



COMMONLIT 360 EFFECTIVENESS REPORT

A STUDY OF COMMONLIT 360 IN NYC MIDDLE SCHOOLS
DURING A YEAR OF REMOTE LEARNING, SCHOOL YEAR 2020-21



DECEMBER 2021



TABLE OF CONTENTS

Executive Summary	2
About the CommonLit 360 Program	3
Study Context: A Pandemic Year in NYC	5
Research Questions	6
Methods	7
Study Findings	8
Acknowledgments	13
Appendix A: Descriptive Statistics by School and Grade Level	14
Appendix B: Hierarchical Linear Model for Investigating Finding #2	18
Limitations and Recommendations	25



EXECUTIVE SUMMARY

During the school year 2020-2021, CommonLit worked with ten middle schools in New York City to study the effectiveness of its comprehensive program for English Language Arts, CommonLit 360, Edition 1.0. The 360 program offers curriculum, teacher training, and integrated assessments.

The study sample included 1,022 students from seven high-need schools in New York City with sufficient data. Participating teachers taught the CommonLit 360 Curriculum and utilized the technology platform during remote, in-person, and hybrid instruction. A valid reading pre-test and post-test was utilized to measure student reading growth from the beginning to the end of the school year. Backend data from the CommonLit platform was used to gauge utilization of the program.



Evaluators found that:

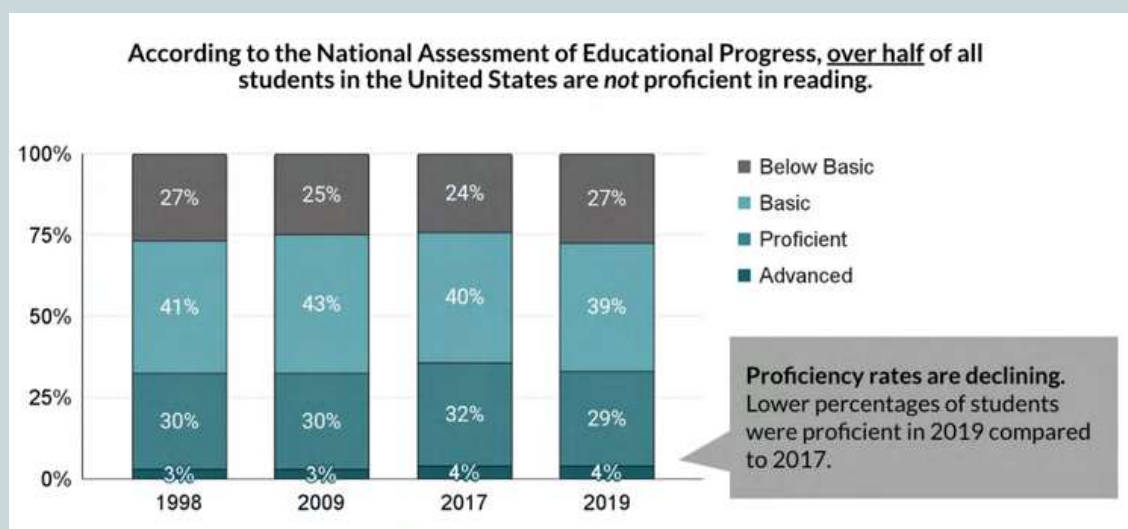
- 1. Students showed a statistically significant increase in reading scores from pre-test to post-test across every grade,**
- 2. Utilization of the 360 program, measured by lesson completion, was a statistically significant predictor of student gains in reading, and**
- 3. Gains for English Language Learners and students with disabilities was similar to students without those designations.**

Across the year, reading gains from pre-test to post-test were positive and statistically meaningful, and fidelity to the program mattered. These small gains in student achievement across grades 6-8 should be viewed within the larger context of a pandemic year, marked by severe learning loss estimated at anywhere from 4-6 months for reading, with disproportionately greater loss for communities of color, and particularly for Black students ([Dorn et al., 2021](#)).



ABOUT THE COMMONLIT 360 PROGRAM

CommonLit 360 was developed as a solution to a problem. Despite the manifest importance of literacy, standard classroom practice in the United States is not sufficient to ensure that all students learn to read and write. According to the Nation's Report Card, over half of all students in the United States are not proficient in reading ([NAEP, 2019](#)). Writing outcomes are also extremely poor; approximately two-thirds of students in grades 8 to 12 scored at or below the basic level on a writing test administered by the National Assessment of Educational Progress (National Center for Educational Statistics, 2012).



Black, Hispanic, Native students, as well as economically disadvantaged students are disproportionately affected by the literacy opportunity gap. Eighty percent of low-income students are unable to read proficiently ([NAEP, 2019](#)). At the same time, students spend half the school year on work that is not at grade level and report feeling bored and disengaged in school work over half the time (The New Teacher Project, 2018). Recent reports suggest that the COVID-19 pandemic set vulnerable students back anywhere from four to six months in reading ([Dorn et al., 2021](#)).



ABOUT THE COMMONLIT 360 PROGRAM

To dramatically change student literacy outcomes at scale, CommonLit developed CommonLit 360, a whole school English Language Arts program with curriculum that is [freely and openly available for grades 6-10](#) to ensure broad access. CommonLit 360 has four main features: (1) curriculum to last a full year of instruction, (2) assessments to inform targeted instruction, (3) on-demand teacher professional learning, and (4) data dashboards for teachers, administrators, and instructional coaches that show student progress in real time.



These components are delivered on a cutting edge digital platform, www.commonlit.org, which currently serves between 7 and 10 million students regularly each month. The 360 innovation is meant to bring coherence to schools, and ensure connectedness between curriculum, training, and assessment – the key features of a strong academic program.

The 360 program is designed to ensure that all students have access to high-quality instruction that is on grade level. [Edition 1.0](#) of the 360 program includes important features to enhance accessibility for students with disabilities and students whose first language isn't English, with options for audio, translation into nearly 40 different languages, and support for English Language Learners like explicit vocabulary instruction and supplemental texts in Spanish. Within Edition 1.0, there are six instructional units per grade, each lasting 5-7 weeks. Within each of the instructional units, students learn about topics that spark genuine curiosity and engagement, gain important vocabulary and knowledge, and explore enduring themes in literature. Within a classroom, the program is designed to help teachers maximize instructional time, minimize teacher talk, and increase the amount of time students spend reading, writing, discussing, and engaging intellectually.

CommonLit 360 was developed iteratively, through extensive feedback over the course of three years in 50 pilot schools based all over the country.



STUDY CONTEXT: A PANDEMIC YEAR IN NYC

During the 2020-2021 school year, CommonLit 360 represented a timely solution during a difficult year marked by frequent closures, hybrid learning, staffing shortages, and health challenges related to COVID-19. All ten partner schools opted to continue piloting the CommonLit 360 program, and agreed to support our study of its effectiveness. During the 2020-2021 school year, participating teachers utilized and relied on the 360 program heavily while schools shifted in and out of remote learning.

Several features of that are unique to the CommonLit 360 program helped to ensure continuity of learning in remote, in-person, and hybrid contexts:

- **Interactive Student Platform:** To help teachers digitally monitor student work and provide feedback through the platform asynchronously.
- **Integrated Learning Management System:** To digitally assign materials, and manage grading, missing assignments, and more – all through the CommonLit platform.
- **Real-Time Data:** To give teachers insight into student performance, in real time while they are working on an assignment.
- **24/7 Tech Support:** Offered via live chat and email to help teachers troubleshoot technical challenges that can disrupt learning.
- **Virtual Training:** Delivered via webinar and video, and on demand, to be viewed whenever is convenient.
- **Mobile Optimization:** The entire platform and all activities are compatible with mobile devices to support students who are sharing devices or access school via smartphone.
- **Intermittent Internet Access Support:** Including by autosaving student responses, to ensure continuity of learning even during outages.

Further context about the 2020-2021 school year is important to consider. 62.8% of NYC public school students were either fully remote or hybrid during the 2020-2021 school year. The New York State Education Department reported that during SY 2020-2021, about ¹8 percent of New York students lacked devices, and about 6 percent lacked sufficient internet access ([NYSED](#)). The data revealed that districts serving majority Black or Hispanic students were four times as likely to have inadequate or no internet in SY2020-21 ([NYSED](#)).

¹ According to [this source](#), 650,000 NYC public school students opted into remote learning in April 2021. These 650,000 students comprise 62.8% of K-12 students enrolled (1,033,669 students) in NYC public schools during SY 2020 - 2021.



RESEARCH QUESTIONS

1

Did students' reading scores improve from beginning to end of year through the use of the CommonLit 360 reading program?

2

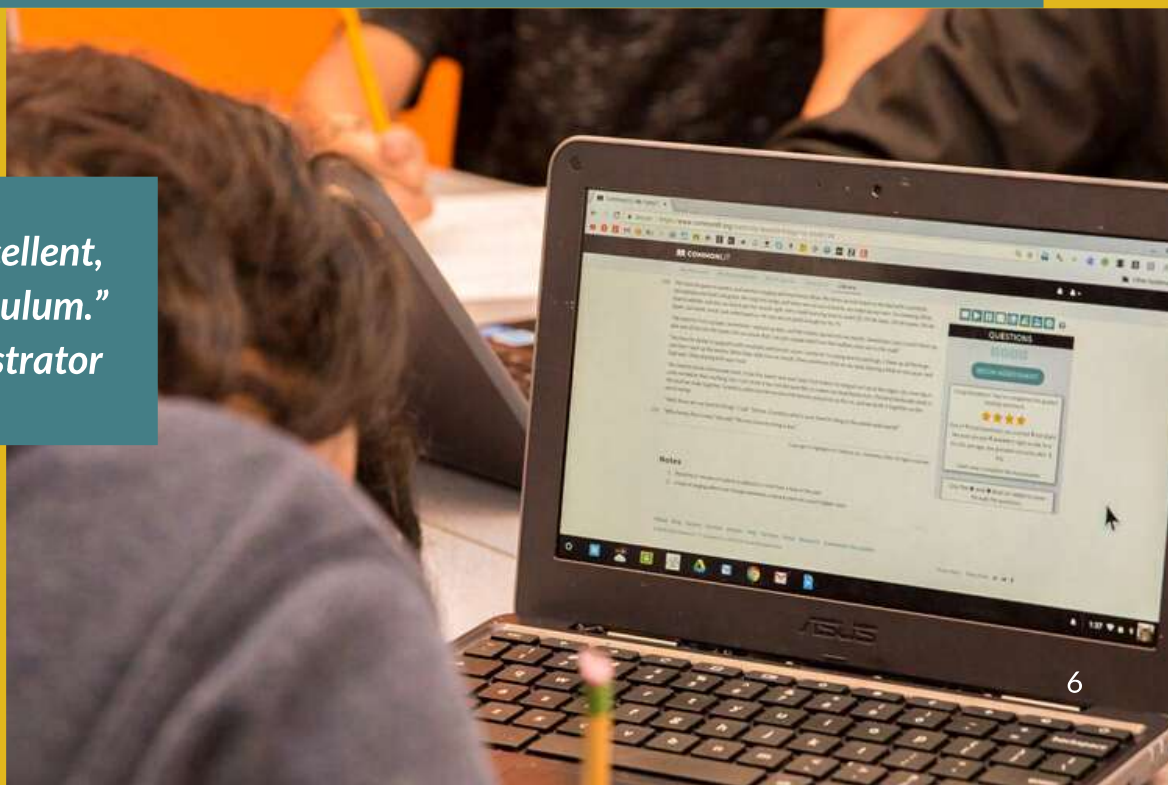
How does utilization of CommonLit 360 materials and lessons impact student reading gains?

3

How did these reading gains differ for special populations of students, including students receiving special education services and students designated as English Language Learners?

*"[CommonLit 360] has been an amazing resource for our teachers and students. We have appreciated the clarity, organization, and accessibility of the resources."
- NYC School Administrator*

*"CommonLit is an excellent, comprehensive curriculum."
- NYC School Administrator*





METHODS

What were the characteristics of the schools in the study?

CommonLit partnered with ten New York City public schools across four boroughs during the school year 2020-2021 to implement CommonLit 360. All schools opted into the program. All ten schools were large city public schools classified as Title I schools and had a school-wide Title I program. Across the schools, the percentage of students who were eligible for free lunch ranged from 63% to 94% (median = 85%). The percentage of students who were eligible for free or reduced lunch ranged from 71% to 96% (median = 88%).

All ten schools served a large population of high-need, at-risk students. All but one of the schools enrolled more than 88% Black and Hispanic students. For those schools, each of the remaining racial categories consisted of 1 to 4% students (i.e., Native American, Asian, Pacific Islander, White, or two or more races). One school slightly differed in racial composition, enrolling 54% Black and Hispanic students, 29% Asian students, 11% Native American students, and the remaining percentage a mixture of other racial categories. For the purposes of this study, we had complete data in seven of the ten schools.

How did we measure reading growth?

Valid pre- and post-assessments, developed by CommonLit, were utilized to measure reading growth from beginning to end of year. CommonLit Growth Assessments are scored on a valid IRT scale (150 to 250), with each grade on a discrete scale (e.g. a 175 on a 6th grade assessment is not equivalent to a 175 on a 7th grade assessment). Growth is calculated based on points gained on this scale from pre-test to post-test.

Scaled scores take item difficulty into account, and weight those items accordingly. Items on the assessments have undergone rigorous field testing, rounds of revision, psychometric evaluation, and expert review. Assessments have met the threshold of 0.8 reliability, an industry standard. Previous research has shown that CommonLit Growth Assessments are highly associated with standardized state test scores, like the Florida Standards Assessment (FSA).

CommonLit collected the baseline pre-test data in August-September 2020, and collected post-test data in June 2021. State standardized test data were not available during SY 2020-2021 due to COVID related test cancellations.



STUDY FINDINGS

FINDING 1: STUDENTS SHOWED A STATISTICALLY SIGNIFICANT INCREASE IN READING SCORES FROM THE BEGINNING TO END OF THE SCHOOL YEAR, ACROSS ALL GRADES.

SUMMARY OF FINDINGS

Students' average CommonLit Growth Assessment scores increased from the beginning to end of the 2020-2021 school year, across all grades. The average (aggregate) gain was small, but statistically significant.

Further examination of class and individual student gain illustrated that many students saw dramatic gains, and some classes saw outsized gains unexpected during a pandemic year.

Students participating in the CommonLit 360 program showed statistically significant gains in reading scores on average from pre-test to post-test. These gains were seen in all schools, and across every grade level. The average gain was 2.76 scaled score points.² In the context of a pandemic year, even a small gain is notable, considering that learning loss in reading is estimated to be four to six months, with vulnerable students bearing the brunt of the effect of school closures ([Dorn et al., 2021](#)).

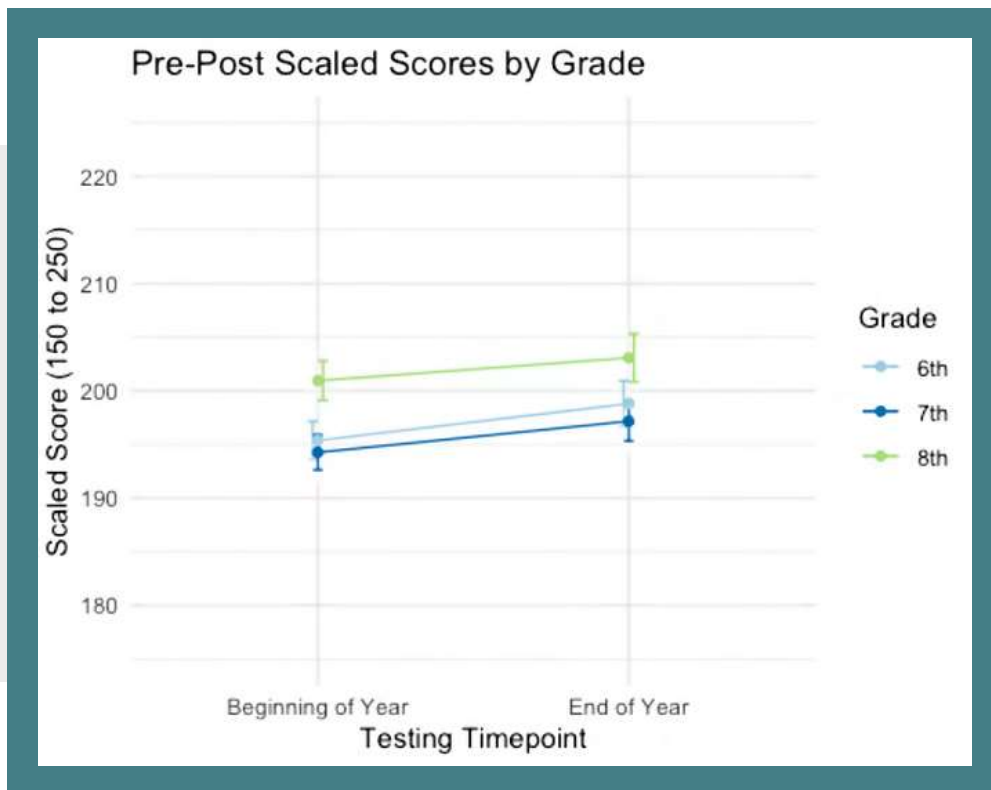


² On the standardized metric (Cohen's d), this is within the range of middle-school year-to-year reading gain scores on seven nationally-normed tests examined by Bloom et al. (2008), Table 1, page 10 (<https://files.eric.ed.gov/fulltext/ED503202.pdf>)



STUDY FINDINGS

Gain scores were also analyzed individually by school and grade level (6th, 7th, and 8th grade). In all but two classes evaluated, there was an increase in average reading scores from the beginning to the end of the school year. Many classes saw dramatic gains. In fact, 32% percent of the students gained 10 or more scaled score points, and 11.4% gained 20 or more scaled score points on the CommonLit Growth Assessment. The highest average gain was 13.74 points in one school's 6th grade class. Additional details on individual class analysis may be found in Appendix A.



Students participating in the CommonLit 360 program showed gains from pretest to posttest, 2.76 scaled score points on average. Note: Each grade uses a discrete scale.



STUDY FINDINGS

FINDING 2: TEACHER UTILIZATION OF COMMONLIT 360 MATERIALS WAS A MEANINGFUL PREDICTOR OF STUDENT READING GAINS.

SUMMARY OF FINDINGS

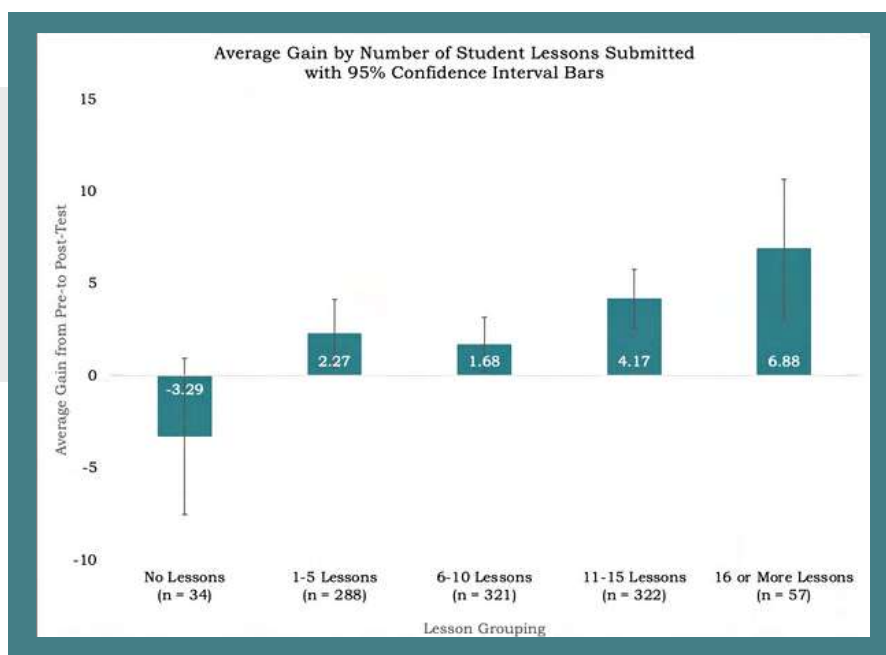
Teacher utilization of CommonLit materials and student submission of assignments was a statistically significant predictor of student outcomes. The more CommonLit 360 lessons that students submitted, the greater the gains from beginning to end of year on reading assessments. Students who submitted ten or more CommonLit 360 lessons showed a statistically significant higher gain.

The effect of teacher utilization was significant, even when accounting for student level characteristics including English Language Learner designation, special education designation, and pre-test scores.

Students who submitted more than ten assignments showed a greater average gain in reading scores than students who submitted fewer assignments. As the number of lessons increased, students' average gain from pre- to post-test also increased. Average reading assessment score gain for students submitting more than ten assignments was statistically significantly higher than for those who did not submit assignments.³

Students submitting more than ten CommonLit 360 lessons showed statistically significantly higher gain than students who did not submit lessons.

Technical details may be found in Appendix B.



³ Gain scores across the categories differed statistically significantly, $F(4,1017) = 3.85$, $p = .004$, $\eta^2 = 0.015$. Tukey's post-hoc findings suggested a statistically-significant difference in gain scores for students who submitted no assignments and those who submitted more than ten assignments (i.e., the 11 to 15 and the 16 + categories).



STUDY FINDINGS

FINDING 3: STUDENTS WITH DISABILITIES AND STUDENTS CLASSIFIED AS ENGLISH LANGUAGE LEARNERS SHOWED GROWTH ON READING ASSESSMENTS NEARLY ON PAR COMPARED TO THEIR PEERS WITHOUT THESE DESIGNATIONS.

SUMMARY OF FINDINGS

Students with disabilities and students classified as English Language Learners showed growth on reading assessments nearly on par compared to their peers without these designations. A student's special education or English Language Learner designation was not a statistically significant predictor of gains from beginning to end of year.

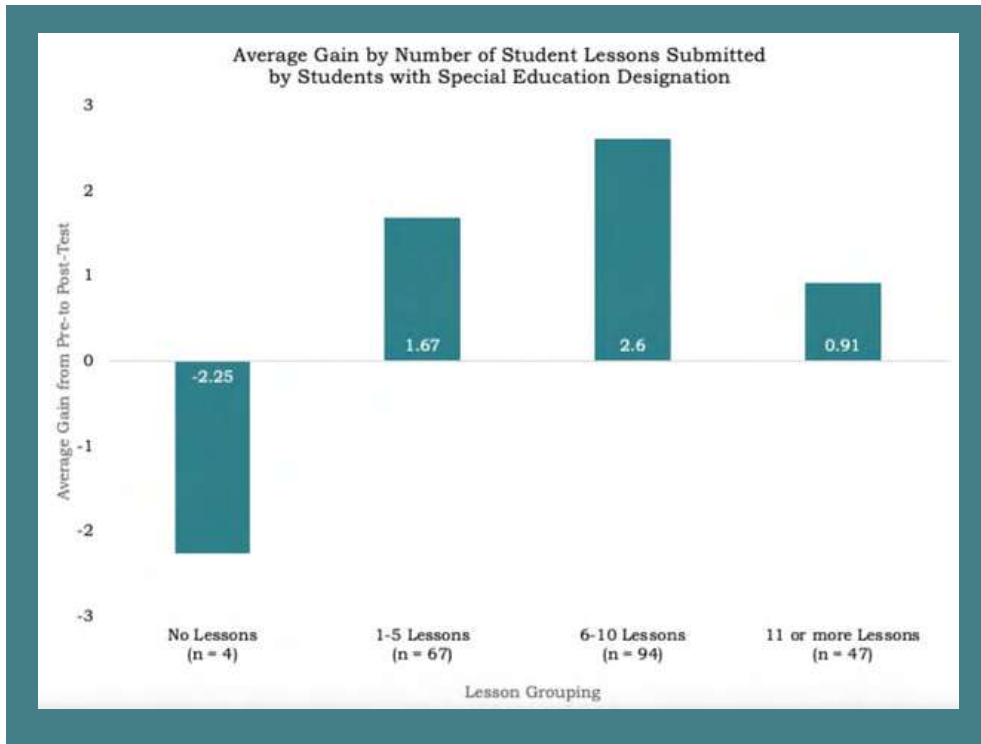
Students receiving special education services showed growth on reading assessments nearly on par with their peers who were not receiving special education services. Average reading gain for students receiving special education services was positive, a gain of 1.84 scaled score points. However, given high variability in students' gain scores, the findings should be interpreted cautiously. Similarly, students with English Language Learner designation showed gains similar to the gains made by their peers without the designation. Average reading gain scores for students with English Language Learner designation were also positive, a gain of 1.95 scaled score points. Given high variability of the gain scores, however, it is important to approach these findings cautiously as well.

In the final model, CommonLit usage predicted reading gain, even when accounting for English Language Learner and Special Education designations, pre-test score, and differences in grade (6th, 7th, 8th), teacher, and school. English Language Learner and Special Education designation did not have a unique impact on scaled gain scores, meaning these students had similar success with CommonLit 360 when compared to the general population of students.



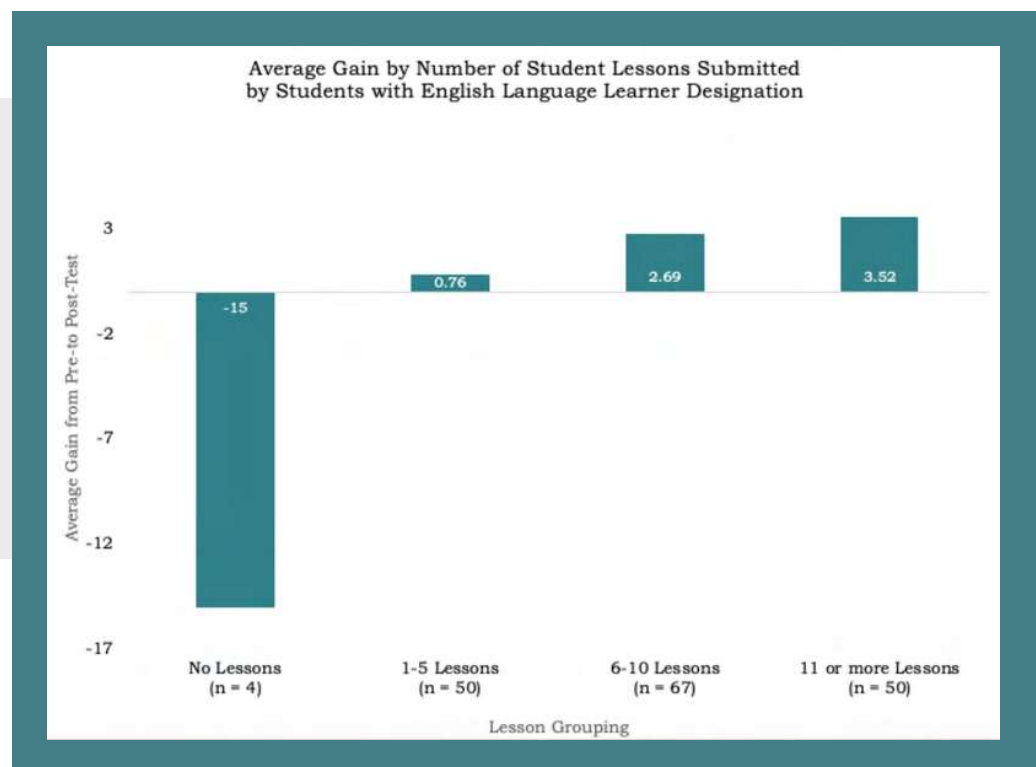


STUDY FINDINGS



On average, positive reading gain was seen for students receiving special education services who submitted one or more CommonLit 360 lessons.

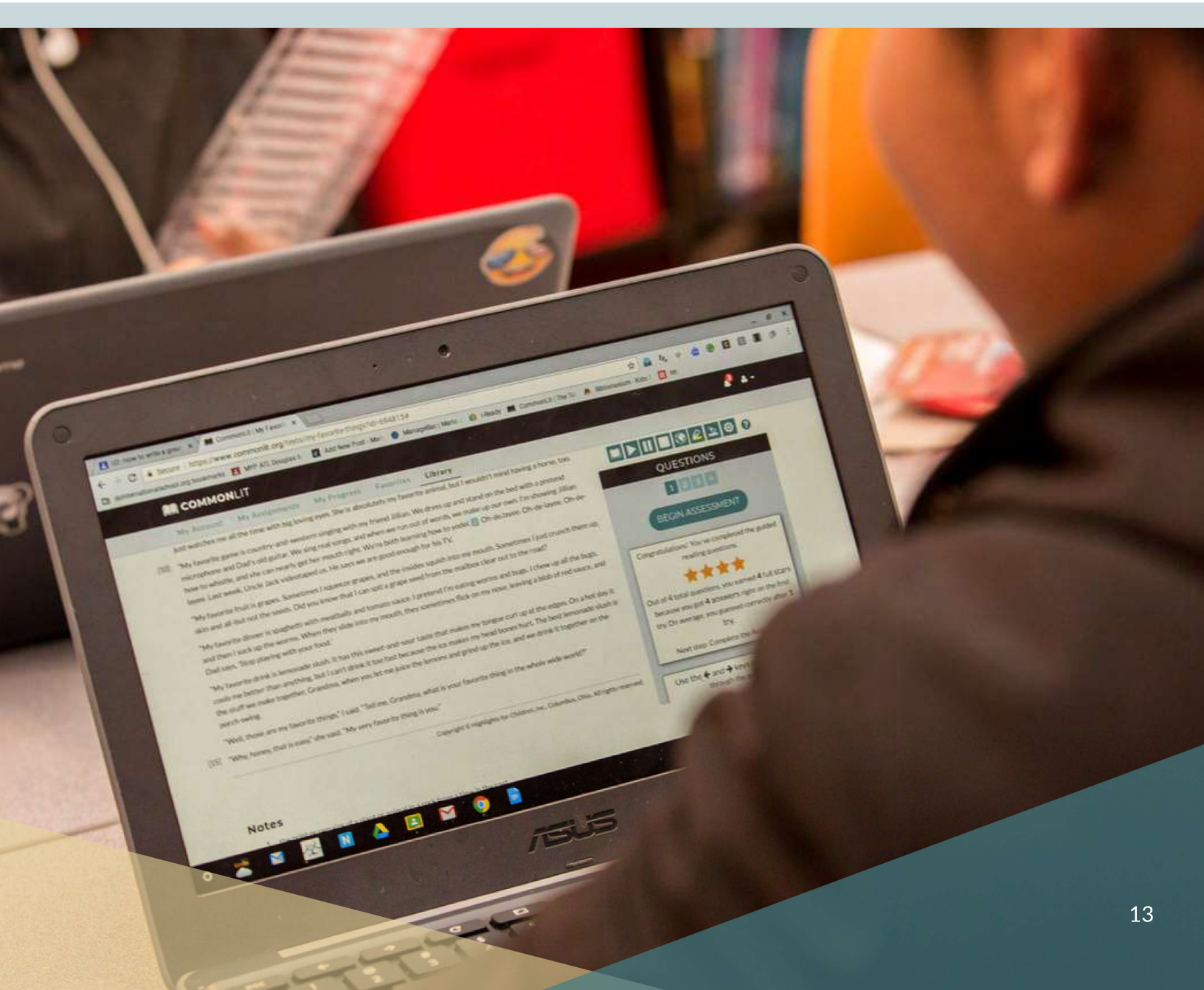
On average, positive reading gain was seen for students with English Language Learner designation who submitted one or more CommonLit 360 lessons.





ACKNOWLEDGEMENTS

This report would not be possible without the generous support of CommonLit’s philanthropic donors, including the Robin Hood Learning + Technology Fund.





APPENDIX A:

DESCRIPTIVE STATISTICS BY SCHOOL AND GRADE LEVEL

Of the 21 classes studied, 19 showed gains from pre-test to post-test. Of the two instances in which there was no gain, one school's 8th grade class decreased by 0.23 points and another school's 7th grade class by 4.81 points.⁴ Table 1 in Appendix A summarizes individual gain scores by school and grade level. Note the large standard deviations, suggesting considerable variability in gain within schools and classes. Considering the low-stakes nature and end-of-year administration of post-tests during a pandemic year, it is probable that gain scores are attenuated.⁵

The aggregate average score does not capture the variability in individual students' scores. Many students experienced substantial gain across the school year while some showed decline from pre-test to post-test. The scatterplot figures, below, illustrate four classes in which the majority of students showed dramatic growth from the beginning to end of the school year. Dots above the horizontal '0' line on the scatterplots, below, indicate gain (i.e., students whose assessment scores increased from pre- to post-test). In sum, there was positive growth for most students in many of the classrooms, even during a year in which some experts have estimated a four to six month decline in reading ([Dorn et al., 2021](#)).



⁴ Anecdotal evidence from the class with the 4.81 point decrease was that the students took two standardized tests earlier in the same week. Test burnout likely had an impact on scores. It was communicated that the CommonLit post-test assessment was low stakes comparatively.

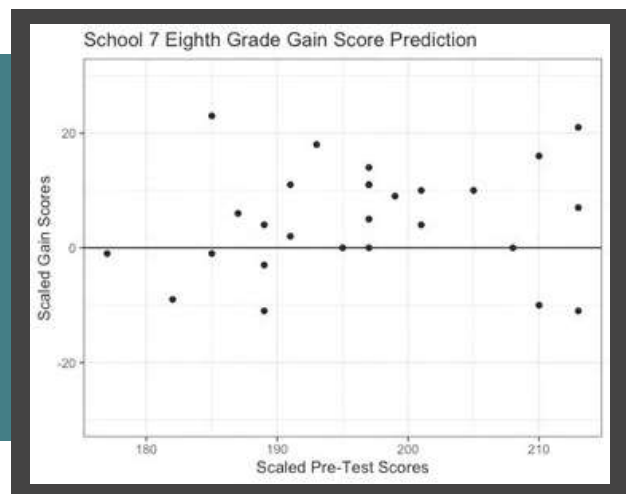
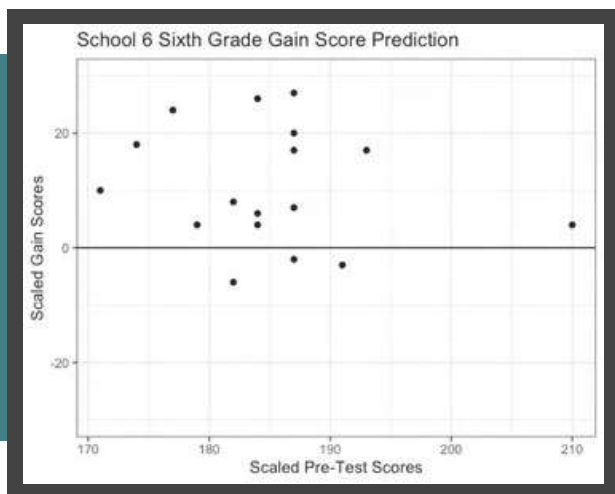
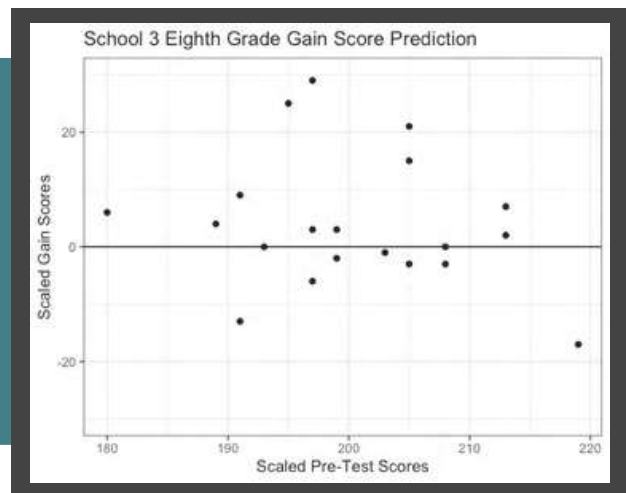
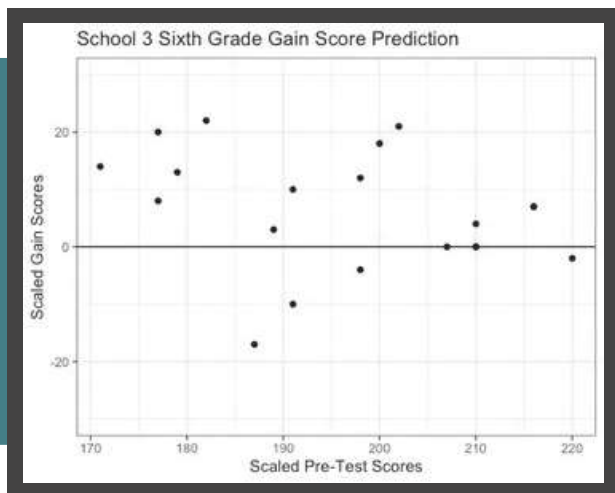
⁵ By mathematical necessity, we also expect regression toward the mean where, on average, low scores will rise on post-test ('gain') and high scores will be lower on post-test ('decline').



APPENDIX A:

DESCRIPTIVE STATISTICS BY SCHOOL AND GRADE LEVEL

Examples of classes in which most students' assessment scores increased from the beginning to end of the school year

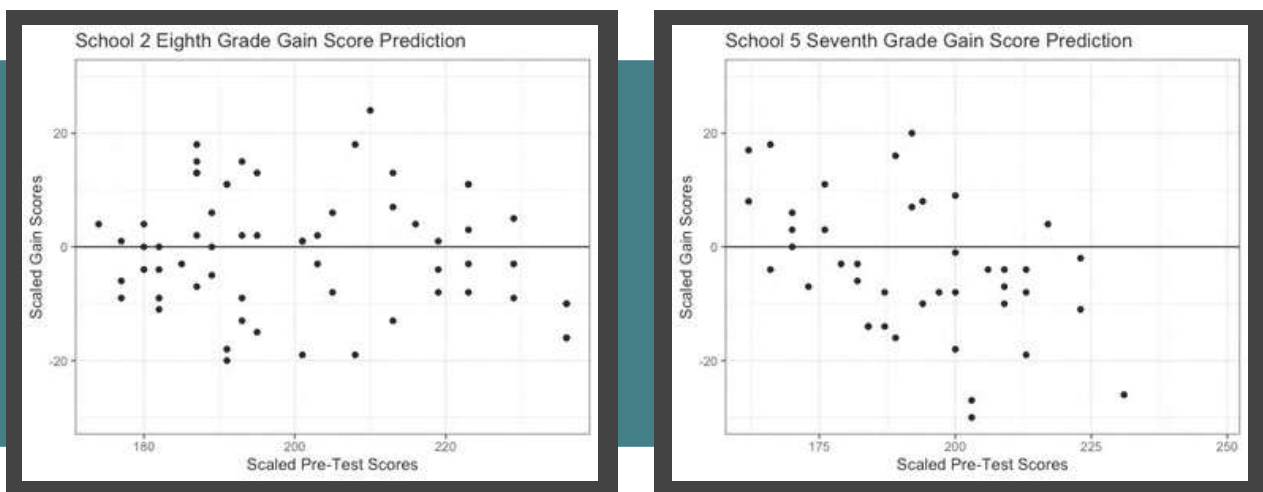




APPENDIX A:

DESCRIPTIVE STATISTICS BY SCHOOL AND GRADE LEVEL

In contrast, as may be seen in the scatterplots below, in some classes there was a greater mix of student score gains and declines from pre- to post-test. These examples of high variability classrooms illustrate the importance of understanding factors related to student growth. The scatterplots below are for the two classes with negative gain scores.⁶

Examples of classes with average assessment scores decline

Given the variation in student growth across school and grade level, we examined factors that could help us better understand student reading growth (e.g., pre-test scores, teacher utilization, and English Language Learner and Special Education designations). Table 1, below, includes means and standard deviations of teacher utilization, scaled pre-test and post-test scores, and gain scores.

⁶ Average gain scores for School 2 Eighth Grade (left) was -0.23, and average gain scores for School 5 Seventh Grade (right) was -4.81.



APPENDIX A:

DESCRIPTIVE STATISTICS BY SCHOOL AND GRADE LEVEL

TABLE 1

Mean and Standard Deviations (SD) of Utilization, Scaled Pre-Test, Scaled Post-Test, and Gain Scores (N= 1,022)

School	Grade	n	Utilization		Scaled Pre-Test		Scaled Post-test		Gain	
			Mean	SD	Mean	SD	Mean	SD	Mean	SD
1	6	---	---	---	---	---	---	---	---	---
1	7	15	0.76	0.20	190.33	14.91	194.33	15.58	4.00	10.70
1	8	19	0.81	0.24	211.68	20.61	214.26	17.68	2.58	11.49
2	6	70	0.56	0.28	189.34	14.94	195.59	19.48	6.24	15.34
2	7	108	0.62	0.31	190.92	18.81	194.81	18.09	3.90	13.75
2	8	62	0.67	0.32	201.19	17.82	200.97	19.95	-0.23	11.14
3	6	20	0.52	0.22	196.55	14.78	202.85	15.42	6.30	10.41
3	7	13	0.66	0.37	196.08	14.67	201.77	20.64	5.69	14.40
3	8	22	0.64	0.34	199.32	9.58	204.27	19.04	4.95	18.25
4	6	115	0.52	0.32	197.85	16.81	198.86	19.84	1.01	17.20
4	7	151	0.64	0.31	197.34	15.76	199.72	20.11	2.37	14.94
4	8	167	0.65	0.28	201.32	16.98	203.44	21.33	2.11	14.10
5	6	73	0.75	0.24	198.26	16.18	198.74	19.86	0.48	17.35
5	7	42	0.89	0.17	193.88	19.76	189.07	16.26	-4.81	13.52
5	8	---	---	---	---	---	---	---	---	---
6	6	19	0.76	0.15	185.26	9.21	199.00	15.22	13.74	13.76
6	7	19	0.65	0.37	185.37	10.97	187.63	15.36	2.26	11.08
6	8	17	0.63	0.31	191.35	12.58	195.24	19.67	3.88	18.65
7	6	11	0.85	0.23	207.00	10.84	207.73	19.77	0.73	14.39
7	7	52	0.88	0.22	196.87	12.96	204.19	15.62	7.33	12.66
7	8	27	0.91	0.17	196.70	10.04	201.74	14.57	5.04	9.37

Note: Utilization is the average proportion of lessons submitted by students out of the total number of lessons assigned by the teacher. Pre-test and Post-test scores are on a 150 to 250 scale. Gain was calculated by subtracting students' scaled pre-test score from their scaled post-test scores. A positive score indicates average gain for that school and grade level, whereas a negative score indicates decline.



APPENDIX B:

HIERARCHICAL LINEAR MODEL FOR INVESTIGATING FINDING #2

For purposes of examining Finding #2, we estimated two hierarchical linear models (HLM; random intercept models) predicting student reading from a set of student-level and classroom-level predictors. Rather than traditional ordinary least squares regression models, the HLM models were chosen in order to account for the fact that students were nested within classrooms and schools. Findings from the two models are reported below.

MODEL 1

Research question: How does utilization of CommonLit 360 materials and lessons impact student reading gains?

Note: In Model 1, utilization was operationalized as the proportion of lessons submitted by the student out of the total number of lessons assigned by the teacher.

Outcome: The outcome for the Model 1 HLM analysis was students' CL360 assessment gain score on the scaled score metric. Specifically, gain scores were calculated by subtracting individual students' scaled pre-test score from their scaled post-test scores. Positive gain scores indicate an increase in scores from pre- to post-test. Negative gain scores indicate a decrease in scores from pre- to post-test. A gain of '0' indicates no gain. The gain score means and standard deviations for each school and grade level may be found in Appendix A, Table 1.



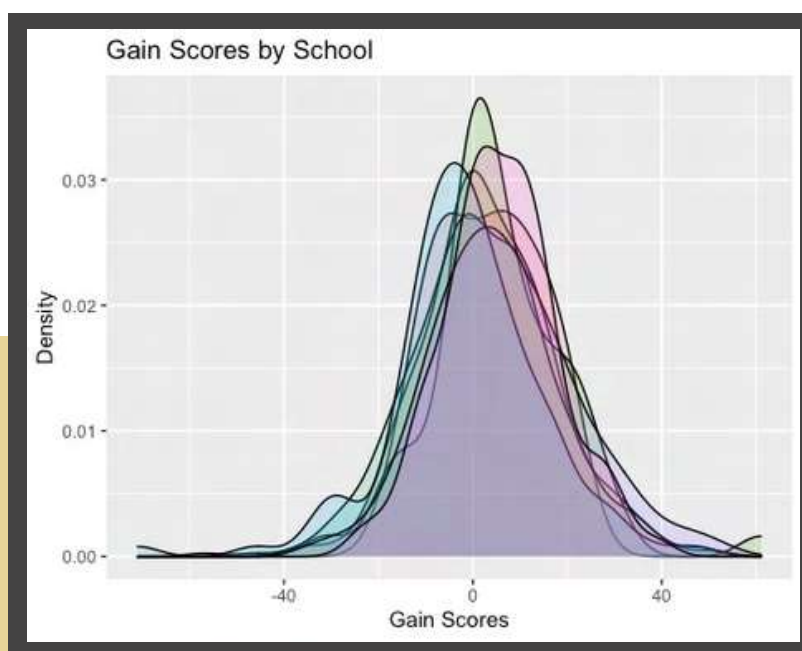


APPENDIX B:

HIERARCHICAL LINEAR MODEL FOR INVESTIGATING FINDING #2

Student-Level Predictors: Four student-level predictors were included in the model: scaled pre-test scores, English Language Learner status, Special Education status, and teacher utilization (aka fidelity). Students' scaled pre-test scores (scale = 150 to 250) were the students' scores from the beginning of the year assessment test ('pre-test'). English Language Learner and Special Education classifications were coded as 0 or 1, where 1 signified English Language Learner or Special Education student designation, respectively. Teacher usage of CommonLit (i.e., fidelity) was defined as the proportion of lessons submitted by the student out of the lessons assigned by the student's teacher. A high value suggested high teacher utilization, and vice versa. Scaled pre-test and teacher utilization means and standard deviations for each school and grade level may be found in Appendix A, Table 1.

School-level Predictors: There were two forms of classroom-level predictors included in the model: school and grade. School was dummy-coded with School 4 serving as the reference school. Grade was also dummy-coded with Grade 8 serving as the reference school. The intra-class correlation for the school effect was 0.026. Note the density plot, below, illustrating gain score distributions by school.





APPENDIX B:

HIERARCHICAL LINEAR MODEL FOR INVESTIGATING FINDING #2

FINDINGS

Model 1: Findings from the Model 1 HLM analysis may be found in Table 2, below. Note that there were two statistically-significant student-level predictors of gain scores: Pre-Test Score and teacher utilization. Pre-test scores were a negative predictor of gain scores. In other words, those who scored higher on pre-test tended to show less (or negative) gain, and those who scored lower on pre-test tended to show higher gain. This finding is consistent with ceiling effects found in educational research and past evaluation studies of CommonLit. That is, students with higher initial scores have less room for growth compared to students with lower initial scores.

Of importance to this study, is that once controlling for the other predictors in the model, **teacher utilization statistically significantly predicted student gain.** This finding is consistent with past CommonLit evaluation reports which found higher student usage to be associated with greater performance gains. Of the variance explained by the model, student-level predictors explained 93.3% of that variance.

Although the other student- and classroom- variables were not statistically significant, there is good news with this type of finding. For example, we see that designation of English Language Learner and Special Education status did not have a unique impact on scaled gain scores, meaning that students with either designation showed no statistically significant differences in growth or decline, compared to their peers.





APPENDIX B:

**HIERARCHICAL LINEAR MODEL
FOR INVESTIGATING FINDING #2**

TABLE 2

Student- and classroom-level effects of CommonLit 360 participation

Student-Level	Coefficient	Standard Error	Statistical Significance
Pre-Test Score	-0.27	0.03	***
English Language Learner	-1.83	1.87	
Special Education	-2.13	1.68	
Teacher Utilization	5.74	2.16	**
Intercept	2.98	0.73	***
Classroom-Level			
School 1	3.23	3.90	
School 2	1.67	1.95	
School 3	3.48	2.73	
School 5	-2.71	2.45	
School 6	1.71	2.89	
School 7	2.76	3.01	
Grade 6	2.03	1.88	
Grade 7	-0.30	1.82	
Variance Explained	Total	Proportion of Variance Explained by Variables	
		Kraft & de Leeuw (1998)	Snijders & Bosker (1999)
Student-Level (e)	93.3%	9.2%	7.4%
Classroom-Level (u)	6.7%	0.0%	0.0%

Note: Reference categories for this HLM model were School 4 and Grade 8. Total number of students included in the analysis was $n=1,014$. The 1,014 sample size was eight fewer students than the other analyses in this report, because scores from students who completed pre- and post-assessments less than 90 days apart were deleted from the data set for validity reasons.

* $p < .05$; ** $p < .01$; *** $p < .001$



APPENDIX B:

HIERARCHICAL LINEAR MODEL FOR INVESTIGATING FINDING #2

MODEL 2

Research question: How does utilization of CommonLit 360 materials and lessons impact student reading gains?

Note: In Model 2, utilization was operationalized as the total number of lessons that were submitted by the student.

Outcome: As with Model 1, the outcome for the Model 2 HLM analysis was students' gain score, from pre- to post-test. Gain scores, scaled pre- and post-test score means and standard deviations for each school and grade level may be found in Appendix A, Table 1.

Student-Level Predictors: Five student-level predictors were included in the model: scaled pre-test scores (centered), English Language Learner, Special Education designation, grade level (dummy-coded with 8th grade as the reference group), and teacher utilization (centered) as measured by the actual number of lessons submitted by the student. Differences between Model 1 and 2 student-level predictors were 1) including grade level as a fixed effect at the student level, rather than the classroom-level and 2) the number of lessons submitted (rather than proportion in Model 1) as a measure of teacher utilization. The raw number of lessons submitted was investigated, as a more intuitive way of approaching teacher utilization compared to Model 1.





APPENDIX B:

**HIERARCHICAL LINEAR MODEL
FOR INVESTIGATING FINDING #2**

MODEL 2

Teacher and School-level Predictors: Given that students were nested within teachers, who are nested within schools, we wanted to investigate the effect of student-level predictors when controlling for teacher effects, which Model 1 did not permit.⁷ Therefore, teachers (n = 52) nested within schools (n = 7) served as the second level in a random intercepts model with random intercepts estimated for 1) teachers nested within schools and 2) schools. Another difference from Model 1 was that we included grade level as a fixed effect at the student level. The intra-class correlation for the teacher effect was 0.22, which was substantially higher than the 0.026 intra-class correlation for school effects.

When determining model selection, we compared three nested models: 1) an intercept-only model with teachers nested in schools, 2) a fixed effects OLS model, and 3) a random intercepts model with students nested within teachers nested within schools. The full random intercepts model had the lowest AIC value and explained a statistically significant proportion of null deviance, over and above the reduced models (See Table 3, below).

TABLE 3

Comparison Between Intercept-Only, Fixed-Effects and Random Intercepts Models

Model	Parameters	AIC	BIC	LL	Deviance	χ^2	<i>p</i>
Intercept-only	4	8388.3	8408.6	-4190.2	8380.3	1.46	---
Fixed effects	8	8328.4	8367.9	-4156.2	8312.4	67.93	<.001
Random Intercepts	10	8303.5	8352.8	-4141.8	8283.5	96.84	<.001

N = 1022. AIC = Aikake's Information Criterion, BIC = Bayesian Information Criterion, LL = -2 Log-Likelihood, *p* = null hypothesis probability value (i.e., statistical significance).

⁷ In addition to controlling for teacher and school effects, the HLM model also accommodated for violation of OLS assumptions of local independence and homoscedasticity.



APPENDIX B:

**HIERARCHICAL LINEAR MODEL
FOR INVESTIGATING FINDING #2**

FINDINGS

Model 2: Model 2 was estimated using restricted maximum likelihood estimation (REML) using the lme4 package in R (Bates et al., 2015).⁸ Findings from the Model 2 HLM analysis may be found in Table 4, below.

Note that, overall, the Model 1 findings were replicated by Model 2. There were two statistically significant student-level predictors of gain scores: 1) pre-test scores (grand mean centered), and 2) utilization (fidelity), operationalized as the number of lessons submitted by the student (grand mean centered). Of particular interest in Model 2 is that the number of student CL360 lessons submitted was a statistically significant predictor of student gain, controlling for scaled pre-test, English Language Learning and Special Education designation, grade level, as well as teacher and school-level variance.

TABLE 4

Student- and Classroom-Level Effects of CommonLit 360 Participation

Student-Level	Coefficient	Standard Error	Statistical Significance
Pre-Test Score	-0.27	0.03	***
English Language Learner	-1.41	1.66	
Special Education	-2.07	1.57	
Grade 6	2.69	1.82	
Grade 7	1.45	1.58	
Number of Lessons Submitted by Students	0.41	0.12	***
Intercept	1.95	1.55	
Random Effects	Variance	SD	N
Teacher-Level Intercept Variance	15.47	3.93	52
School-Level Intercept Variance	3.70	1.924	7
Residual Variance	186.09	13.64	1022
Proportion Total Variance Explained	Student (e)	Teacher (u_1)	School (u_2)
Value	90.7%	7.5%	1.8%

Note: Reference categories for this HLM model Grade 8. Total number of students included in the analysis was n=1,022.
***p < .001

⁸ Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1-48. doi:10.18637/jss.v067.i01. Note: Models 1 and 2 were estimated by two separate researchers. Model 1 was estimated via IBM SPSS Statistical Software by an independent researcher from Ad Hoc Analytics.



LIMITATIONS AND RECOMMENDATIONS

Although the findings from this study are consistent with past evaluations of the impact of CommonLit on student performance, the analysis was limited in several ways.

First, the number of students with both valid pre- and post-assessments (Model 1 $n = 1,014$; Model 2 $n = 1,022$) was a little more than half of the number of students who took just the pre-assessment ($n=1,848$). We conclude that pandemic-related school closures had an impact on assessment administration during this study. Closing this gap in future efforts will help to both increase the amount and (potential) variety of data being analyzed (e.g., more school coverage) and reduce the lost effort of obtaining pre-assessment data without subsequent post-assessments. We recommend that future evaluation efforts include significant follow up with teachers of students who completed a pre-assessment to ensure completion of the post-assessment. Even if students are ultimately unable to complete a post-assessment, learning why will help subsequent efforts.

Second, in order to make causal claims about effectiveness, a randomized control trial study or strict quasi-experimental design with a balanced comparison group is desirable. A mixed methods quasi experiment is in motion for the current school year (SY 2021-2022), with plans to conduct a larger scale experimental study to learn more about the efficacy of CommonLit 360 as a core English Language Arts program, its impact on subgroups of students, and the factors contributing to its success in the classroom.



CommonLit, Inc.
660 Pennsylvania Avenue SE
Suite 302
Washington, DC 20003

info@commonlit.org
www.CommonLit.org