Total Learning Classroom Impact Evaluation

Amanuel Medhanie

12/13/22

Table of Contents

Executive Summary

The Total Learning Classroom (TLC) is an AmeriCorps program designed to strengthen school capacity to deliver educational support to students. Schools that participate in TLC receive whole-school support by a Systems Coach who embeds into the school and works with leadership to make system-level improvements. In addition, schools receive AmeriCorps members who embed into classrooms and directly augment teacher capacity to support students.

The current evaluation examined year-end literacy outcomes for two types of classrooms in TLC schools: those that had an AmeriCorps member, called a Scholar Coach; and those that did not. Because the Systems Coach provides school- and grade-level support designed to benefit all classrooms, the current evaluation approach focused on analyzing the effect of having a Scholar Coach in a TLC classroom over and above the effect of having a Systems Coach assigned to the school.

Data from 39 classrooms were analyzed both descriptively and inferentially, with findings showing a potential effect for Scholar Coaches in kindergarten as well as a potential effect for evidence-based tutoring in kindergarten and second grade. However, findings were mixed overall, and should also be contextualized in light of limitations, particularly with respect to how Scholar Coaches are assigned to classrooms and a high percentage of missing data. Recommendations for the TLC program, including future evaluations, are noted.

TLC Impact Evaluation 2021-2022

Background

The Total Learning Classroom (TLC) program was developed to support whole-school learning outcomes in grades kindergarten through three. It does this in two ways. First, it provides the school AmeriCorps members trained in evidence-based literacy interventions, called Scholar Coaches, with up to one Scholar Coach per classroom. Those individuals deliver effective literacy tutoring to as many students as possible; and when they're not delivering tutoring they support teacher-led learning priorities, with a focus on language, literacy, and math.

The second way TLC supports whole-school learning is through an assigned Systems Coach. Systems Coaches add capacity to do several things that schools otherwise may be limited in accomplishing. First, they provide ongoing coaching to the Scholar Coaches, ensuring interventions are delivered effectively, assisting in the interpretation of student progress data, and helping tutors establish strong and responsive relationships with each student. Systems Coaches also help school leaders with decisions and practices that affect entire grades or the whole school, such as improving systems of data use, setting targets for learning improvement, and establishing classwide interventions when needed.

The TLC Program is an intense allocation of resources; thus, partner schools must meet at least three primary criteria: (1) 70% or more students must receive free or reduced price lunch (FRPL), (2) 70% or more students must have scored below proficient levels on the previous year's state test in reading, and (3) schools must sign an agreement indicating readiness to partner with the TLC Program. Schools that meet these criteria also tend to face other challenges (e.g., high student mobility) associated with being systematically under-resourced and marginalized by broader communities within which they exist.

Study Purpose and Design

The purpose of this study is to evaluate the literacy outcomes of students attending schools that participate in the TLC program. More specifically, the current study compared learning outcomes for two groups of students. The first group of students were in classrooms at TLC schools that had a Scholar Coach supporting learning activities. The second group of students were in classrooms at TLC schools but did not have a Scholar Coach supporting their classroom learning activities. Spring outcomes on student literacy data were evaluated for both groups.

This design provides insights into the effect of having a Scholar Coach in a TLC classroom over and above the effect of having a Systems Coach assigned to the school. Not all classrooms in a TLC school receive a Scholar Coach, but all classrooms in a TLC school do benefit from the system-level support provided by the Systems Coach, which includes a limited amount of capacity for evidence-based tutoring. Scholar Coaches can often provide evidence-based tutoring for a limited number of students outside their assigned classroom during times when

students are unavailable (e.g., lunch, recess, specialist). Thus, in a given TLC school all students benefit from the Systems Coach working with school leadership, but those with a Scholar Coach in their room likely have accelerated access to those benefits, as well as a greater likelihood of receiving evidence-based tutoring.

Methodology

Evaluation questions

This evaluation study was designed to answer the following question: *To what extent do students in classrooms assigned a Scholar Coach demonstrate higher literacy scores relative to similar students in classrooms without a Scholar Coach?*

Population and Sample

The population of interest for this study consists of kindergarten through 3rd grade students enrolled in high-poverty elementary schools in Minnesota that also have a high percentage of students that do not meet proficiency in reading[1]. The sample consisted of 1,235 students enrolled in one of the 10 schools participating in the TLC program.

Table 1: School characteristics

School Characteristic	Mean (N)
English Language Learners	31.3% (10)
Free or Reduced Priced Lunch	74.2% (10)
Met ELA Standards in 3rd Grade	20.4% (9)
Students Consistently Attend	82.9% (9)

A power analysis (using Optimal Design 3.01)[2] indicated that, in order to detect a difference (i.e. effect size) of 0.4 (Cohen's d, with a power of 0.85) on post-treatment assessment scores, and assuming the use of two-level hierarchical linear model[3], the necessary sample size for this study is 35 classrooms.

Sixteen classrooms were not assigned a Scholar Coach and 23 classrooms were assigned a Scholar Coach. Below is a table describing the Fall, Winter, and Spring assessment scores for

---

[1] For this study, "high-poverty" was defined as having 70% or more students must receive free or reduced price lunch (FRPL), and "high percentage of students that do not meet proficiency in reading" was defined as 70% or more students must have scored below proficient levels on the previous year's state test in reading. Schools that received TLC had to also sign an agreement indicating readiness to partner with the TLC program.

[2] Raudenbush, S. W., & Liu, X. (2000). Statistical power and optimal design for multisite randomized trials. Psychological methods, 5(2), 199..

[3] Raudenbush, S. W., & Bryk, A. S. (2002). Hierarchical linear models: Applications and data analysis methods (Vol. 1). sage.

the sample of students in this study. Fall assessments for the treatment and comparison groups were administered between 2021-09-02 and 2021-09-24. Winter assessments for the treatment and comparison groups were administered between 2022-01-11 and 2022-02-18, and Spring assessments for the treatment and comparison groups were administered between 2022-05-05 and 2022-05-27. All assessments were administered through FastBridge Learning[4].

Table 2: Average Assessment Scores by Student Group and Time of School Year

Student Characteristics	Fall Mean (SD, N)	Winter Mean (SD, N)	Spring Mean (SD, N)
Kindergarten - Correct Letter Sounds (Comparison)	6.7 (9.6, 53)	20.1 (17, 77)	25.9 (15.2, 57)
Kindergarten - Correct Letter Sounds (Treatment)	4 (6.2, 125)	24.9 (16.4, 126)	37.7 (18, 126)
1st Grade - Correct Nonsense Words (Comparison)	23.4 (28.5, 82)	20.1 (22.1, 65)	
1st Grade - Correct Nonsense Words (Treatment)	22.3 (24.5, 280)	43.1 (35.4, 275)	
1st Grade - CBM-Reading (Comparison)		26.2 (32.7, 68)	44 (33, 69)
1st Grade - CBM-Reading (Treatment)		27.6 (32.9, 280)	47.2 (39.3, 279)
2nd Grade - CBM-Reading (Comparison)	43 (36.4, 112)	58.7 (38.8, 98)	73.3 (44.6, 92)
2nd Grade - CBM-Reading (Treatment)	33.4 (37.2, 122)	50.9 (43.3, 118)	67.5 (46.2, 116)
3rd Grade - CBM-Reading (Comparison)	74.1 (48.4, 116)	81.4 (48.1, 107)	104.4 (50.1, 101)
3rd Grade - CBM-Reading (Treatment)	56.8 (38.7, 93)	67 (46, 148)	82.8 (46.3, 134)

The table above shows that, in general, comparison students had higher average baseline scores in the fall. However, the baseline differences between treatment (i.e. students in classrooms paired with a Scholar Coach) and comparison students (in classrooms not paired with a Scholar Coach) was relatively small for kindergartners and 1st graders (i.e. 2.7 and 1.1 respectively) and somewhat larger for students in 2nd and 3rd grade (i.e. 9.9 and 17.3 respectively).

Treatment vs. Comparison Classrooms

· All schools in the current study served students in kindergarten through grade three or beyond. Per the TLC program, the Systems Coach assigned to each school focused on the following factors designed to improve the school's capacity to facilitate improved learning:

---

[4] Thornblad, S. C., & Christ, T. J. (2014). Curriculum-based measurement of reading: Is 6 weeks of daily progress monitoring enough?. School Psychology Review, 43(1), 19-29.

(1) The Systems Coach embedded into the school's literacy leadership and data review practices. This typically took the form of attending meetings to advise on data practices, classwide intervention opportunities, or school/grade-level data.

(2) The Systems Coach completed a "multi-tiered systems of support" (MTSS) checklist in partnership with the school. MTSS is a research-based framework for organizing school resources, and the MTSS checklist used by TLC is designed to assist schools in prioritizing high-value schoolwide changes to improve learning.

(3) The Systems Coach interpreted class- and grade-level data to help allocate evidence-based tutoring services for students not in classrooms with Scholar Coaches.

· 23 classrooms were assigned TLC Scholar Coaches. The assignment process varied by school, but was typically led by principals, who in turn worked with individual teachers or grade-level teams to determine the best assignments for TLC Scholar Coaches. Some sites reported assigning Scholar Coaches to the classrooms with the highest level of reported instructional needs; other sites assigned Scholar Coaches to teachers who expressed the most interest in having additional support. For students in the treatment classrooms:

(1) Scholar Coaches provided evidence-based literacy tutoring to as many students as allowed by the classroom and school schedule (which was often 9 or more students in a given day). These tutoring sessions lasted 20 minutes and were scheduled daily until data indicated students no longer needed them.

(2) Scholar Coaches supported teacher-led learning activities and assisted with classroom management (e.g., sat with students during large-group instruction to facilitate engagement and learning).

(3) students received supports determined by the Systems Coach, school leadership, and/or teachers designed to improve whole-class learning outcomes (e.g., a teacher-administered classwide intervention)

· 16 classrooms were not assigned TLC Scholar Coaches. Students in these comparison classrooms received:

(1) typical classroom instruction without the support of a Scholar Coach

(2) supports determined by the Systems Coach, school leadership, and/or teachers designed to improve whole-class learning outcomes (e.g., a teacher-administered classwide intervention)

(3) evidence-based tutoring provided by a Scholar Coach from a different classroom. *(Note: this occurred in select cases, and is not considered to be a substantive contamination between treatment and comparison conditions, because the principle purpose of the current evaluation is to determine the unique effect of having the Scholar Coach embedded into the classroom.)*

Independent and Dependant Variables

There were three dependent variables in this study:

1) The number of Correct Letter Sounds measured during the spring administration of the FastBridge Test of Letter Sounds. The score ranges from 0 to 100+ correct sounds in one minute. It reflects a student's ability to efficiently recognize a common sound associated with a given grapheme (i.e., the letters that correspond to a discrete sound in the English language).

2) The number of Correct Nonsense Words measured during the winter administration of the FastBridge Test of Nonsense Words. The score ranges from 0 to 100+ correct nonsense words in one minute and reflects a student's ability to efficiently decode common (but unfamiliar) sequences of basic consonant-vowel-consonant words (e.g., nim).

3) The number of words read correctly per minute during the spring administration of the FastBridge Curriculum-based measure of reading fluency (CBM-Reading). The score ranges from 0 to 250+ words read correctly in one minute and reflects a student's ability to efficiently read connected text that is at a controlled level of difficulty.

The independent variables in this study were as follows:

· The score from the fall administration of the corresponding FastBridge assessment.

· Scholar Coach: A dichotomous variable indicating whether the student was in a class that was assigned a Scholar Coach (1 = yes, 0 = no).

· Tutored: A dichotomous variable indicating whether the student received direct tutoring through a Scholar Coach during the study period (1 = yes, 0 = no).

· ClassID: A numeric code indicating the unique classroom in which the student was enrolled during the timeframe of this study.

Analysis

The evaluation question – *"To what extent do students in classrooms paired with a Scholar Coach demonstrate higher literacy scores relative to similar students in classrooms without a Scholar Coach?"* – was answered using the lme function in the nlme[5] package for the statistical programming language R[6]. Separate models were fit for each of the four outcomes (i.e. Correct Letter Sounds for Kindergarten, Correct Nonsense Words for 1st grade, CBM-Reading for 2nd

---

[5] Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., Heisterkamp, S., Van Willigen, B., & Maintainer, R. (2017). Package 'nlme'. Linear and nonlinear mixed effects models, version, 3(1).

[6] R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

grade, and CBM-Reading for 3rd grade)[7]. Baseline differences were controlled for by including these assessment scores as covariates in the model.

The two-level hierarchical linear models took the general form:

$Y_{ij} = \beta_{0j} + \beta_{1j}X_{1ij} + \sum \beta_{pj}X_{pij} + e_{ij}$ Student - level

$\beta_{0j} = \gamma_{00} + \gamma_{01}W_{1j} + b_{0j}$ Class - level

Where Y_{ij} is the ith student's post-treatment assessment score (i.e. Correct Letter Sounds, Correct Nonsense Words, etc.), β_{0j} is the average value for the outcome for class j, β_{1j} is a coefficient reflecting the relationship between a student-level variable X_{1ij} (e.g. student's baseline assessment score) and the outcome, γ_{00} is the overall average value for the outcome, γ_{01} is a coefficient reflecting the relationship between a class-level variable W_{1j} (e.g. whether the class was assigned a Scholar Coach) and the average value for the outcome.

The authors hypothesized that the results would show a positive and statistically significant relationship between the Scholar Coach variable and the post-treatment assessment scores for each grade. This would support the hypothesis that classrooms paired with a Scholar Coach demonstrate higher literacy scores relative to similar students in classrooms without a Scholar Coach.

Results

The evaluation team performed descriptive and inferential analyses to answer the evaluation question *"To what extent do students in classrooms assigned a Scholar Coach demonstrate higher literacy scores relative to similar students in classrooms without a Scholar Coach?"*. This section presents these results.

The figure below presents the average Correct Letter Sounds by treatment group for kindergarteners in the sample. It shows that treatment and comparison kindergarten students in the sample began with similar baseline average correct letter sounds. After treatment (i.e. the classroom being paired with a Scholar Coach), the students in the treatment classes had higher average correct letter sounds compared to comparison students.
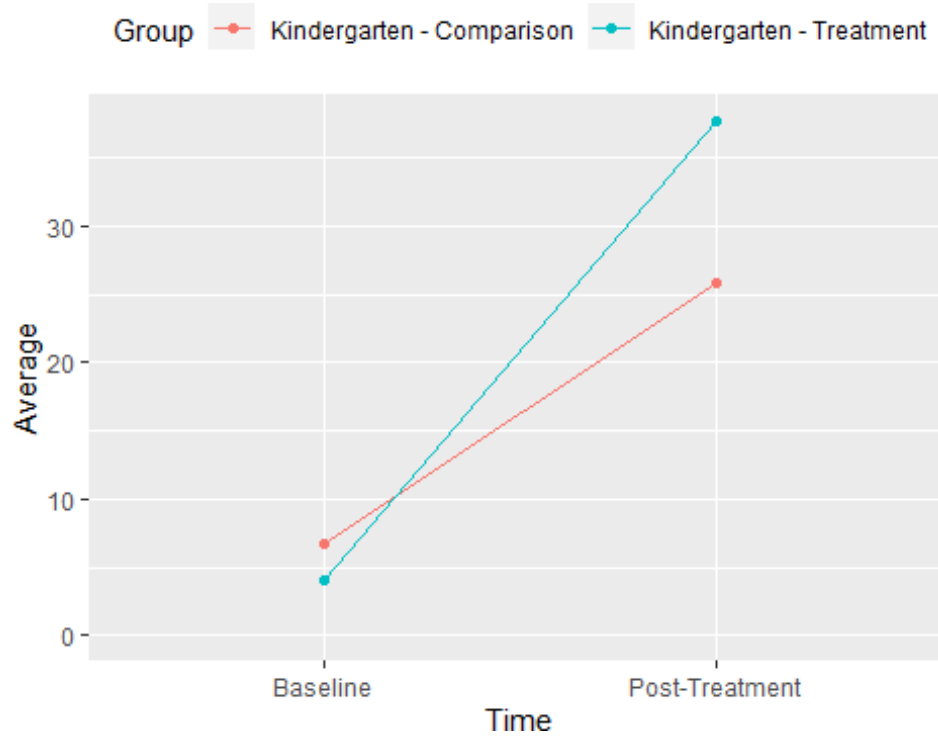
---

[7] 1st grade CBM-Reading scores wer not modeled because these data were only collected during winter (after treatment classrooms had already been served by a Scholar Coach) and spring.

Figure 1: Kindergarten Average Correct Letter Sounds by Treatment Group

The figure below presents the average Correct Nonsense Words by treatment group for 1st graders in the sample. It shows that treatment and comparison 1st grade students in the sample began with similar baseline average correct correct nonsense words. After treatment, the students in the treatment classes had higher average correct nonsense words compared to comparison students.
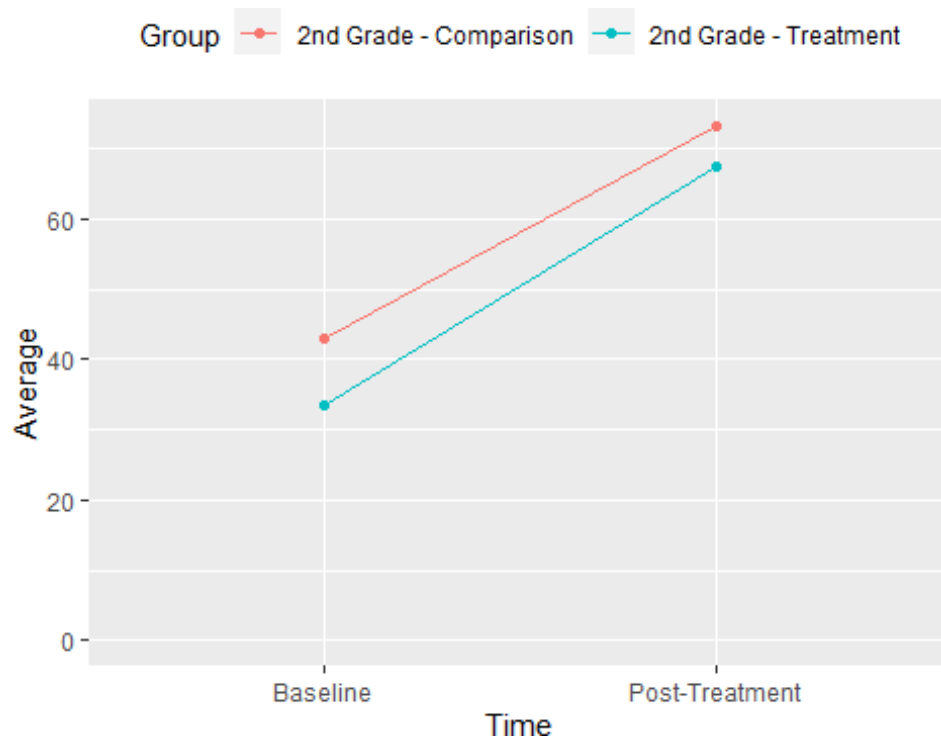
Figure 2: 1st Grade Average Correct Nonsense Words by Treatment Group
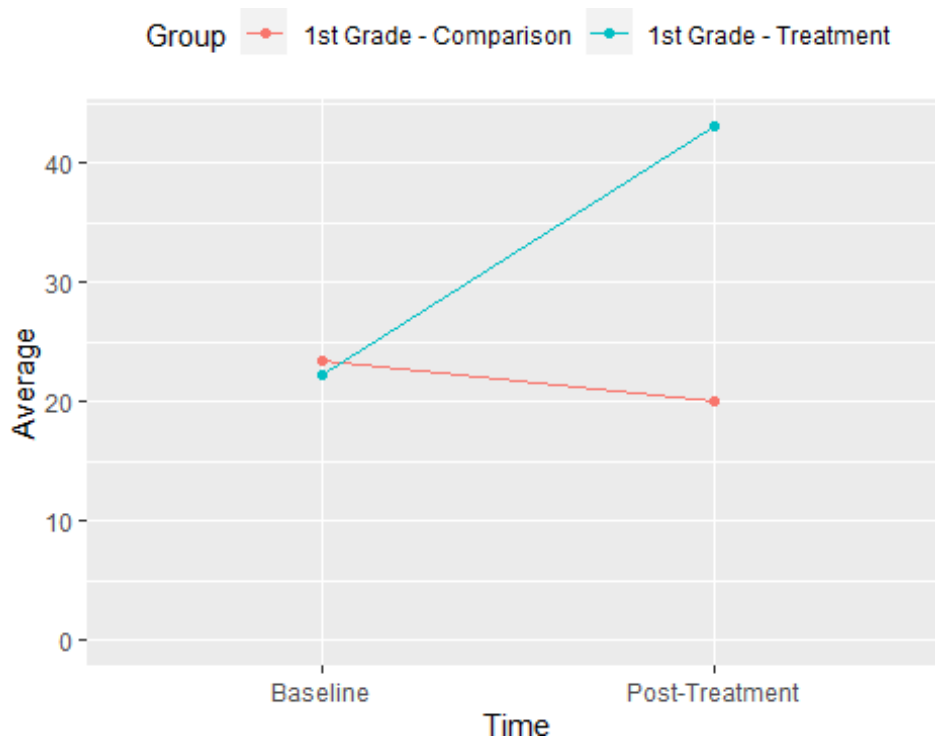
The figure below presents the average CBM-Reading scores by treatment group for 2nd graders in the sample. It shows that treatment and comparison 2nd grade students in the sample began with different baseline average CBM-Reading scores. The size of this gap between treatment and comparison students did not substantively change after one year of TLC program implementation.

Figure 3: 2nd Grade Average CBM-Reading Score by Treatment Group

The figure below presents the average CBM-Reading score by treatment group for 3rd graders in the sample. It shows that treatment and comparison 3rd grade students in the sample began with different baseline average CBM-Reading scores. The size of this gap between treatment and comparison students did not substantively change after one year of TLC program implementation.
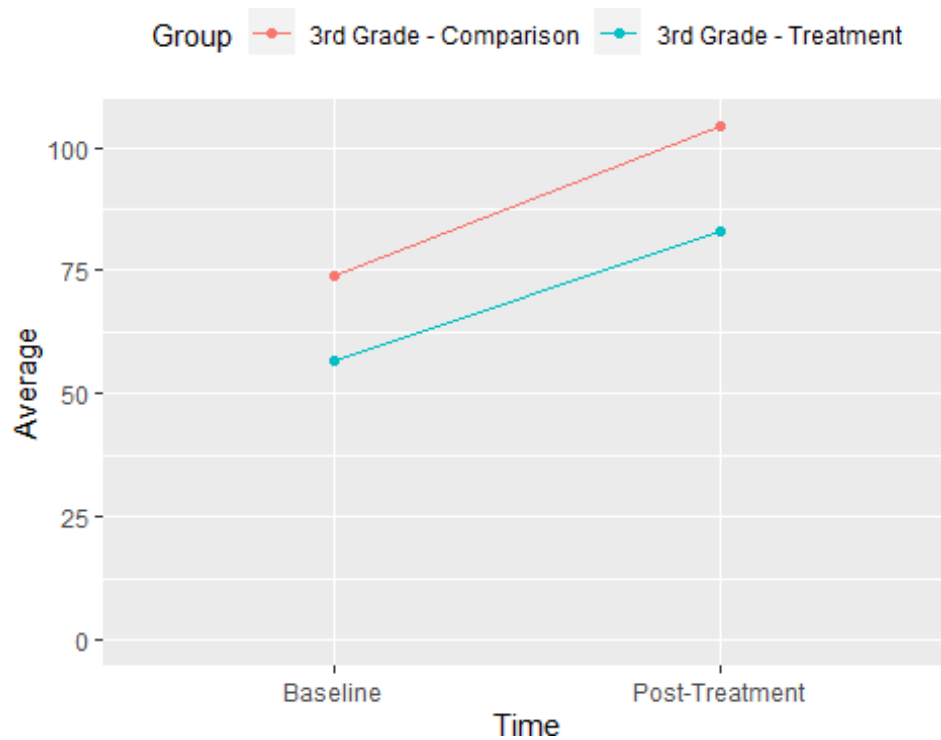
Figure 4: 3rd Grade Average CBM-Reading Score by Treatment Group

The table below presents the results from the hierarchical linear models used to answer the evaluation question for each of the grades. Cells containing an "*" indicate that a variable was a statistically significant (p < 0.05) predictor of that student outcome for that grade level.

Table 3: Model results by grade

Variable	Coefficient (SE) Kindergarten Correct Letter Sounds	Coefficient (SE) 1st Grade Correct Nonsense Words	Coefficient (SE) 2nd Grade CBM Reading	Coefficient (SE) 3rd Grade CBM Reading
Intercept	17.54** (2.69)	6.7 (8.44)	27.08** (4.07)	22.31** (4.06)
Baseline Score	1.42** (0.18)	0.92** (0.05)	1.07** (0.05)	1.01** (0.04)
Tutored	11.87** (2.96)	2.11 (2.68)	9.05* (3.71)	7.1 (3.8)
Scholar Coach	7.8* (3.04)	13.72 (8.26)	1.5 (4.19)	5.4 (3.37)

Note: * denotes statistical significance at p < 0.05, ** denotes statistical significance at p < 0.01. These results are based on a model assuming the normal distribution[8].

The results in the table are described for each grade and outcome below:

Kindergarten Correct Letter Sounds: The results of the model suggest that the baseline number of correct letter sounds of kindergarten students was positively associated with post scores (B = 1.42, p < 0.01). Also, on average, students who received tutoring had a higher number of correct letter sounds during the last observation, even after controlling for other factors (B = 11.87, p < 0.01). Finally, classrooms with a Scholar Coach had a higher number of correct letter sounds on average compared to classrooms without a Scholar Coach (B = 7.8, p < 0.05).

1st Grade Correct Nonsense Words: The results of the model suggest that the baseline number of correct nonsense words of 1st grade students was positively associated with post scores (B = 0.92, p < 0.01). However, neither tutoring nor having a scholar coach assigned to the classroom were associated with the final number of correct nonsense words after controlling for the other variables (B = 2.11, p > 0.05; B = 13.72, p > 0.05).

2nd Grade CBM-Reading: The results of the model suggest that the baseline CBM-Reading score of 2nd grade students was positively associated with post scores (B = 1.07, p < 0.01). Also, on average, students who received tutoring had a higher CBM-Reading score during the last observation, even after controlling for other factors (B = 9.05, p < 0.05). There was no association between whether classrooms had a Scholar Coach and students' post CBM-Reading score (B = 1.5, p > 0.05).

3rd Grade CBM-Reading: The results of the model suggest that the baseline CBM-Reading score of 3rd grade students was positively associated with post scores (B = 1.01, p < 0.01). However,

---

[8] Because the dependent variables reflect count data, the authors used Poisson and negative binomial models in addition to the Gaussian/normal model. The conclusions from these results did not change and so the results from the Gaussian/normal model are presented for simplicity.

neither tutoring nor having a Scholar Coach assigned to the classroom were associated with the final CBM-Reading score after controlling for the other variables (B = 7.1, p > 0.05; B = 5.4, p > 0.05).

Findings, Limitations, and Conclusions

Main Findings: The current evaluation sought to better understand the effects of the TLC program on student learning outcomes. Specifically, it investigated the effect of different classroom-level contexts on year-end literacy outcomes. Overall, the findings were mixed and need to be interpreted in the context of the evaluation's limitations (as noted below).

The following findings are worth highlighting: First, baseline assessment scores were correlated with post assessment scores, an expected outcome as it is common for students who score higher in a pretest assessment to also score higher at posttest. Second, the presence of a Scholar Coach was positively associated with an increase in correct letter sounds in kindergarten, even after controlling for baseline scores as well as whether a student received tutoring, but that was the only grade where this finding was observed. Third, students who received tutoring tended to have higher scores in kindergarten and 2nd grade, although this relationship was not present among 1st graders (for correct nonsense words) or 3rd graders (for CBM-Reading).

Taken together, these findings provide tentative evidence that some aspects of the TLC program may be associated with improved student outcomes; however, what aspects support learning and in which grades are unclear. Only in kindergarten did the Scholar Coach presence have a statistically significant impact on year-end outcomes. That may be due to something unique to the kindergarten experience (e.g., many students accessing formal schooling for the first time), which in turn allows for a greater effect of the Scholar Coach, but that is a tentative interpretation. A similarly tentative interpretation may be made with respect to tutoring in kindergarten and second grade, where access to tutoring was significantly associated with year-end outcomes. That finding is consistent with robust research that shows evidence-based tutoring is effective, so it is less surprising. But in the context of non-significant associations in the two other grades, the extent to which TLC facilitates effective tutoring is unclear (though small sample sizes urge caution in any interpretation).

Limitations: Three primary limitations are important to consider when interpreting the above findings. First, the assignment mechanism behind what TLC classrooms get a Scholar Coach and what classrooms do not is largely unknown. Thus, it is unclear whether other factors (e.g., teacher quality or experience) might be influencing the results. In the future, it would be beneficial to either control Scholar Coach assignment (e.g., randomization) or fully understand the assignment mechanisms so as to establish a stronger quasi-experimental comparison. Second, incomplete data resulted in approximately 40% of the sample missing class-specific data. This was due to the fact that several schools desired to not identify their classroom teacher (all data in this report are fully aggregated and anonymized), which resulted in the main inferential results being based on only 60% of all students in the sample. Third, students in both treatment and comparison conditions potentially benefit from the school-wide support

provided by Systems Coach, as described above. The evaluation was designed to not control for the presence of a Systems Coach and, therefore, did not capture a possible significant benefit the program has on student literacy outcomes.

Conclusions for TLC Evaluation and Practice: Despite the limitations and mixed findings, the current study provides tentative evidence that the TLC program produces a net benefit on learning outcomes. Kindergarten students and second grade students who were tutored demonstrated improved outcomes. When considered in light of existing research on tutoring–as well as descriptive outcomes in the current study and elsewhere that indicate benefits from the program–a pattern emerges that efforts to augment school capacity for improved learning outcomes are promising. The current findings can be interpreted as evidence that such efforts should continue, and that stronger research designs might better identify the true effects they have on learning outcomes. In particular, determining feasible designs that control for Scholar Coach classroom assignments holds particular promise.