38%	-2%
	ELL	2%	2%	0%	2%	2%	0%
	IEP	11%	11%	0%	11%	11%	0%
	Mathematics scale score	246	245	1	246	244	2
	Reading scale score	249	247	2	249	247	2

Note. The full sample includes 76 TLC Cohort 2 districts and 221 Cohort 3 districts. The optimal matching sample includes 76 TLC Cohort 2 districts and 76 Cohort 3 districts.

**Table C4. Matched Districts, TLC Cohort 1, Year 1, 2014–15**

TLC Cohort 1			Matched Districts From Comparison Pool of All Other Districts		
District	District Name	Propensity	District	District Name	Propensity
1863	Dubuque	1.00	3330	Keota	0.70
1053	Cedar Rapids	1.00	4725	Newton	0.69
3231	Johnston	1.00	1337	College	0.67
1611	Davenport	1.00	3141	Iowa City	0.61
1953	Earlham	1.00	5283	Pocahontas Area	0.58
5607	Rock Valley	1.00	225	Ames	0.54
5643	Roland-Story	1.00	1278	Clinton	0.52
6957	West Des Moines	1.00	2673	Nodaway Valley	0.42
3195	Greene County	0.99	4617	Nevada	0.40
5049	Ottumwa	0.99	1737	Des Moines Independent	0.37
6615	Van Meter	0.99	4149	MOC-Floyd Valley	0.36
6961	Western Dubuque	0.97	2727	Grundy Center	0.36
4104	Marshalltown	0.92	4131	Mason City	0.36
6516	Twin Rivers	0.92	5250	Pleasant Valley	0.35
5166	Pella	0.90	6579	Urbandale	0.33
6101	Southeast Polk	0.84	1332	Colfax-Mingo	0.31
4536	Mount Pleasant	0.83	540	BCLUW	0.28
5805	Saydel	0.80	5256	Pleasantville	0.27
1675	Delwood	0.78	63	Akron Westfield	0.23

TLC Cohort 1			Matched Districts From Comparison Pool of All Other Districts		
District	District Name	Propensity	District	District Name	Propensity
2466	Gilbert	0.74	4788	Northwood-Kensett	0.23
7056	Winterset	0.71	6992	Westwood	0.22
6795	Waterloo	0.63	2113	Essex	0.22
1476	Council Bluffs	0.63	2376	Galva-Holstein	0.21
1968	East Marshall	0.62	2766	H-L-V	0.21
4581	Muscatine	0.60	2988	Hinton	0.21
621	Bettendorf	0.57	6534	Underwood	0.20
4869	Oelwein	0.51	4978	Orient-Macksburg	0.19
882	Burlington	0.44	6943	West Central	0.18
4779	North Polk	0.40	2520	Glidden-Ralston	0.16
6039	Sioux City	0.40	729	Boone	0.16
4797	Norwalk	0.39	6138	Springville	0.14
5121	Panorama	0.35	6762	Wapsie Valley	0.12
3042	Hudson	0.28	2124	Estherville Lincoln	0.12
3715	Linn-Mar	0.24	5508	Riceville	0.12
1359	Colo-NESCO School	0.23	981	Carlisle	0.11
3600	Le Mars	0.21	2205	Farragut	0.10
1970	East Union	0.18	5877	Sergeant Bluff-Luton	0.10
3060	Humboldt	0.12	4644	Newell-Fonda	0.09
609	Benton	0.00	6822	Waukee	0.00

**Table C5. Matched Districts, TLC Cohort 1, Year 2, 2015–16**

TLC Cohort 1			Matched Districts From Comparison Pool of TLC Cohort 3 Districts		
District	District Name	Propensity	District	District Name	Propensity
1863	Dubuque	1.00	4122	Martensdale-St Marys	0.89
3231	Johnston	1.00	4725	Newton	0.84
3715	Linn-Mar	1.00	27	Adel DeSoto Minburn	0.44
1611	Davenport	1.00	594	Belmond-Klemme	0.39
1053	Cedar Rapids	1.00	2520	Glidden-Ralston	0.39
6795	Waterloo	1.00	3105	Independence	0.36
6039	Sioux City	1.00	126	Algona	0.36
6101	Southeast Polk	1.00	747	Boyden-Hull	0.31
6957	West Des Moines	1.00	6094	Southeast Warren	0.25
621	Bettendorf	1.00	2727	Grundy Center	0.23
1476	Council Bluffs	1.00	1233	Clear Lake	0.23
4581	Muscatine	1.00	1602	Danville	0.23
6961	Western Dubuque	0.99	6943	West Central	0.23

TLC Cohort 1			Matched Districts From Comparison Pool of TLC Cohort 3 Districts		
District	District Name	Propensity	District	District Name	Propensity
5166	Pella	0.99	5256	Pleasantville	0.23
5049	Ottumwa	0.99	6175	Starmont	0.22
2466	Gilbert	0.98	5949	Sheldon	0.22
882	Burlington	0.96	5184	Perry	0.21
4104	Marshalltown	0.95	7002	Whiting	0.19
4779	North Polk	0.91	6219	Storm Lake	0.19
7056	Winterset	0.87	4527	Mount Ayr	0.17
3195	Greene County	0.87	6534	Underwood	0.17
1675	Delwood	0.87	3906	Lynnvile-Sully	0.15
5805	Saydel	0.86	6100	South Winneshiek	0.14
3060	Humboldt	0.85	1071	Centerville	0.14
4797	Norwalk	0.74	4033	Maple Valley-Anthon Oto	0.14
5643	Roland-Story	0.57	2754	Guthrie Center	0.14
609	Benton	0.53	4203	Mediapolis	0.14
5607	Rock Valley	0.49	4271	Mid-Prairie	0.14
6516	Twin Rivers	0.48	6990	West Sioux	0.13
1953	Earlham	0.48	1044	Cedar Falls	0.13
6615	Van Meter	0.38	6937	West Burlington Ind	0.12
5121	Panorama	0.33	1965	Easton Valley	0.12
1968	East Marshall	0.30	4905	Olin Consolidated	0.12
3600	Le Mars	0.28	2205	Farragut	0.11
4536	Mount Pleasant	0.17	4086	Marion Independent	0.11
4869	Oelwein	0.16	2763	Clayton Ridge	0.10
1359	Colo-NESCO School	0.12	2846	Harris-Lake Park	0.10
3042	Hudson	0.08	3330	Keota	0.07
1970	East Union	0.00	6921	West Bend-Mallard	0.00

**Table C6. Matched Districts, TLC Cohort 2, Year 1, 2015–16**

TLC Cohort 2			Matched Districts From Comparison Pool of TLC Cohort 3 Districts		
District	District Name	Propensity	District	District Name	Propensity
1737	Des Moines Independent	1.00	4086	Marion Independent	0.89
261	Ankeny	1.00	1044	Cedar Falls	0.84
1337	College	1.00	3744	Lisbon	0.78
6822	Waukee	1.00	4725	Newton	0.76
6579	Urbandale	1.00	1503	Creston	0.76
5250	Pleasant Valley	1.00	1965	Easton Valley	0.74
4784	North Scott	1.00	5256	Pleasantville	0.73

TLC Cohort 2			Matched Districts From Comparison Pool of TLC Cohort 3 Districts		
District	District Name	Propensity	District	District Name	Propensity
5013	Oskaloosa	1.00	2834	Harmony	0.68
3114	Indianola	0.99	549	Bedford	0.65
4774	North Fayette	0.97	4271	Mid-Prairie	0.65
225	Ames	0.97	4554	Mount Vernon	0.61
1576	Dallas Center-Grimes	0.94	4203	Mediapolis	0.56
1278	Clinton	0.93	135	Allamakee	0.55
1619	Davis County	0.92	6096	Southeast Webster Grand	0.51
6462	Tri-County	0.92	1080	Central	0.49
6854	Wayne	0.91	2709	Grinnell-Newburg	0.48
3141	Iowa City	0.90	4662	New Hampton	0.45
1972	Eastern Allamakee	0.90	2862	Hartley-Melvin-Sanborn	0.43
1107	Chariton	0.89	2493	Gilmore City-Bradgate	0.43
3150	Iowa Falls	0.88	126	Algona	0.43
2511	Glenwood	0.88	6768	Washington	0.42
1221	Clear Creek Amana	0.87	4491	Moravia	0.42
916	CAL	0.84	472	Ballard	0.41
3645	Lewis Central	0.82	6094	Southeast Warren	0.41
81	Albia	0.82	576	Belle Plaine	0.40
1062	Center Point-Urbana	0.81	6120	Spirit Lake	0.39
6660	Vinton-Shellsburg	0.80	4787	North Winneshiek	0.38
4777	North Linn	0.76	3816	Lone Tree	0.38
4773	Northeast	0.75	234	Anamosa	0.37
729	Boone	0.74	6099	South O'Brien	0.36
2313	Fort Dodge	0.73	1926	Durant	0.35
4043	Maquoketa Valley	0.72	423	Aurelia	0.35
2169	Fairfield	0.72	1917	Boyer Valley	0.34
6591	Valley	0.71	6175	Starmont	0.34
6759	Wapello	0.69	5163	Pekin	0.34
7029	Williamsburg	0.68	5949	Sheldon	0.31
513	Baxter	0.68	5508	Riceville	0.31
999	Carroll	0.67	27	Adel DeSoto Minburn	0.29
981	Carlisle	0.66	4778	North Kossuth	0.28
6246	Stratford	0.64	6867	Webster City	0.28
5463	Red Oak	0.63	2846	Harris-Lake Park	0.27
1638	Decorah Community	0.62	6985	West Marshall	0.25
2988	Hinton	0.57	6408	Tipton	0.25
1082	Central DeWitt	0.53	5877	Sergeant Bluff-Luton	0.25

TLC Cohort 2			Matched Districts From Comparison Pool of TLC Cohort 3 Districts		
District	District Name	Propensity	District	District Name	Propensity
414	Audubon	0.49	2124	Estherville Lincoln	0.24
243	Andrew	0.49	2763	Clayton Ridge	0.24
7038	Wilton	0.47	5832	Schleswig	0.23
2088	Emmetsburg	0.45	657	Eddyville-Blakesburg-	0.23
4890	Okoboji	0.43	3029	Howard-Winneshiek	0.23
3691	North Cedar	0.42	5328	Prescott	0.22
1350	Collins-Maxwell	0.42	99	Alburnett	0.22
6840	Waverly-Shell Rock	0.40	3330	Keota	0.21
108	Alden	0.39	4776	North Mahaska	0.21
6561	United	0.39	7110	Woodward-Granger	0.20
6093	Solon	0.39	5184	Perry	0.20
7047	Winfield-Mt Union	0.38	4527	Mount Ayr	0.19
4419	MFL MarMac	0.36	1079	Central Lee	0.19
6930	West Branch	0.35	918	Calamus-Wheatland	0.18
1989	Edgewood-Colesburg	0.32	2403	Garner-Hayfield-Ventura	0.18
4446	Monticello	0.31	1368	Columbus	0.17
4617	Nevada	0.29	6012	Sigourney	0.16
387	Atlantic	0.29	3119	Interstate 35	0.16
2781	Hampton-Dumont	0.25	3375	Knoxville	0.16
6950	West Delaware County	0.24	4572	Murray	0.16
1093	Central Decatur	0.22	936	Camanche	0.16
1218	Clay Central-Everly	0.21	4599	Nashua-Plainfield	0.15
3154	Iowa Valley	0.20	6937	West Burlington Ind	0.15
6138	Springville	0.19	4068	Marcus-Meriden-Cleghorn	0.13
4149	MOC-Floyd Valley	0.19	6030	Sioux Center	0.13
6987	West Monona	0.18	6762	Wapsie Valley	0.12
4689	New London	0.18	7002	Whiting	0.12
6102	Spencer	0.16	4995	Osage	0.12
1095	Central Lyon	0.15	720	Bondurant-Farrar	0.11
6536	Union	0.10	914	CAM	0.10
977	Cardinal	0.09	6592	Van Buren	0.08
2673	Nodaway Valley	0.05	5895	Seymour	0.05

## Regression Discontinuity Design Sensitivity Analysis

To estimate student achievement outcomes in the first year of program implementation, we calculated within-year TLC Cohort 1 effects and within-year TLC Cohort 2 effects using an RD design analysis. This analysis examines differences in outcomes of students in TLC districts that were

eligible and received funding to students in noneligible districts based on their application scores, where eligibility was based on a cutoff score of 73 for Cohort 1 and 78 for Cohort 2. The RD sensitivity analysis compares students in districts just below and just above the application cutoff score, producing a local average treatment effect for those schools.<sup>37</sup> Because of this restriction and because the approach does not control for preintervention performance, the RD analysis has less statistical power to detect changes in performance than both the CITS confirmatory analysis and the CITS sensitivity analysis.

A two-stage least-squares (2SLS) modeling approach was conducted separately for Cohort 1 (2014–15) and Cohort 2 (2015–16). The first stage of this model is represented by the following equation:

$$TLC_k = \beta_0 + \beta_1 Eligible_k + \beta_2 Distance_k + \beta_3 Eligible_k Distance_k + Tier_k + Grade_{ik} + X_{ik}$$

where  $TLC_k$  is an indicator of whether a district  $k$  is a member of the given TLC cohort (Cohort 1 in the 2014–15 school year and Cohort 2 in the 2015–16 school year);  $Eligible_k$  is an indicator of whether a district  $k$  became eligible to be a member of the given TLC cohort based on the application score (i.e., obtained a score greater than or equal to 73 for Cohort 1 and a score greater than or equal to 78 for Cohort 2);  $Distance_k$  is the difference between district  $k$ 's application score and the cutoff score;  $Tier_k$  represents district tier size fixed effects;  $Grade_{ik}$  represents grade fixed effects; and  $X_{ik}$  is a set of student-level characteristics (i.e., gender, race, ELL status, FRPL status, and IEP status). We accounted for the nesting of students in districts by calculating robust standard errors at the district level  $k$ .

The second stage of this model is represented by the following equation:

$$Y_{ik} = \beta_0 + \beta_1 TLC_k + \beta_2 Distance_k + \beta_3 Eligible_k Distance_k + Tier_k + Grade_{ik} + X_{ik}$$

where  $Y_{ik}$  is the outcome measure (i.e., the 2014–15 standardized scaled achievement score for reading or mathematics for Cohort 1 or the 2015–16 standardized scaled achievement score for Cohort 2) for a student  $i$  in district  $k$ ;  $TLC_k$  is the predicted probability of TLC membership for the cohort of interest, estimated in the Stage 1 equation; and the remaining variables are the same as the first stage equation described.

Similar to the CITS analyses, to estimate a pooled Year 1 effect, we combined the treatment effects ( $TLC_k$ ) for Cohort 1 and Cohort 2 using meta-analysis. Because Cohort 2 includes a larger sample, the effect from Cohort 2 is weighted more heavily in the pooled effect.

<sup>37</sup> Restricting TLC and comparison districts to those just below and just above the application cutoff score produces a local average treatment effect. For Cohort 1, the analysis was restricted to districts within seven points of the application cutoff score of 73 (i.e., between 66 and 80 points), resulting in 23 TLC Cohort 1 districts and 44 noneligible comparison districts. For Cohort 2, the analysis was restricted to districts within eight points of the application cutoff score of 78 (i.e., between 70 and 86 points), resulting in 73 TLC Cohort 2 districts and 73 noneligible comparison districts.

## Appendix D. TLC Impacts

The comparative interrupted time series (CITS) confirmatory analysis impact estimates for the Teacher Leadership and Compensation (TLC) program and student, school, and district subgroups are presented in Table D1. The estimates represent the effects of TLC in the first and second years of program implementation. Specifically, the estimates represent the change in the outcomes of students in TLC-implementing districts in the first and second years of program implementation compared with changes in outcomes of students in the comparison districts.<sup>38</sup>

**Table D1. TLC Impact Estimates and Subgroup Effects**

Subgroup	Reading		Mathematics	
	Year 1	Year 2	Year 1	Year 2
<b>All Students</b>				
Overall TLC impact	-0.02* (0.01)	-0.03** (0.01)	-0.02** (0.01)	-0.04** (0.01)
<b>Cohorts</b>				
Cohort 1	-0.04** (0.01)	-0.02 (0.01)	-0.05** (0.01)	-0.03** (0.01)
Cohort 2	0.01 (0.01)	-	0.01 (0.01)	-
Cohorts 1 and 2 Combined	-0.01* (0.01)	-	-0.02** (0.01)	-
<b>Special Populations</b>				
ELL students versus non-ELL students	0.07** (0.02)	0.07** (0.02)	0.07** (0.02)	0.07** (0.02)
FRPL students versus non-FRPL students	0.01 (0.01)	0.02* (0.01)	0.02** (0.01)	0.04** (0.01)
IEP students versus non-IEP students	0.01 (0.01)	0.02 (0.01)	0.02* (0.01)	0.05** (0.01)
<b>NYCLA (2015–16)</b>				
TLC versus non-TLC, controlling for NYCLA		-0.02* (0.01)		-0.03** (0.01)
TLC + NYCLA versus non-TLC		-0.02 (0.02)		-0.06* (0.02)
<b>District Size Tiers</b>				
2,500 to 8,999 students	-0.03* (0.01)	-0.09** (0.02)	-0.08** (0.02)	-0.16** (0.03)
1,000 to 2,499 students	0.00 (0.02)	0.04 (0.02)	0.01 (0.02)	0.04 (0.03)

<sup>38</sup> As noted in Appendix C, all models control for student-level demographics (i.e., gender, race, ELL status, FRPL status, and IEP status), grade and AEA fixed effects, and linear outcome trends across time. Due to space constraints, we only present the postintervention treatment impact estimates.

Subgroup	Reading		Mathematics	
	Year 1	Year 2	Year 1	Year 2
600 to 999 students	-0.01 (0.02)	0.05 (0.03)	0.02 (0.02)	0.19** (0.04)
300 to 599 students	0.02 (0.03)	0.04 (0.06)	-0.01 (0.03)	-0.01 (0.07)
Fewer than 300 students	-0.03 (0.05)	-0.06 (0.12)	0.07 (0.06)	0.60** (0.14)
<b>Grade Bands</b>				
Grades 3–5	-0.02* (0.01)	-0.06** (0.01)	-0.03** (0.01)	-0.07** (0.02)
Grades 6–8	0.00 (0.01)	0.02 (0.02)	0.01 (0.01)	0.03 (0.02)
Grades 10–11	-0.01 (0.02)	0.01 (0.02)	-0.02 (0.02)	0.00 (0.03)

Note. Unless the TLC cohort is specified, Year 1 effects include Cohorts 1 and 2 in the treatment group, and Year 2 effects include only Cohort 1 in the treatment group. Effects for Tier 1 size district (districts with 9,000 or more students) were not estimated because all Tier 1 districts were implementing TLC by Year 2 (i.e., all Tier 1 districts are members of Cohorts 1 and 2, thus Tier 1 does not have a comparison group). NYCLA was implemented in 2015–16 for both Cohorts 1 and 2, and thus represent overall NYCLA effects for both cohorts combined.

Standard errors are presented in parentheses.

\* $p \leq .05$ ; \*\* $p \leq .01$ .

Table D2 presents the CITS sensitivity analysis impact estimates for TLC Cohort 1 and 2 outcomes for student achievement. Again, these estimates represent the effects of TLC in the first and second years of program implementation. However, instead of including all districts, students in TLC-implementing districts were compared to students in matched comparison districts.

**Table D2. Sensitivity Analysis, TLC Impacts**

Subgroup	Reading	Mathematics
Cohort 1 Year 1 versus matched comparison districts	-0.04** (0.01)	-0.05** (0.02)
Cohort 2 Year 1 versus matched comparison districts	0.01 (0.01)	0.01 (0.01)
Combined Year 1 effect	-0.01 (0.01)	-0.02 (0.01)
Cohort 1 Year 2 versus matched comparison districts	-0.03* (0.02)	-0.05** (0.02)

Note. Standard errors are presented in parentheses.

\* $p \leq .05$ ; \*\* $p \leq .01$ .

Table D3 presents the regression discontinuity (RD) analysis impact estimates for TLC Cohort 1 and 2 outcomes for student achievement in the first year of program implementation. Note that RD provides a local average treatment effect, using data from districts just below and just above the cutoff. Because of this feature, the effects are somewhat different than effects in the CITS



framework, which estimates an average treatment effect for all districts. In addition, due to the smaller sample size, the power to detect an effect is much lower with the RD design than with the CITS design.

**Table D3. RD Outcomes for Reading and Mathematics Student Achievement: TLC Eligible and Funded Districts and Noneligible Districts**

Subgroup	Reading	Mathematics
Cohort 1 Year 1 versus noneligible districts	-0.05 (0.38)	0.01 (0.57)
Cohort 2 Year 1 versus noneligible districts	0.07 (0.07)	0.10 (0.09)
Combined Year 1 versus noneligible districts	0.07 (0.07)	0.10 (0.09)

Note. Robust standard errors are presented in parentheses.  
\* $p \leq .05$ ; \*\* $p \leq .01$ .

Examining the results across Tables D1 to D3, we find that the findings are robust across all modeling approaches. First, the CITS sensitivity analysis estimates in Table D2 are nearly identical to those of the confirmatory CITS analysis in Table D1, albeit with larger standard errors due to the smaller sample sizes. Second, although the local average treatment effect estimates from the RD analysis are not significant, the direction of the Cohort 1 and Cohort 2 estimates from the RD analysis is the same as the direction of the estimates from both CITS analyses (except for Cohort 1 Year 1 in mathematics). This cross-modeling comparison suggests that our findings are not dependent on the modeling approach taken, but rather robust across all analyses.