CommonLit

Digital CommonLit Lesson Usage and NJSLA ELA Achievement

Results from a New Jersey District, School Year 2024-2025



November 2026

Table of Contents

Executive Summary 3

Introduction 4

- Research Questions 5

Method 5

- Sample 5
- Measures 6
  - CommonLit Usage 6
  - NJSLA ELA Assessment 7
  - Covariates 8

Results 9

- How Did Students Perform on the NJSLA ELA Assessment in the 2023-2024 and 2024-2025 School Years? 9
- How Did Students Use Digital CommonLit Lessons During the 2024-2025 School Year? 10
- Was Digital CommonLit Lesson Usage Associated with Students' Achievement on the NJSLA ELA Assessment in the 2024-2025 School Year? 11
  - Continuous Outcome: Scale Scores 12
  - Binary Outcome: Performance Levels of Meeting or Exceeding Expectations 14
  - Tests of Moderation and Robustness 16

Discussion 17

Limitations 18

- Observational Study Design 18
- Usage and Instructional Context 18
- Outcome Scaling and Sample 18
- Ongoing Research 19

References 20

Appendix 21

Executive Summary

This report examines whether students' use of digital CommonLit lessons was associated with their achievement on the New Jersey Student Learning Assessment for English Language Arts (NJSLA ELA) assessment in the 2024-2025 school year (SY). The study was conducted in partnership with a large, urban district in New Jersey. The sample included 2,495 students in Grades 4-9 who took the NJSLA ELA assessment in both 2023-2024 and 2024-2025 and submitted at least one digital CommonLit lesson in 2024-2025. These students were taught by 84 teachers across 18 schools.

In 2024-2025, 40% of students in the sample met or exceeded expectations on the NJSLA and students submitted an average of 19 digital CommonLit lessons. Students who submitted more digital CommonLit lessons than their classmates had higher NJSLA ELA scale scores, after accounting for prior achievement and student characteristics. The association was modest: a difference of about 5 lessons (25th to 75th percentile of relative usage) corresponded to about 1.2 scale score points (0.03 *SD*). These students were also more likely to meet or exceed expectations. The same 5-lesson difference corresponded to a 4-percentage-point difference in the predicted likelihood of meeting expectations (35% vs. 39%). Results held across robustness checks and did not differ significantly by student group. Because this was an observational study, these findings showed association, not causal effect.

This study was rated as meeting the Every Student Succeeds Act (ESSA) Level 3 standards ("Promising Evidence") by Instructure's Research team, a third-party edtech research group. This treatment-only study used correlational analyses with statistical controls for prior achievement and student characteristics and found that students who submitted more digital CommonLit lessons than their classmates had significantly higher NJSLA ELA scores and a greater likelihood of meeting or exceeding expectations. Ongoing CommonLit studies aim to meet ESSA Level 2 standards ("Moderate Evidence") by examining the impact of CommonLit by comparing students at schools that use CommonLit to a comparison group of students at schools that do not use CommonLit.



Introduction

In 2024, the National Assessment of Educational Progress (NAEP) reading assessment showed significant declines since 2019 in average reading scores and in the percentage of students who scored at or above the NAEP Proficient level (National Center for Education Statistics, 2025). Only 30% of eighth-grade students scored at or above the NAEP Proficient level, with 37% scoring at the NAEP Basic level and 33% scoring below NAEP Basic. This was significantly lower than the pre-pandemic result in 2019, when 34% of eighth-grade students scored at or above NAEP Proficient, 39% scored at NAEP Basic, and 27% scored below NAEP Basic. Declines in reading proficiency have also been documented across several other reading assessments (Kuhfeld & Lewis, 2025; Kuhfeld et al., 2023), with widening achievement gaps for economically disadvantaged and traditionally underserved students (Fahle et al., 2024). These national trends make it especially important for districts to understand whether the literacy tools they have adopted are associated with better student reading outcomes.

CommonLit is a nonprofit education technology organization dedicated to ensuring that all students—especially those in economically disadvantaged schools—graduate with the reading, writing, communication, collaboration, and problem-solving skills they need to be successful. CommonLit's theory of change is that when schools—including those in underserved communities—have access to a coherent curriculum, aligned professional learning, meaningful assessments, and implementation support, all students will be able to read on grade level and succeed in school and beyond. To this end, CommonLit has developed a free, year-long core English Language Arts (ELA) curriculum for Grades 6-12, rated as all-green for Grades 6-8 by EdReports, and offers wraparound support services through its paid packages for schools and districts. CommonLit also continues to offer a free Text Library with thousands of articles and reading passages for supplemental use.

In 2013, CommonLit began as a research project with 30 published online lessons (Brown, 2014). Since then, CommonLit has conducted studies of how its programs relate to student outcomes, including studies rated as meeting ESSA Level 2 and Level 3 standards. These studies found that higher CommonLit usage was associated with greater reading growth on the CommonLit Assessment Series (CommonLit, 2026; CommonLit & Mathematica, 2023) and greater writing growth (CommonLit, 2024). Studies with district

partners also found that higher usage was associated with higher state ELA scores in Maryland (CommonLit, 2025) and Florida (Lee & Shah, 2024). This technical report describes a study conducted in partnership with a large, urban district in New Jersey that examined the association between students' use of digital CommonLit lessons and their achievement on the New Jersey Student Learning Assessment (NJSLA) ELA assessment in the 2024-2025 school year.

Research Questions

The study addressed three research questions:

- How did students perform on the NJSLA ELA assessment in the 2023-2024 and 2024-2025 school years?
- How did students use digital CommonLit lessons during the 2024-2025 school year?
- Was digital CommonLit lesson usage associated with students' achievement on the NJSLA ELA assessment in the 2024-2025 school year?

Method

Sample

The study took place in a large, urban school district in New Jersey that serves more than 8,500 students from preschool through high school. The analytic sample included students who took the NJSLA ELA assessment in both school years, had complete demographic data, and submitted at least one digital CommonLit lesson in 2024-2025. This resulted in 2,495 students in Grades 4-9 taught by 84 teachers across 18 schools (Table 1). On average, each teacher taught about 30 students (range: 1 to 114).

Table 1. Number of Students in Each Grade Level in the Sample

Grade (SY 2024-2025)	*N*	%
Grade 4	372	15%
Grade 5	408	16%
Grade 6	445	18%
Grade 7	461	19%
Grade 8	408	16%
Grade 9	401	16%
Total	2,495	100%

Note. Percentages were adjusted to sum to 100.

Of the students in the sample, 51% were female and 49% were male. By race/ethnicity, 80% of students were Black (not Hispanic/Latino), 17% were Hispanic/Latino, 1% were White (not Hispanic/Latino), and 2% were Asian, American Indian or Alaska Native, Native Hawaiian or Other Pacific Islander, or two or more races. Ten percent of students received special education services, 12% were eligible for the Language Instruction Educational Program (LIEP), and 8% were identified as gifted and talented.

Measures

CommonLit Usage

The primary predictor in this study was digital CommonLit usage, as measured by backend platform data. Students and teachers can access lessons across the CommonLit platform through either the CommonLit 360 curriculum, Text Library, or Target Lessons. Notably, the CommonLit 360 curriculum consists of six units per grade level across Grades 6-12. Each unit consists of Core Unit Texts that students engage in throughout the unit through Reading Lessons, Writing Lessons, and more. The CommonLit Text Library consists of thousands of articles and reading passages for Grades 3-12. CommonLit's Target Lessons are skill-focused lessons that are designed to close specific learning gaps. This study counted only digital lessons submitted by students (CommonLit 360, Library, and Target Lessons) as digital lessons can be attributed to individual students. Lessons downloaded by teachers were not included in the analysis.

Usage was measured as the number of digital lessons each student submitted during the 2024-2025 school year. To distinguish individual usage from classroom-level usage, each student's lesson count was separated into two components: (1) Relative usage was the student's usage relative to their classroom average, or how many more or fewer lessons they submitted than their classmates, and (2) Classroom average usage was the average usage of the student's classroom.

Classrooms were defined by each student's main teacher. The main teacher was the teacher through whom the student submitted the most digital lessons during the 2024-2025 school year, combined across CommonLit 360, Library, and Target lessons. Classroom average usage was calculated across students in the analytic sample who shared a main teacher, and relative usage was each student's lesson count minus that average.

NJSLA ELA Assessment

Student achievement was measured with the New Jersey Student Learning Assessment (NJSLA) ELA assessment, the state's annual summative assessment for students in Grades 3 through 9, designed to measure whether students are meeting state academic standards (New Meridian, 2025). Scores from the 2024-2025 school year were used to create two outcome measures.

The first was each student's overall NJSLA ELA scale score, which ranged from 650 to 850. Because NJSLA scale scores are not vertically scaled across grades, the scale scores were standardized (*z*-scored) within grade using the means and standard deviations of the sample, so that each score reflects a student's standing relative to other students in the same grade in the sample.

The second was a binary indicator of whether the student met or exceeded expectations. The NJSLA places students in one of five performance levels: Level 1 (Did Not Yet Meet Expectations), Level 2 (Partially Met Expectations), Level 3 (Approached Expectations), Level 4 (Met Expectations), and Level 5 (Exceeded Expectations). Students at Levels 4 and 5 were coded as meeting expectations, and students at Levels 1 through 3 were coded as not meeting expectations. For 2023-2024, meeting expectations was determined from the scale score (≥ 750, the Level 4 minimum in every grade); for 2024-2025, it was determined

from the reported performance level. The 2023-2024 binary indicator was used only for the descriptive comparison in Table 2.

In addition, each student's 2023-2024 NJSLA ELA scale score was *z*-scored within each grade and served as the prior-achievement control (see Covariates). Statewide scale score reliability for the NJSLA ELA assessment was high in both years. In 2024-2025, average reliability ranged from .90 to .93 across Grades 4-9, form-level reliability ranged from .84 to .95, and average standard errors of measurement ranged from 12.50 to 14.02 scale score points (New Meridian, 2025). For the 2023-2024 prior-year scores (Grades 3-8), average reliability ranged from .90 to .93, form-level reliability ranged from .83 to .95, and average standard errors of measurement ranged from 12.10 to 13.78 scale score points (New Meridian, 2024).

Covariates

To account for differences that could influence NJSLA achievement regardless of CommonLit usage, the analysis controlled for prior achievement, grade level, and several student characteristics. Prior achievement was each student's 2023-2024 NJSLA ELA scale score, standardized within grade (see NJSLA ELA Assessment). Classroom mean prior achievement (the average 2023-2024 *z*-score of students associated with the same teacher) was also included so that student-level and classroom-level prior achievement could be separated. Grade level was the student's grade in 2024-2025 (Grades 4-9). Student characteristics included LIEP eligibility, special education status (whether the student received special education services), gifted and talented status (whether the student was identified as gifted and talented), gender (female or male), and race/ethnicity (Black, not Hispanic/Latino vs. all other students). Because 80% of students were Black (not Hispanic/Latino) and no other group was large enough to model separately (Hispanic/Latino, 17%; all remaining categories combined, 3%), race/ethnicity was coded as Black, not Hispanic/Latino vs. all other students. NJSLA scores and data on demographic characteristics were provided by the district.

Results

How Did Students Perform on the NJSLA ELA Assessment in the 2023-2024 and 2024-2025 School Years?

Overall, 40% of students in the sample met or exceeded expectations in 2024-2025. Figure 1 shows NJSLA Performance Levels for the 2024-2025 school year. Table 2 provides a breakdown of NJSLA performance by grade level across both school years.

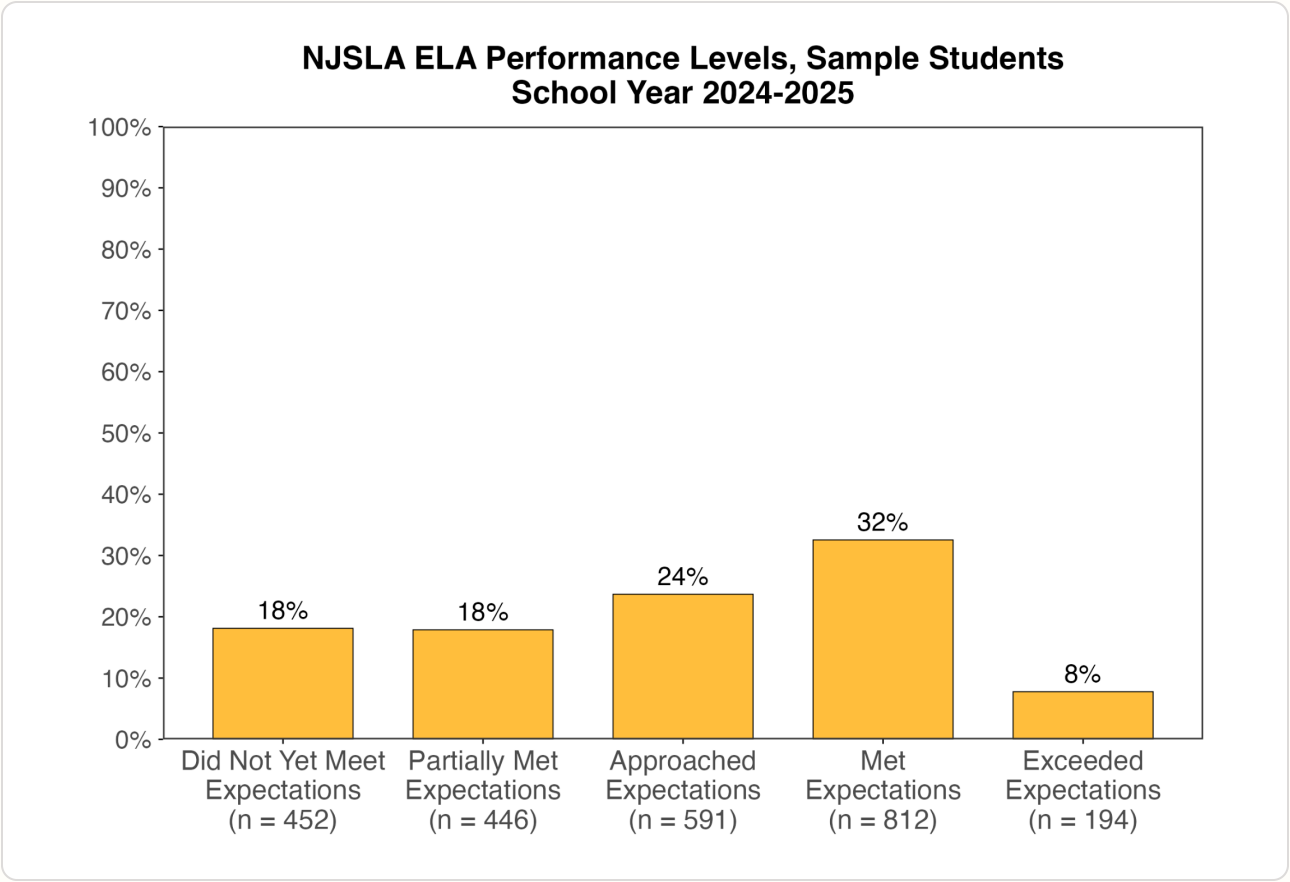


Figure 1. Percent of students in the sample at each performance level on the NJSLA ELA assessment in school year 2024-2025. Percentages were adjusted to sum to 100.

Table 2. NJSLA ELA Achievement in SY 2023-2024 and SY 2024-2025

		SY 2023-2024		SY 2024-2025	
Grade (SY 2024-2025)	*N* (%)	Average NJSLA ELA Scale Score	% met or exceeded expectations	Average NJSLA ELA Scale Score	% met or exceeded expectations
Grade 4	372 (15%)	726	30%	739	44%
Grade 5	408 (16%)	736	41%	740	44%
Grade 6	445 (18%)	742	43%	741	45%
Grade 7	461 (19%)	730	28%	736	38%
Grade 8	408 (16%)	730	32%	732	36%
Grade 9	401 (16%)	728	29%	735	36%
All Students	2,495 (100%)	–	34%	–	40%

Note. Grade refers to students' grade in 2024-2025; 2023-2024 results are from the previous grade. Average scale scores are not reported for all students combined because NJSLA scale scores are not vertically scaled and should not be compared across years or grades. Percentages were adjusted to sum to 100.

How Did Students Use Digital CommonLit Lessons During the 2024-2025 School Year?

Students in the sample (Grades 4-9) submitted an average of 19 digital CommonLit lessons, of which 9 were CommonLit 360 Lessons, 6 were Library Lessons, and 4 were Target Lessons (Table 3; see Appendix Table A1 for more details). Usage varied widely across students (*SD* = 12.05; range = 1 to 127) and by grade, from 15 lessons in Grade 7 to 24 in Grade 5. The mix of lessons also differed by grade. CommonLit 360 is designed for Grades 6-12, so Grades 4 and 5 usage consisted entirely of Library and Target lessons, while CommonLit 360 accounted for most usage in Grades 6 through 9.

Table 3. Average Number of Digital CommonLit Lessons Submitted During the 2024-2025 School Year

Grade (SY 2024-2025)	*N* (%)	Average Digital CommonLit Lessons	Average Digital CommonLit 360 Lessons	Average Digital CommonLit Library Lessons	Average Digital CommonLit Target Lessons
Grade 4	372 (15%)	18	-	9	8
Grade 5	408 (16%)	24	-	15	8
Grade 6	445 (18%)	20	12	2	5
Grade 7	461 (19%)	15	12	2	1
Grade 8	408 (16%)	20	16	2	2
Grade 9	401 (16%)	20	14	6	0
All Students	2,495 (100%)	19	9	6	4

Note. Grade refers to students' grade in 2024-2025. A dash indicates no available lessons for that grade. Sample size percentages were adjusted to sum to 100 and lesson components may not sum to the digital CommonLit lessons average because of rounding.

Was Digital CommonLit Lesson Usage Associated with Students' Achievement on the NJSLA ELA Assessment in the 2024-2025 School Year?

To test the relationship between digital CommonLit lesson usage and students' performance on the NJSLA ELA assessment, we conducted two statistical models. The first model examined NJSLA ELA scale scores as a continuous outcome variable. The second model examined NJSLA performance levels as a binary outcome variable, testing whether students were more likely to meet or exceed expectations (Levels 4 and 5). Both models accounted for students being nested within classrooms and the possibility that students in the same classroom may perform more similarly than students in a different classroom.

Usage was entered as two components: relative usage and classroom average usage. Both models also controlled for the following covariates: prior NJSLA ELA scale scores from the 2023-2024 school year, classroom mean prior achievement, grade level (reference group: Grade 4), gender (reference group: female), race/ethnicity (Black, not Hispanic/Latino vs. all other students; reference group: all other students), special

education status (reference group: no), Language Instruction Educational Program (LIEP) eligibility (reference group: no), and gifted/talented status (reference group: no). NJSLA ELA scale scores from both school years were standardized (*z*-scored) within grade prior to analysis, because the NJSLA is not vertically scaled.

Continuous Outcome: Scale Scores

Table 4 displays the results of the model that examined NJSLA ELA scale scores. Prior-year NJSLA ELA scores were the strongest predictor of students' 2024-2025 scores, $b = 0.69$, $SE = 0.01$, $t = 47.58$, $p < .001$. Importantly, there was a statistically significant relationship between relative usage and NJSLA ELA scores, $b = 0.01$, $SE = 0.002$, $t = 3.61$, $p < .001$, 95% CI [0.003, 0.01]. That is, even after controlling for prior achievement and student characteristics, students who submitted more digital CommonLit lessons than their classmates scored slightly higher on the NJSLA ELA assessment. This result can be translated into about 1.2 scale score points (95% CI [0.6, 1.9]; 0.03 *SD*, 95% CI [0.01, 0.05]) for a difference of about 5 lessons (the interquartile range of relative usage). Classroom average usage was not significantly related to scale scores, $b = 0.003$, $SE = 0.002$, $t = 1.19$, $p = .236$, but the 95% confidence interval [−0.002, 0.01] is wide enough that a between-classroom association as large as the within-classroom estimate cannot be ruled out. Figure 2 shows the relationship between relative usage and NJSLA ELA scale scores. The intermediate models, sensitivity analyses, and moderation tests are in Appendix Tables A2, A3, and A5 respectively.

Table 4. Hierarchical Linear Model Results: NJSLA ELA Scale Scores

Random Effect	Variance	*SD*	95% CI for *SD*
Classroom intercept	0.04	0.20	[0.15, 0.23]
Residual	0.30	0.55	[0.53, 0.56]

Predictor	Estimate	*SE*	95% CI	*df*	*t*-value	*p*-value
Intercept	0.02	0.06	[−0.10, 0.14]	165.20	0.34	.732
Relative usage (digital lessons vs. classroom average)	**0.01**	**0.002**	**[0.003, 0.01]**	**2425.53**	**3.61**	**< .001**
Classroom average usage	0.003	0.002	[−0.002, 0.01]	83.43	1.19	.236

Predictor	Estimate	*SE*	95% CI	*df*	*t*-value	*p*-value
Prior NJSLA scores (2023-2024 *z*-score)	0.69	0.01	[0.66, 0.72]	2430.33	47.58	< .001
Classroom mean prior NJSLA scores	0.23	0.05	[0.14, 0.32]	109.82	4.95	< .001
Grade 5 (vs. Grade 4)	−0.04	0.07	[−0.18, 0.10]	175.87	−0.51	.608
Grade 6	−0.02	0.07	[−0.17, 0.12]	112.78	−0.32	.752
Grade 7	0.08	0.07	[−0.06, 0.22]	111.93	1.07	.288
Grade 8	0.03	0.08	[−0.12, 0.17]	117.52	0.34	.731
Grade 9	0.00	0.09	[−0.17, 0.17]	81.62	0.00	.999
Male	−0.07	0.02	[−0.12, −0.03]	2445.39	−3.21	.001
Black (not Hispanic/Latino) vs. all other students	0.01	0.03	[−0.05, 0.07]	2461.19	0.38	.706
Special education	−0.22	0.05	[−0.31, −0.13]	1374.47	−4.62	< .001
LIEP eligible	0.08	0.04	[−0.004, 0.17]	2158.66	1.86	.063
Gifted and talented	0.22	0.05	[0.14, 0.31]	2455.99	4.97	< .001

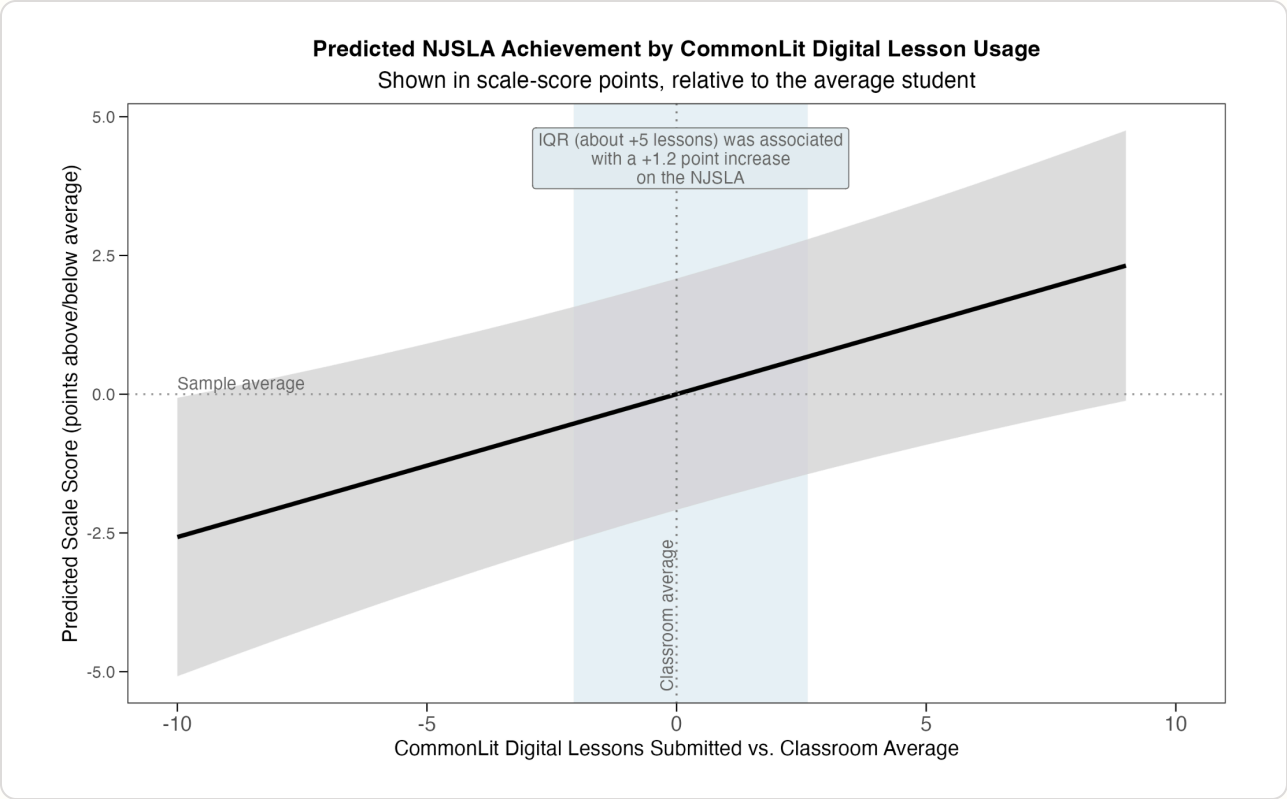


Figure 2. Predicted NJSLA ELA scale score points above or below the average student in the sample, based on digital CommonLit lessons submitted relative to one's classmates.

Note. Predicted scale score points relative to the average student, converted from *z*-scores using the overall standard deviation (39.13). The shaded blue band marks the 25th to 75th percentile of relative usage (−2.06 to +2.63 lessons), and the gray band is the 95% confidence interval.

Binary Outcome: Performance Levels of Meeting or Exceeding Expectations

Table 5 displays the results of the model that examined whether students met or exceeded expectations on the NJSLA ELA assessment. Similar to the first model, prior-year NJSLA ELA scores were the strongest predictor, $b = 2.60$, $SE = 0.13$, $z = 20.20$, $p < .001$. Notably, the relationship between relative usage and meeting or exceeding expectations was statistically significant, $b = 0.04$, $SE = 0.01$, $z = 3.19$, $p = .001$ (odds ratio [OR] = 1.04, 95% CI [1.01, 1.06]). That is, even after controlling for prior achievement and student characteristics, students who submitted more digital CommonLit lessons than their classmates were more likely to meet or exceed expectations. A difference of about 5 lessons (the interquartile range of relative usage) corresponds to 1.19 times the odds (95% CI [1.07, 1.32]). For a reference-group student (defined in the Figure 3 note), this is a predicted likelihood of 35% at the 25th percentile of relative usage and 39% at the 75th, a

difference of about 4 percentage points. Classroom average usage was not significantly related to the outcome, b = 0.01, SE = 0.01, z = 1.53, p = .125 (OR = 1.01, 95% CI [1.00, 1.03]). Figure 3 shows the relationship between relative usage and the predicted likelihood of meeting or exceeding expectations. The sensitivity analyses are in Appendix Table A4, and the moderation tests are in Appendix Table A5.

Table 5. Mixed-Effects Logistic Regression Results: Meeting or Exceeding Expectations on the NJSLA ELA Assessment

Random Effect	Variance	*SD*
Classroom intercept	0.41	0.64

Predictor	Estimate	*SE*	OR	95% CI	*z*-value	*p*-value
(Intercept)	−0.55	0.28	0.58	[0.33, 1.01]	−1.93	.054
Relative usage (digital lessons vs. classroom average)	**0.04**	**0.01**	**1.04**	**[1.01, 1.06]**	**3.19**	**.001**
Classroom average usage	0.01	0.01	1.01	[1.00, 1.03]	1.53	.125
Prior NJSLA scores (2023-2024 *z*-score)	2.60	0.13	13.49	[10.48, 17.36]	20.20	< .001
Classroom mean prior NJSLA scores	0.89	0.24	2.44	[1.53, 3.88]	3.75	< .001
Grade 5 (vs. Grade 4)	−0.29	0.33	0.75	[0.39, 1.44]	−0.86	.387
Grade 6	−0.02	0.32	0.98	[0.52, 1.84]	−0.08	.939
Grade 7	−0.49	0.31	0.61	[0.33, 1.13]	−1.57	.116
Grade 8	−0.66	0.33	0.52	[0.27, 0.99]	−2.00	.046
Grade 9	−0.83	0.38	0.44	[0.21, 0.93]	−2.16	.031
Male	−0.37	0.13	0.69	[0.54, 0.88]	−2.96	.003
Black (not Hispanic/Latino) vs. all other students	0.08	0.17	1.08	[0.77, 1.52]	0.44	.657
Special education	0.07	0.30	1.07	[0.60, 1.92]	0.23	.816
LIEP eligible	0.56	0.24	1.76	[1.10, 2.81]	2.37	.018
Gifted and talented	1.08	0.33	2.93	[1.53, 5.64]	3.23	.001

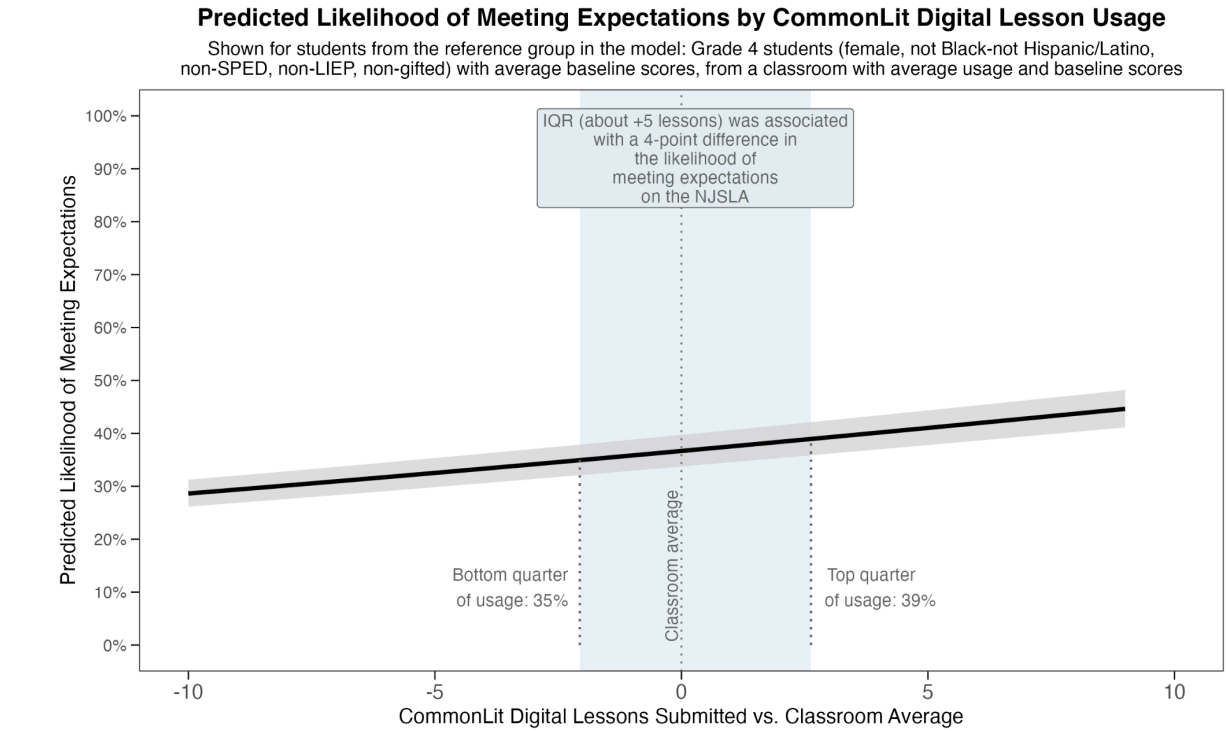


Figure 3. Predicted likelihood of meeting expectations on the NJSLA ELA assessment by relative digital CommonLit lesson usage.

Note. Reference-group student: Grade 4, female, not Black (not Hispanic/Latino), not receiving special education services, not LIEP-eligible, not gifted and talented, with average prior achievement in a classroom with average usage and average prior achievement. Predictions are population-level (fixed-effects) predictions. The shaded blue band marks the 25th to 75th percentile of relative usage.

Tests of Moderation and Robustness

We conducted several additional analyses to test whether these results were robust (see Appendix Tables A3 and A4). The first set of analyses tested whether the results of both models held under different scenarios: excluding small classrooms (n < 10 students), excluding students with usage outliers (> 80 lessons; n = 5), and excluding the five classrooms flagged as influential in the continuous model (Cook's D). Under all three scenarios, relative usage remained a significant predictor of both NJSLA ELA scale scores (b's = 0.01, all p < .001) and meeting or exceeding expectations (OR = 1.04, all $p \leq .002$), demonstrating that results were robust. The second set of analyses tested whether the relationship between relative usage and each outcome was moderated by prior achievement, gender, race/ethnicity, special education status, LIEP eligibility, or

gifted/talented status (see Appendix Table A5). None of the 12 interaction tests (6 per outcome) was statistically significant (unadjusted p = .051 to .832), and all Holm-adjusted p-values were ≥ .61. The third set of analyses tested whether the relationship between relative usage and scale scores varied across classrooms by adding a random slope for relative usage (see final row of Appendix Table A3). This did not significantly improve model fit, $\chi^2(2) = 4.79$, p = .091, and produced a singular fit, so we retained the random-intercept model. In sum, these analyses were consistent with the main results.

Discussion

This study examined whether 2,495 students in Grades 4-9 in a large, urban New Jersey district who submitted more digital CommonLit lessons than their classmates had higher 2024-2025 NJSLA ELA achievement, after accounting for their 2023-2024 NJSLA performance. In 2024-2025, 40% of students met or exceeded expectations and submitted an average of 19 digital CommonLit lessons. Students who submitted more digital lessons than their classmates scored higher on the NJSLA ELA, after controlling for prior-year achievement, classroom mean prior achievement, grade, and student characteristics. A difference of about 5 lessons corresponded to about 1.2 scale score points (0.03 *SD*), a modest difference, and to a 4-percentage-point difference in the predicted likelihood of meeting or exceeding expectations (35% vs. 39%) for a reference-group student.

Importantly, the association between NJSLA ELA scores and digital CommonLit usage was consistently positive and significant for both outcomes (scale scores and performance levels). These findings held after excluding small classrooms, students with very high usage, and influential classrooms. Tests of moderation found no statistically significant differences by prior achievement, gender, race/ethnicity, special education status, LIEP eligibility, or gifted and talented status. In sum, students who submitted more digital CommonLit lessons than their classmates tended to have higher NJSLA ELA achievement.

Limitations

Observational Study Design

This was an observational study. Students were not randomly assigned to levels of CommonLit usage, so students who submitted more lessons may differ from their classmates in ways beyond usage, such as motivation, attendance, or access to technology. This study attempted to address potential selection bias by comparing students who were taught by the same teacher and by statistically controlling for prior-year NJSLA achievement, classroom mean prior achievement, grade, and student characteristics. Because unmeasured differences between students (such as differences in motivation or attendance) cannot be ruled out, claims from this study should be made with caution, as this design did not support causal conclusions. However, because random assignment of instructional materials may not be feasible in practice, this design provides a way to examine the association between CommonLit usage and a state assessment outcome while balancing demands on district and school leaders.

Usage and Instructional Context

This study was not a comprehensive test of all instructional resources used in the district in 2024-2025. Usage was measured as digital lessons submitted by students, so lessons that teachers downloaded were not counted. The study also has no information on which core ELA curriculum students received or what other resources, such as supplemental intervention programs, they used. CommonLit may have served as a supplement or as part of core instruction. Comparing students who were taught by the same teacher likely reduces differences in curriculum across classrooms, but it cannot show how CommonLit's association with achievement compares with or combines with other resources. Future research could examine all literacy resources used in a classroom and estimate their independent and combined associations with ELA achievement.

Outcome Scaling and Sample

NJSLA scale scores are not vertically scaled, so scores were standardized within grade using the means and standard deviations of this sample. As a result, findings describe students' relative standing within this sample, not absolute change or standing relative to the state. The sample included only students in Grades 4-9 who submitted at least one

digital lesson and had NJSLA scores in both years, so the study cannot compare CommonLit users with non-users. The analyses assumed a linear relationship between usage and outcomes. Nonlinearity was not tested, although a larger CommonLit study found that reading growth increased more gradually at the highest levels of usage (CommonLit, 2026). Finally, the sample came from a single district, so findings may not generalize to other districts, grades, or school years.

Ongoing Research

CommonLit regularly evaluates the effectiveness of its program on student achievement. Including this study, CommonLit has published 12 ESSA Level 3 studies that provided "Promising Evidence" and one ESSA Level 2 study that provided "Moderate Evidence" for the efficacy of CommonLit. CommonLit is currently conducting a two-year, quasi-experimental study under the Education Innovation and Research grant program. This two-year study will examine the impact of CommonLit 360 on middle school students' state ELA test scores, using a comparative interrupted time series study design that addresses many limitations in the current study. Moreover, the comparison group will be identified prior to the collection of state test data and matched on baseline ELA scores and available school-level demographics to address baseline equivalence. When possible, information about instructional alternatives will be collected to understand the contrast between treatment and comparison groups. This study also incorporates surveys and interviews to understand teachers' experiences with implementing the curriculum. Additionally, findings will be analyzed by external evaluators at the Consortium for Policy Research in Education (CPRE) at Teachers College, Columbia University. This two-year study represents the most rigorous evaluation of CommonLit to date, and is eligible to meet ESSA Level 2 (Moderate Evidence) standards with positive findings. Future studies may examine the impact of CommonLit across additional states and on measures beyond student achievement. Through research partnerships and grants, CommonLit aims to evaluate the effectiveness of its programs for students and teachers.

References

Brown, M. (2014). *Investigating the effects of CommonLit on student engagement in middle school reading classrooms* (ERIC No. ED662965). Harvard Graduate School of Education. [Link]

CommonLit. (2024). *CommonLit usage and writing growth: SY 2022-2023* [Technical report]. [Link]

CommonLit. (2025). *Results from Harford County Public Schools (MD): School year 2023-2024* [Technical report]. [Link]

CommonLit. (2026). *Report on school year 2025-2026: Students using CommonLit saw significant reading gains* [Technical report]. [Link]

CommonLit & Mathematica. (2023). *Effects of CommonLit 360 on reading comprehension, student engagement, and cultural awareness* [Technical report]. [Link]

Fahle, E., Kane, T. J., Reardon, S. F., & Staiger, D. O. (2024). *The first year of pandemic recovery: A district-level analysis*. Education Recovery Scorecard. [Link]

Kuhfeld, M., & Lewis, K. (2025). *5 years after COVID-19 hit: Test data converge on math gains, stalled reading recovery*. Brookings. [Link]

Kuhfeld, M., Lewis, K., & Peltier, T. (2023). Reading achievement declines during the COVID-19 pandemic: Evidence from 5 million US students in grades 3–8. *Reading and Writing, 36*(2), 245–261. [Link]

Lee, A., & Shah, M. (2024). *CommonLit: Examining Impacts among Florida Middle School Students (2022-23): Study Type: ESSA Evidence Level II.* [Technical report]. [Link]

National Center for Education Statistics. (2025). *NAEP report card: Reading* [Web report]. U.S. Department of Education, Institute of Education Sciences. [Link]

New Meridian. (2024). *New Meridian summative assessments Spring 2024: New Jersey technical report*. New Jersey Department of Education. [Link]

New Meridian. (2025). *New Meridian summative assessments Spring 2025: New Jersey technical report*. New Jersey Department of Education. [Link]

Appendix

Table A1 presents descriptive statistics for digital CommonLit lesson usage and the NJSLA ELA outcome variable.

Table A1. Descriptive Statistics for Key Variables (*N* = 2,495)

Variable	Mean	*SD*	Median	Min	Max
Total digital lessons	19.30	12.05	18	1	127
CommonLit 360 lessons	9.25	8.25	10	0	34
Library lessons	6.04	9.10	3	0	90
Target lessons	4.00	5.26	2	0	37
Relative usage (vs. classroom average)	0.00	6.10	0.33	−34.20	76.12
NJSLA ELA scale score, 2024-2025	737.36	39.13	738	650	850
NJSLA ELA scale score, 2023-2024	732.10	40.53	733	650	850

Note. The interquartile range of relative usage was −2.06 to +2.63. Thirty students submitted more than 60 lessons, and five submitted more than 80.

Within-grade *z*-scores had a mean of 0 and *SD* of 1. We fit a sequence of multilevel models with a random intercept for classroom: an unconditional model to estimate the intraclass correlation (ICC); a model adding prior achievement and classroom mean prior achievement; a model adding the two usage components; and a final model adding grade and student characteristics (gender, race/ethnicity (Black, not Hispanic/Latino vs. all other students), special education status, LIEP eligibility, and gifted and talented status). This final model is reported in the Results section of the main report, while the others are reported below in Appendix Table A2.

Table A2. Progressive Models for the Continuous Outcome (Within-Grade *z*-Score)

Predictor	Unconditional	+ Prior achievement	+ Usage	Final (Table 4)
Intercept	−0.21 (0.07)	0.001 (0.03)	0.00 (0.03)	0.02 (0.06)
Prior achievement		0.73 (0.01)	0.73 (0.01)	0.69 (0.01)
Classroom mean prior achievement		0.26 (0.04)	0.26 (0.05)	0.23 (0.05)
Relative usage			0.01 (0.002)	0.01 (0.002)
Classroom average usage			0.002 (0.002)	0.003 (0.002)
Demographics and grade				Included
Classroom intercept variance	0.42	0.04	0.04	0.04
Residual variance	0.69	0.31	0.31	0.30
AIC (ML)	6381.9	4271.9	4259.7	4214.1
χ^2 vs. previous model		2114 (*df* = 2), *p* < .001	16.25 (*df* = 2), *p* < .001	65.60 (*df* = 10), *p* < .001
Adjusted ICC	.38	.12	.12	.12

Note. Cells show *b* (*SE*). Relative usage was significant ($p < .001$) in the + Usage and Final models; classroom average usage was not significant in either. AIC values are from maximum likelihood fits, and χ^2 tests compare each model with the model in the previous column.

To assess robustness, we refit both models after excluding (a) classrooms with fewer than 10 students (18 classrooms, 113 students), (b) students who submitted more than 80 lessons ($n = 5$), and (c) five classrooms flagged as influential in the continuous model by Cook's distance (> 4/84, where 84 is the number of classrooms; 312 students), re-centering predictors in each restricted sample. Table A3 shows the results for the continuous outcome, and Table A4 shows the results for the binary outcome. We also tested whether the usage association varied across classrooms by adding a random slope for relative usage to the continuous model (final row of Table A3).

Table A3. Sensitivity Analyses for the Continuous Outcome

			Relative usage		Classroom average usage	
Model	**Students**	**Classrooms**	**Estimate (*SE*)**	***p***	**Estimate (*SE*)**	***p***
Final model	2,495	84	0.01 (0.002)	< .001	0.003 (0.002)	.236
Excluding classrooms with < 10 students	2,382	66	0.01 (0.002)	< .001	0.002 (0.003)	.446
Excluding students with > 80 lessons	2,490	84	0.01 (0.002)	< .001	0.004 (0.002)	.125
Excluding 5 influential classrooms	2,183	79	0.01 (0.002)	< .001	0.003 (0.002)	.228
Random slope for relative usage	2,495	84	0.01 (0.002)	< .001	0.004 (0.002)	.074

Note. Predictors were re-centered within each restricted sample. Influential classrooms were identified using Cook's distance (> 4/84) in the continuous model. The random-slope model produced a singular fit (likelihood ratio test vs. final model: $\chi^2(2) = 4.79$, $p = .091$).

Table A4. Sensitivity Analyses for the Binary Outcome

			Relative usage		Classroom average usage	
Model	**Students**	**Classrooms**	**OR [95% CI]**	***p***	**OR [95% CI]**	***p***
Final model	2,495	84	1.04 [1.01, 1.06]	.001	1.01 [1.00, 1.03]	.125
Excluding classrooms with < 10 students	2,382	66	1.04 [1.01, 1.06]	.002	1.01 [0.99, 1.03]	.196
Excluding students with > 80 lessons	2,490	84	1.04 [1.02, 1.07]	.001	1.02 [1.00, 1.04]	.104
Excluding 5 influential classrooms	2,183	79	1.04 [1.02, 1.07]	.001	1.01 [0.99, 1.03]	.235

Note. Predictors were re-centered within each restricted sample. Odds ratios are per additional lesson. The five influential classrooms were identified from the continuous model (Cook's distance > 4/84); two of them were also flagged in the model for the binary outcome.

To test moderation, we added interactions between relative usage and prior achievement, gender, race/ethnicity, special education status, LIEP eligibility, and gifted and talented

status (one model per outcome). The adjusted *p*-values for these 12 interaction tests using the Holm procedure are reported in Table A5.

Table A5. Interactions Between Relative Usage and Student Characteristics

	Continuous		Binary	
Interaction with relative usage	**Estimate (*SE*)**	**Holm *p***	**Estimate (*SE*)**	**Holm *p***
Prior achievement	−0.001 (0.002)	1.000	−0.02 (0.02)	1.000
Male	−0.002 (0.004)	1.000	−0.01 (0.02)	1.000
Black, not Hispanic/Latino (vs. all other students)	0.003 (0.005)	1.000	0.03 (0.03)	1.000
Special education	0.004 (0.006)	1.000	0.03 (0.04)	1.000
LIEP eligible	−0.001 (0.006)	1.000	−0.04 (0.04)	1.000
Gifted and talented	−0.004 (0.007)	1.000	0.11 (0.05)	.613

Note. Each interaction was tested in a model that included all interactions with relative usage along with the final-model predictors. Continuous estimates are *b* (*SE*) on the within-grade *z*-score scale; binary estimates are *b* (*SE*) on the log-odds scale. Holm adjustment was applied across all 12 tests (6 per outcome).

Model Diagnostics

For the continuous model, residuals were approximately normal apart from modest departures in the tails, showed no systematic pattern across fitted values or grades (residual *SD*s ranged from 0.49 to 0.59), and classroom random intercepts were approximately normal with modest tail departures. The final model's marginal R^2 was .67 and its conditional R^2 was .71. For the logistic model, a binned residual check gave mixed results (94% of 50 equal-sized bins within binomial error bounds; 68% under the default settings of the performance package). Allowing the prior achievement slope to vary by grade or modeling it with a spline did not improve the logistic model's fit (both $p > .25$) and did not change the estimate for relative usage (OR = 1.04), and classroom random intercepts had heavier tails than a normal distribution, driven largely by one classroom that was also flagged as influential in the continuous model. Multicollinearity was low in both models (all variance inflation factors ≤ 1.31). Five classrooms exceeded the Cook's distance cutoff in the continuous model, and two of them also exceeded it in the logistic model, including the classroom with the highest logistic intercept. Excluding all five did not change the results for either outcome (Appendix Tables A3 and A4).