Investigating the Effects of CommonLit on Student Engagement in Middle School Reading Classrooms

Michelle Brown, M.Ed. Principal Investigator
Dr. Ronald Ferguson, Faculty Advisor
Harvard Graduate School of Education
May 16, 2014

Introduction

Our school system struggles to prepare students to meet the growing literacy demands of college and the workforce. In recent years, federal initiatives, such as Reading First, have focused on supporting literacy at the early years. But while early literacy rates have steadily improved since 1970, literacy scores at age thirteen remain flat (Murnane, Sawhill, & Snow, 2012). According to the National Assessment of Educational Progress, only 36% of eighth graders read at proficient or advanced levels (NCES, 2013).

The vast majority of the research on reading achievement uses test scores as an outcome measure. However, a considerable body of research has been done in recent years to show the promise of student engagement as an outcome measure. Students who are more engaged in school have higher test scores (Fredericks, 2004), drop out at lower rates (Finn & Rock, 1997), are more resilient, exhibit increased strategy use to solve problems (Connell & Wellborn, 1991), and comprehend texts at higher levels (Guthrie et al., 2004).

As our education system is financially pressed to do more with less, our team created a free, open-access reading intervention called CommonLit for middle school English classrooms with the goal of increasing student engagement in reading. CommonLit is a research-based supplemental curriculum designed to give reading classrooms a more authentic focus through high-interest reading materials and cognitively demanding questions, like "Which is more important—freedom or security?" which can be used to foster rich discussion and dialogue. To test the intervention in its nascent stages, we conducted a small-scale study to measure the impact of the intervention on student engagement. The question for our research was:

> **Research Question:** Can CommonLit, a scalable curricular intervention, increase student engagement in middle school English classrooms?

To answer this question, we conducted a four-week quasi-experiment in nine middle school classrooms within the Boston Public School District. We administered the CommonLit intervention, which included one hour of professional development and access to our text collection, to five classroom teachers while the other four teachers resumed normal instructional practices. We used a student survey modeled after the Tripod® survey as a tool to measure changes in student engagement over the course of the study.

Student Engagement: A Conceptual Framework

Figure 1 below is a heuristic of our theory of student engagement. We posit that there are four major factors or "domains" that influence student engagement: i) the curriculum, ii) the teacher and the task, iii) the classroom, and iv) the student. Our theory is that each of these four "domains" plays an important role in the student engagement equation and is directly related to whether or not students find joy in reading.

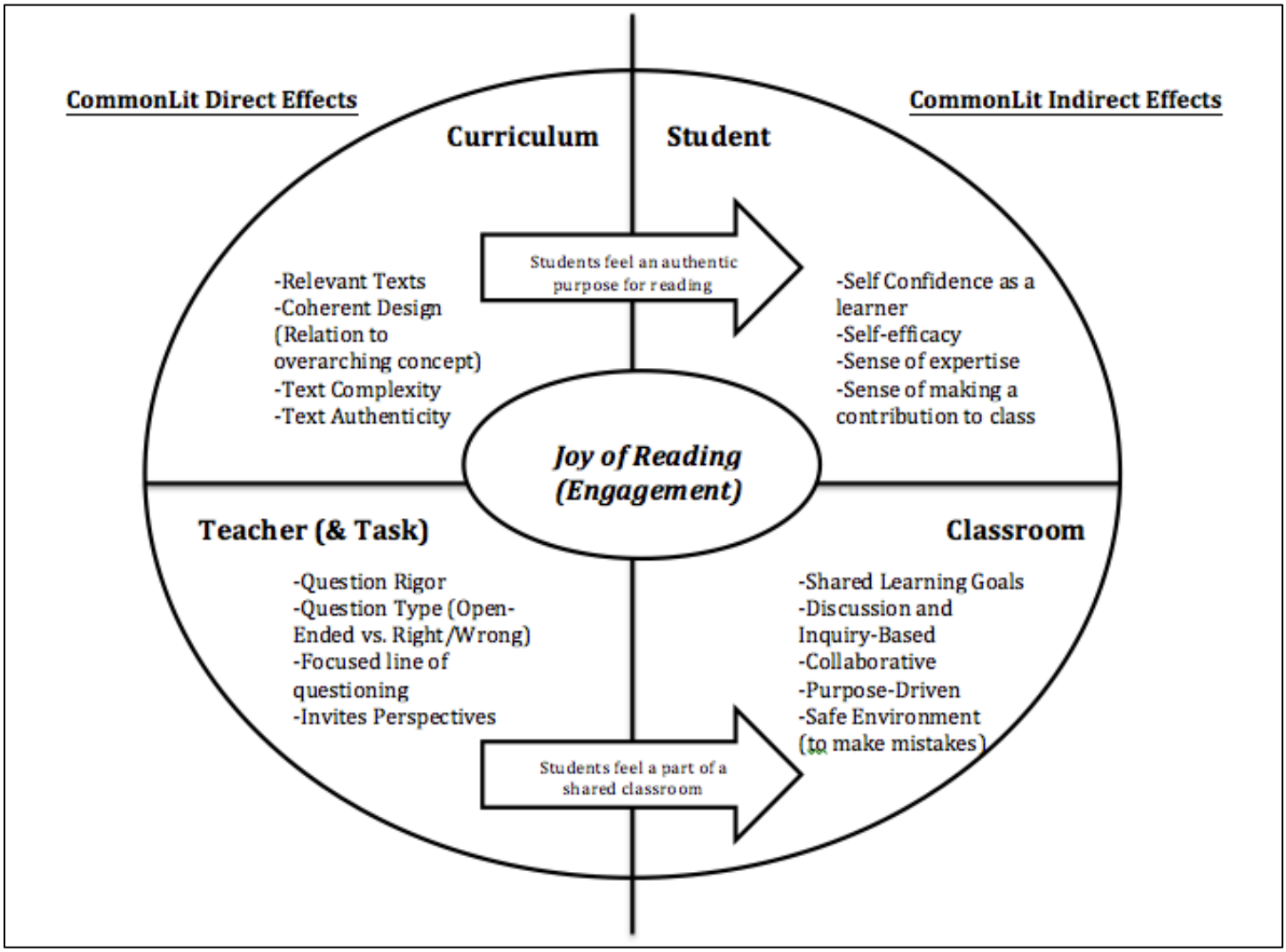

Figure 1. Heuristic of student engagement, a theoretical framework

In the curriculum domain, we posit that the key variable is whether reading content is relevant to students. Experience has shown us that middle school students tend to find more meaning and authentic enjoyment in reading materials that coherently relate, conceptually and topically, to their lives outside of school. We also argue that students are likely to find more relevance in reading materials that add to their cultural background knowledge; students who know something about history and current events have an "entry point" to civic and social engagement. In this same vein, we advocate reading content that has not been inauthentically "doctored" or

"dumbed down" for students. Our approach to curriculum design relies on the theory that students' exposure to authentic and diverse reading content is critical for long-term engagement.

Within the curriculum domain, we believe that in addition to relevance, coherence is another key factor that influences student engagement. Curriculum that allows students to investigate one topic or question deeply has been linked to increases in student engagement, cognitive efforts, and reading outcomes (Guthrie et. al, 1996). Curriculum composed of seemingly random, short readings with no clear relationship will likely do little to excite students. In contrast, curriculum composed of multiple readings on a single topic of inquiry will likely lead to students' gaining a sense of expertise.

The teacher and task domain also plays a key role in fostering student engagement. Standard practice in typical middle school reading classrooms involves low-level questioning, lecture, behavior correction, and unaccountable silent work (Nystrand & Gamoran, 1991). Moreover, this classwork tends to decrease, rather than increase, in cognitive demand from elementary to middle school—at a time when brain development is burgeoning (Eccles et. al, 1993). Moreover, research shows that students tend to show preference for challenge (Newmann et al., 1992), and feel intrinsic value for work that asks them to master a cognitively demanding task (Brophy, 1987). Indeed, the importance of brain development in the middle school years can not be overstated. These years compose the second most plastic period in the human brain's development, after only infancy. During this time, adolescents acquire skills in logical reasoning, executive functioning, and social awareness or "perspective taking" (Eccles et. al, 1993). Classrooms that employ discussion-based approaches in which students are encouraged to speak to one another are not only more developmentally appropriate for adolescents (Eccles et. al, 1993), but they are also associated with statistically significant gains in reading scores for both low and

high-skilled readers (Applebee et al., 2003). For these reasons, we argue that rigorous questioning in the classroom will increase cognitive demands on students, and will lead to increases in student engagement.

Student-level factors also affect student engagement. Student perceptions of the curriculum, the teacher, and the other students, form the basis for a sense of self-efficacy. Our theory is that when students have the opportunity to contribute something of academic value to the class, it will lead to a sense of self-worth and belonging. In turn, students who feel a sense of belonging will likely participate in classroom activities, understand the value of working hard, and therefore, show greater levels of overall enjoyment for reading. As Figure 1 shows, the student domain is inextricably related to curriculum and teacher practice; readings should invite argument, and teachers should foster thought, discussion, and debate.

Research Design

Our review of the research regarding reading achievement at the middle school level suggests a need for more information on specific interventions that that have the potential to improve the quality of typical curriculum and teaching practice. CommonLit, the treatment featured in this study, is a research-based reading intervention developed in September 2013 by former reading teachers and current graduate students at the Harvard Graduate School of Education. The creators of this intervention conducted a small-scale pilot study of this resource to test its utility and collect information to inform ongoing improvement of the curriculum.

Sample

Our sample includes 222 middle school (5^{th}-8^{th} grade) students. These students are nested within 9 teachers' classrooms within 5 schools in the Boston Public School District. To recruit

these teachers, we contacted school leaders and school-level reading specialists, who then gave teachers the option to participate in the study. Because of the limitations we faced in terms of recruiting teachers, we were not able to randomly assign teachers to treatment and control groups. We acknowledge there may be bias in our sample of teachers, who might be more motivated to try a new intervention or teaching technique. The five schools represented in our sample are:

Denver McCormack K-8 School

Denver McCormack is a K-8 school located in Dorchester, Massachusetts. 81.4% of its student population is low-income (defined as eligible for the free and reduced price lunch program). The demographic breakdown of McCormack is 43.7% Hispanic, 36.2% Black, 8.7% White, 8.5% Asian, and 3.2% Other. In 2013, Denver McCormack ranked in the 3rd percentile for overall performance relative to other schools in the district in the same grade-span. On the English Language Arts state assessment, 63.8% of 6th grade students scored in the "Warning" and "Needs Improvement" categories. We recruited one 6th grade teacher from this school for our treatment group sample.

Jackson-Mann K-8 School

Jackson-Mann is a K-8 school located in Allston-Brighton, Massachusetts that serves a racially diverse, predominately low-income (76.6%) student body. On the 2013 MCAS English Language Arts assessment, scores for the 5th grade (now 6th grade) cohort were almost normally distributed on "Advanced" to "Warning" scale, mirroring overall state averages for reading. We recruited one 6th grade general education reading class to be in our treatment group, and one 7th grade general education reading class to be in our comparison group.

Roxbury Preparatory Charter School

Roxbury Prep Mission Hill is a high-performing 5th-8th grade charter school in the

Uncommon Schools network located in Roxbury, Massachusetts. All Roxbury Prep students are students of color, and approximately 72% are eligible for the free and reduced price lunch program. Roxbury Prep has a lottery-based enrollment policy, and 75% of students at the school live in the adjacent neighborhoods of Roxbury. In 2013, 58% of students in 6th grade scored "Advanced" or "Proficient" on the English Language Arts state assessment. In 7th grade, 81% of of students scored "Advanced" or "Proficient" in 2013, and 93% scored "Advanced" or "Proficient" in the 8th grade cohort. In 7th and 8th grade, students reversed the achievement gap in English Language Arts by an average of 5.5 percentage points. We recruited one 6th and one 7th grade teacher to join the comparison group for our pilot study.

Maurice J. Tobin K-8 School

Tobin is a K-8 public school located in Roxbury, Massachusetts. It serves an 87.4% low-income population, and is classified as a "Level 3" school.[1] Collectively, a majority of Tobin students are either Black (24.9%) or Latino (68.9%), and 48.6% of students are classified as English Language Learners (ELL). In 2013, 37% of Tobin students in the 4th grade (now 5th grade) cohort scored in the "warning" category on the state English Language Arts assessment; no students scored "Advanced." We recruited two 5th grade teachers at Tobin to be in our treatment group. One of these teachers taught in a Sheltered English Immersion (SEI) classroom; the other teacher taught in a traditional classroom. A third teacher 5th grade regular education teacher was recruited to participate in our comparison group.

William B. Rogers Middle School

Rogers is a middle school located in Hyde Park, Massachusetts. 82.8% of the student body is classified as low-income. The school serves a predominately Black (65.9%) and Hispanic

[1] Defined as a school in the lowest performing 20% of schools in the BPS district.

(23.5%) population, and 23.2% of the study body is classified as ELL. In 2013, Rogers ranked in the 7th percentile for overall academic performance relative to other schools in the same grade span. We recruited one 6th grade teacher who taught a small, accelerated reading class to be in our treatment group.

We aimed to find schools in multiple neighborhoods that were representative of the public school population in Boston. To verify this, conducted a chi-squared test to ensure that the students in the comparison and treatment classes do not differ significantly in terms of gender or self-reported GPA, our two covariates. We found no statistically significant difference for these variables at the .05 level of significance. This led us to conclude that our sample populations in our comparison and treatment groups are equitable in terms of distribution for GPA and gender.

Comparison & Treatment Conditions

Five teachers in our sample at four of the partner schools were given access to the CommonLit intervention. Four teachers at three of the partner schools made up our comparison group; these teachers did not have access to the curricular materials or professional development. Of the 222 students in our entire sample, 110 were taught by teachers in the comparison group, and 112 were taught by teachers in the treatment group. To avoid contamination of our treatment and comparison conditions, no one teacher attempted to use the intervention in only certain classes of students; teachers either had access to the materials ("treatment"), or they were not made aware of its existence ("comparison").

CommonLit Pilot Study Activities

To answer our research question, we conducted a quasi-experiment in nine teachers' middle school (5th-8th grade) classrooms within the Boston Public School District. For four weeks,

we worked with five of these teachers as outside consultants offering resources, professional training, and optional ongoing curriculum development support. These five teachers administered a student engagement pre-survey to their classes, which we used as a baseline measurement for student-level analysis. After administering these surveys, teachers attended a one-hour training session[2] led by the CommonLit study team at each school site. The professional development had three parts:

Part 1: Introduction to the CommonLit Text Database

Teachers were instructed on how to access our free collection of thematically-organized supplemental texts. Through our website (www.commonlit.org) and through our offline database, teachers were able to download news articles, poems, short stories, speeches, historical documents, and magazine articles. These texts are organized by fourteen "discussion ready" themes, including social pressure, technology and progress, and morality (see Appendix, List of Themes & Essential Questions). Teachers were shown how to find texts that would deepen students' understanding of the themes they encounter in their existing curriculum and reading materials. It is important to note that CommonLit is not a comprehensive curriculum; it is rather a collection of short fiction and non-fiction texts that supplement existing curricula and build student background knowledge about topics of inquiry.

Part 2: Teaching Strategies to Engage Students

In the second part of the professional training, teachers were introduced to strategies to engage all students in classroom discussions. Teachers read transcripts of highly effective classroom discussions and then discussed the approaches that were introduced. These

[2] Due to limitations in scheduling, this professional development was cut short at three of the five school sites in our treatment group.

approaches include: driving the discussion with an open-ended question like, "Which is more important—freedom or security?" holding students accountable for speaking and listening, and pushing students to support claims with evidence from the text.

Part 3: Lesson Planning Time

Finally, teachers were given time to peruse the text collection, ask questions, and plan for an upcoming lesson. We provided teachers with a binder of planning templates, tips, sample lesson plans, and other resources.

Within one week of this professional development, our study team contacted teachers via email offering free one one-on-one lesson plan consultation. In this email, we also included attached "featured texts" and recommendations for implementation. Then, we contacted teachers in our treatment sample to share with us the planned dates for implementation. We observed these classrooms over the course a four-week period to collect additional qualitative data. In the final week of the study, we asked all teachers in our treatment and comparison groups to administer the student engagement survey a second time. We used this data to estimate growth at the teacher and student levels.

Figure 2 below shows how CommonLit supports our theory of student engagement directly and indirectly. CommonLit is primarily a curricular intervention that is meant to inject high-quality, relevant, short texts into middle school classrooms. Through the theme categories, teachers are able to retain instructional coherence, and use supplemental texts to engage students in a deeper inquiry into a particular question of the teacher's choosing. Through the one-hour professional development session, CommonLit is designed to improve teacher instruction by helping teachers integrate research-based best practices into daily lessons. CommonLit does not have direct effects at the student or classroom level; it influences these domains through curriculum and professional teacher coaching.

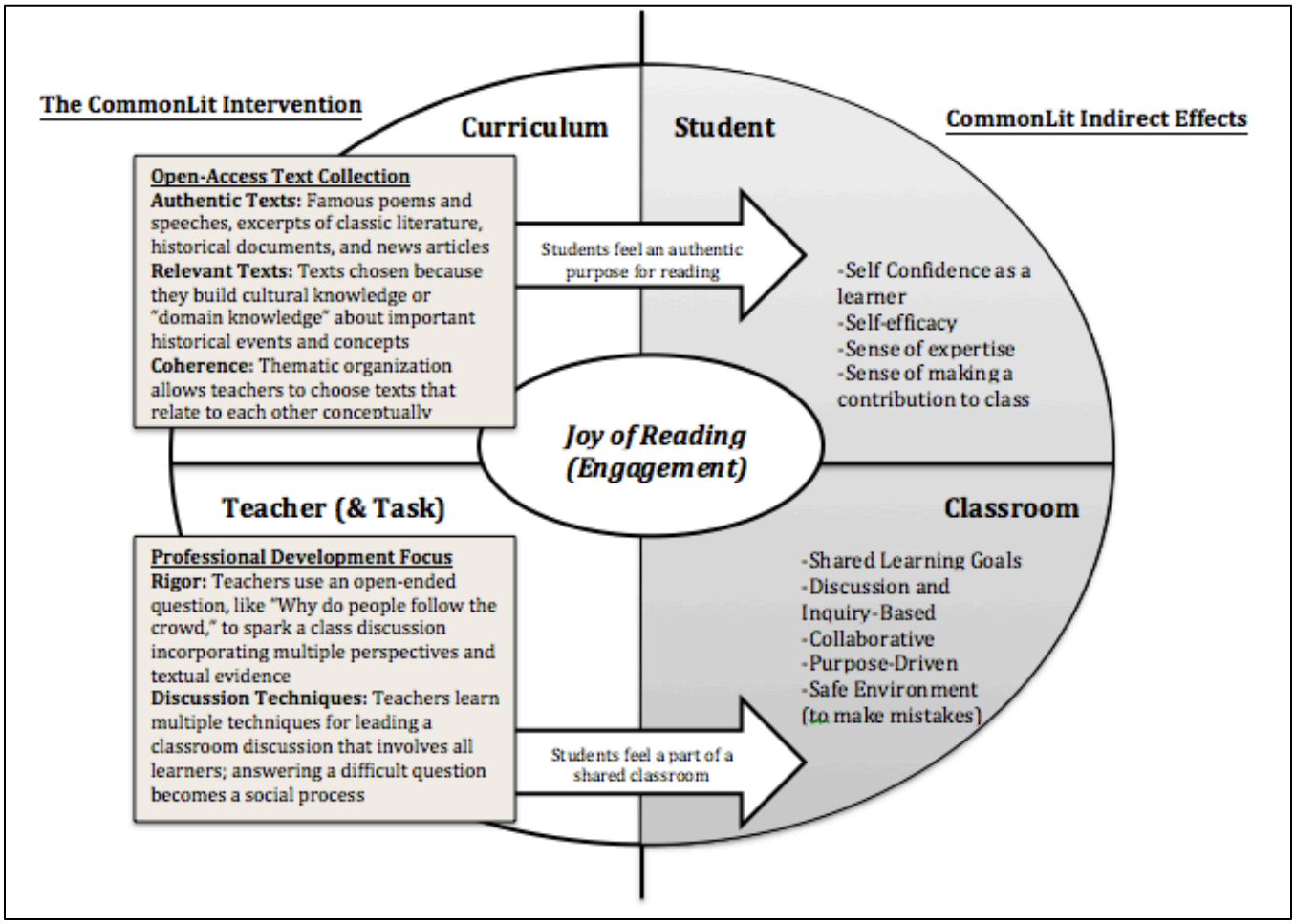

Figure 2. Design features of CommonLit as an intervention to build student engagement in reading

We recognize that teachers implement interventions with varying degrees of fidelity. For this pilot study, we asked but did not require teacher to implement a CommonLit supplemental text at least two times within a four-week period. We were interested in observing how student engagement changed for teachers at both high and low degrees of fidelity of implementation.

Measures

To measure student engagement in reading, we administered a nineteen item student survey (see Appendix). Six of these survey items were directly taken from the Tripod® survey, which is a widely-accepted survey tool used to assess teacher quality in the Measures of Effective Teaching (MET) study. Tripod refers to relationships, pedagogic skill, and content knowledge—three teacher characteristics that have been shown to be positively associated with student outcomes in math and reading (Ferguson, 2012; Bill and Melinda Gates Foundation, 2010). The remaining thirteen survey items were modeled after the Tripod® survey language and made more specific to engagement in reading. Theses items were developed under the guidance of Dr. Ronald Ferguson, the creator of the Tripod® survey.

Figure 3 shows the categorization of each survey item by each of our five engagement constructs. Each engagement construct has at least two, but not more than five, corresponding student survey items.

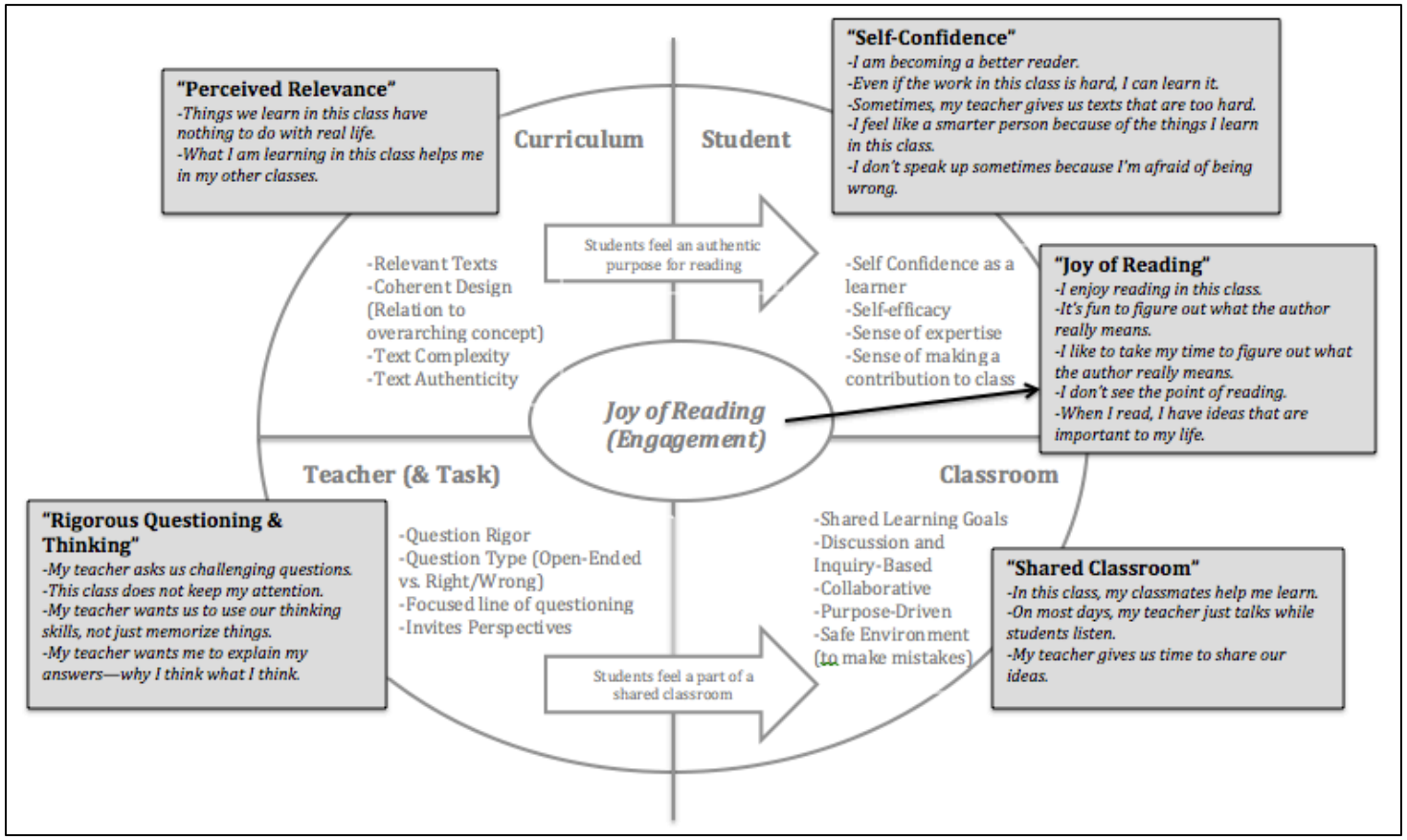

Figure 3. Measuring student engagement in reading through five latent variables, shown with corresponding stude engagement survey items

We conducted Chronbach's Alpha test for internal consistency for each of the five latent variable constructs in order ensure that each of the individual survey items was correlated with the latent variable. For the five items in the "joy of reading" construct, the alpha coefficient for these five items was 0.71 (pre-survey), and 0.74 (post-survey). For the two items in the "relevance" construct, the alpha coefficient was 0.40 (pre-survey), and 0.36 (post-survey). For the four items in the "deep questioning and thinking" construct, the alpha coefficient was 0.60 (pre-survey), and 0.50 (post-survey). For the three items in the "shared class experience" construct, the alpha

coefficient was 0.20 (pre-survey), and 0.41 (post-survey). Last, for the five items in the "self-confidence" construct, the alpha coefficient for was 0.60 (pre-survey), and 0.42 (post-survey). While we recognize that in future iterations of this study, we revise our survey to have higher internal consistency measures. For the purposes of this report, we determined that these alpha coefficients fell within an acceptable range.

Description of Variables

We collected data on the variables described in detail below:

Table 1. Description of Variables and Covariates

Variable Name (range)	**Description**
Student ID (1-222)	To protect the privacy of students in our sample, we assigned an ID code.
Sex	A student's self-reported sex; 1=Male, 0=Female.
GPA (0.0-4.0)	A student's self-reported grade point average (originally A, B, C, or D/F; recoded as a number on a 0-4 scale), collected on the student engagement pre-survey.
Class ID (1-9)	To protect the privacy of teachers, we will assign a class ID code. Each class ID code is associated with a certain teacher in either the treatment and comparison group.
Treatment/Comparison	0= Treatment Group, defined as groups of students taught by teachers exposed to the CommonLit intervention. 1= Comparison Group, defined as groups of students taught by teachers who used a traditional curriculum and were not exposed to the CommonLit intervention.
Student Engagement Survey Score Pre/Post (1-5)	All responses on the student engagement survey are coded on a scale of 1-5, 1 representing a response of "totally untrue" and 5 representing a response of "totally true." To estimate an overall engagement score, we averaged the results on student pre and post surveys. Conceptually, higher scores are associated with higher levels of student engagement. Since our survey included questions written in the negative, we reverse-coded these in our raw data.

Results

Overall Engagement Survey Results

Table 2 (see Appendix) shows the summary statistics of student engagement outcomes for both our treatment and comparison group. The mean change in overall engagement was 0.198 for our treatment group and 0.078 for our comparison group.

Table 3 is a taxonomy of our regression models using average engagement post survey scores as our outcome measure. In all five models, we used average engagement pre survey scores as our predictor of interest recognizing that pre survey scores would likely be a strong predictor of

post survey scores. Using pre survey scores as a predictor allows us to control for pre survey scores in our analysis.

Table 3. Comparison of Regression Models Predicting Student Engagement Scores

	(1)	(2)	(3)	(4)	(5)
	avgengagepost	avgengagepost	avgengagepost	Avgengagepost	avgengagepost
	b/se	b/se	b/se	b/se	b/se
Average Engagement: Pre	0.712***	0.714***	0.719***	0.711***	0.713***
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Comparison/ Treatment		0.115*	0.119*	0.126*	0.127*
		(0.05)	(0.05)	(0.05)	(0.05)
GPA			0.011		0.015
			(0.03)		(0.03)
Male				0.087	0.076
				(0.05)	(0.05)
_cons	1.180***	1.117***	1.054***	1.082***	1.028***
	(0.17)	(0.17)	(0.17)	(0.17)	(0.17)
N	222	222	215	220	213
R^2	0.517	0.527	0.541	0.538	0.550
adj. R^2	0.514	0.523	0.534	0.531	0.542
F	235.1	122.0	82.83	83.71	63.64
df_m	1	2	3	3	4
df_r	220	219	211	216	208

Note: ~p<0.10; *p<0.05; **p<0.01; ***p<0.001

As Model 2 shows, on average in our sample, being in the treatment group is associated with a 0.12 increment change in overall student engagement, controlling for engagement pre-scores. This result is statistically significant at the 0.05 level. Model 2 explains approximately 53% of the variation in average student engagement post-scores.

Model 3 indicates that there is not a statistically significant relationship between GPA and engagement post-scores, controlling for engagement pre-scores. Thus, we did not include GPA in our final model. This is contrary to the long held belief that students who receive high grades are more engaged. The latent variables used in our study are better predictors of higher engagement post-scores than is GPA.

Model 4 indicates that there is not a statistically significant relationship between gender and engagement post-scores, controlling for engagement pre-scores. Thus, we also did include this in our final model. Controlling for gender does not explain more of the variation in overall student engagement.

We chose Model 2 as our best-fit model to represent our final results. Our regression questions is:

$$\widehat{AverageEngagementPost} = 1.1 + .71 AverageEngagement\,Pre + .12 control / treatment$$

To verify the validity of this model, we checked the regression assumptions of linearity and homoscedasticity. Figure 4 (see Appendix) shows random scatter on the residual scatterplot for Model 2, which confirms that this model is a statistically valid representation of the results from our sample.

Figure 5 below is a graphical representation of Model 2, demonstrating the relationship between average engagement pre-scores and average post-scores for students in the comparison and treatment groups. For both the comparison and treatment group, there is a positive relationship between average engagement pre-scores and average engagement post-scores. How a student responded on the pre-survey is strongly predictive of how he or she responded on the post-survey. This graphical representation of our chosen model also shows that on average in our sample, the engagement post-scores were higher for students in the treatment group than in the comparison.

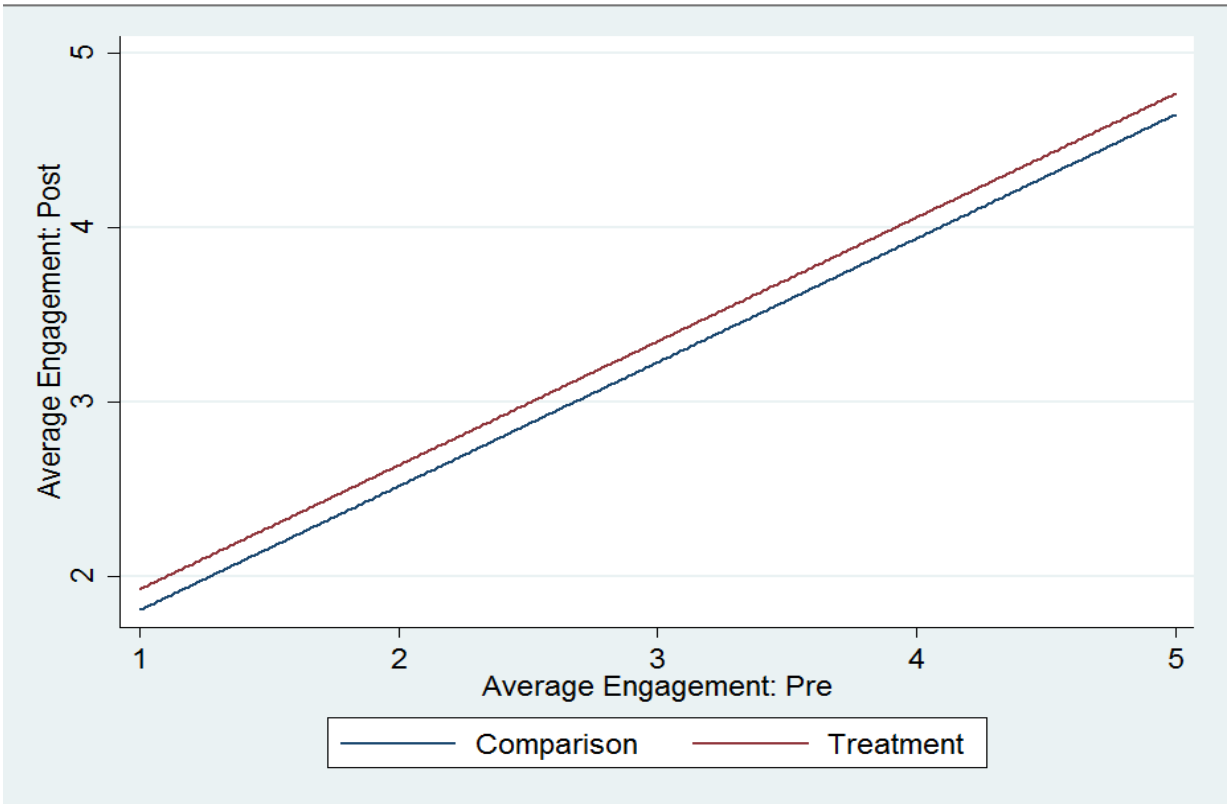

Figure 5. Relationship between student engagement pre-scores and student engagement post-scores for treatment and comparison groups, a graphical representation of Model 2.

Perceived Relevance

"Relevance" is a student engagement construct that refers to the degree to which the curricular materials build cultural background knowledge and relate to themes that are relevant to students' real lives. Figure 3 (see page 14) shows how "relevance" is a measure of the quality of the curriculum which affects teacher practice and student perception. The survey items (1-5 scale) for the "relevance" category were as follows:

1. Things we learn in this class have nothing to do with real life.[3]
2. What I am learning in this class helps me in my other classes.

[3] Note: Questions written in the negative were reverse-coded in the raw data; higher engagement scores are associated with greater student engagement.

The net average change for our comparison group was .12, compared to a net average change of .17 for our treatment group. A t-test revealed that the average change in the "perceived relevance" category was statistically significant for our treatment group (p=0.05), but not for our comparison group (p=0.2). Therefore, we recognize that we cannot assume that the trends we found for our comparison group in this particular category can be assumed for a whole population of students.

As Figure 6 reveals, the average growth in two of our treatment group classes (class 2 and class 7) was more than twice that of the highest average growth for our comparison group. One class in our treatment group, class 4, had the lowest average growth (-0.18) for this category.

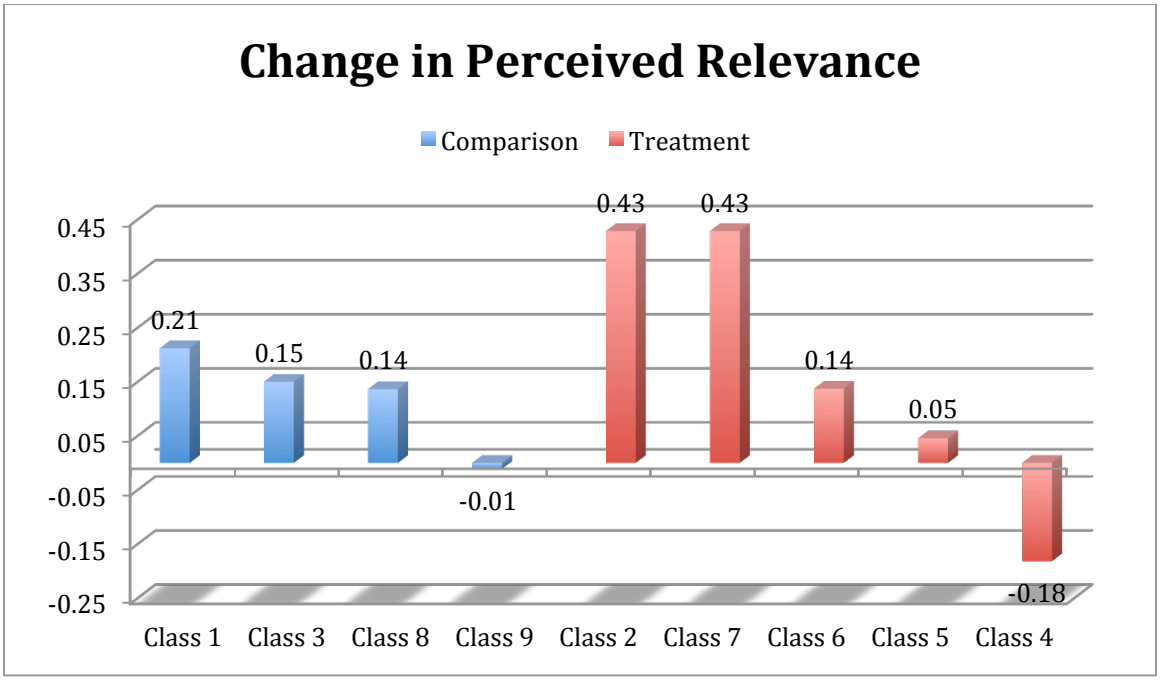

Figure 6. Change in "Perceived Relevance" latent variable for classrooms in treatment and comparison groups.

Classroom observation field notes offer greater insight into how teachers in our treatment group used CommonLit texts to increase curricular coherence and relevance. For example, in one

treatment group class, the teacher created a two-week unit on identity and the questions of how context can influence the way in which a person behaves, and what motivates people to act in certain ways. Throughout this thematic unit, the teacher focused on the skill of identifying character traits, a Massachusetts standard for grade-level reading. During this time, students read articles from the CommonLit text database that related to the guiding questions: an article about the concept of alter egos, an article about Superman/Clark Kent, and a biography of Stanley "Tookie" Williams, the leader of the Crips gang who wrote children's books about avoiding conflict from prison. This excerpt from our field notes transcript shows how the texts invited students to utilize their acquired background knowledge to build an understanding of the concepts in the text:

> **Teacher:** So I want to use what we talked about to infer about what this man was like and how he changed and why. What kind of a character he is (passing out papers). So, you're going to jot down notes…It says, (reads) "Stanley Tooke Williams was the leader of a notorious gang called the Crips, and he spent most of his life in prison…He did most his writing while he was in jail. As you read, write down your notes in the margins to
>
> **Student 1:** Are we talking about like, Superman, like how he can change?
>
> **Teacher:** No, we read about that last week. But, good connection. We're thinking about this man's character traits. Okay, any questions? I'll give you 10 minutes to read, and when you're done, just give me a thumbs up.
>
> (Students read quietly for 10 minutes, they are highlighting, underlining)
>
> **Teacher:** So let's look at what some people inferred about Stanley's character traits. First, raise your hand if you're ever heard of the Bloods and the Crips (all hands go up). So, I think this article is really interesting because it's something that actually happened in the real world.
>
> **Student 2:** (calling out) My dad was in the Bloods.
>
> **Teacher:** Really? James[4]?

[4] A pseudonym.

Student 3: Well, like, there was this guy I knew who was in a gang, and like, one time he was at home, and they drove by and shot at him, and like he almost died. It was crazy for him. Like he said that.

Teacher: So, I bet everyone here has either known someone who is a gang member or had been a gang member. (calls on another student)

Student 4: So like, 3 or 4 weeks ago, there were these people who drove by our house and shot at us. Like around my neighborhood there are a lot of gangs around us.

Teacher: So, this is real life stuff, guys. This is happening not just in California like in the article, but it's happening around here. So I think that since there are so many hands up right now that maybe we can have an open circle next week and
We can talk about things that have happened around our houses or our neighborhoods. But for now, let's learn about Stanley Williams and what happened to him in particular. How do you think Stanley Williams felt about himself when he was a gang member?

Student 5: He probably felt like he had a lot of power over other people around him.

Teacher: Where do you see evidence of that?

Student 5: Well in paragraph 3 it talks about how he like eliminated all the other gangs.

Student 2: (calling out) Ooo! I know!

Teacher: Okay (writes answer on board), another inside character traits about something—

Student 2: Also about the powerful thing, he got powerful because he, like, developed a reputation.

Teacher: (writes on board) So, if we were to write about this in a paragraph, we would have to cite evidence for this from the text. (calls on another student)

Student 6: Also, when he was little, like, he used to sell drugs and everything and he used to carry about a knife. And like, when he was young—

Teacher: So, he did those things because he had to protect himself, right? What about some of the ways he grew up? (calls on students)

Student 4: He used to hang around in an abandoned house. I underlined this. (Reading) "And vacant lots where he would sell drugs and gamble"

Student 7: So, when he was little, his dad left him, and his mom was never around. So, like he basically had to take care of himself.

> **Teacher:** So, that must have really affected him when he was a child, right? So we can infer that he felt scared when he was little (writes on board). But when he was a gang leader, it seems like he wasn't scared—is that true? Was he fearless?

Rigorous Questioning & Thinking

"Rigorous questioning and thinking," is a student engagement construct that refers to the degree to which teachers challenge students to use higher-order thinking skills. Figure 3 (see page 14) shows how "rigorous questioning and thinking" is primarily meant as a measure of teacher practice. The general level of rigor of teacher-level practices in turn affects the classroom environment, students' self-perceptions, and students' perceptions of the curricular content. The survey items (1-5 scale) for the "rigorous questioning and thinking" category were as follows:

1. My teacher asks us challenging questions.
2. This class does not keep my attention.[5]
3. My teacher wants us to use our thinking skills, not just memorize things.
4. My teacher wants me to explain my answers.

A t-test revealed that average changes in neither our comparison nor treatment were statistically significant at the .05 level of significance. However, at the 0.1 alpha level of significance, average change in the "rigorous questioning and thinking" variable was statistically significant for our treatment group (p=0.66).

Our treatment saw a net change of .09; the net change for our comparison group was -.025. As Figure 7 shows, for this latent variable, three out of the five treatment group classes had positive average growth, compared to one out of four for the comparison group. Class 2 showed

[5] Note: Questions written in the negative were reverse-coded in the raw data; higher engagement scores are associated with greater student engagement.

the highest overall average growth in the "rigorous questioning and thinking" category; this average growth was 1.83 times larger than the highest average growth for the comparison group.

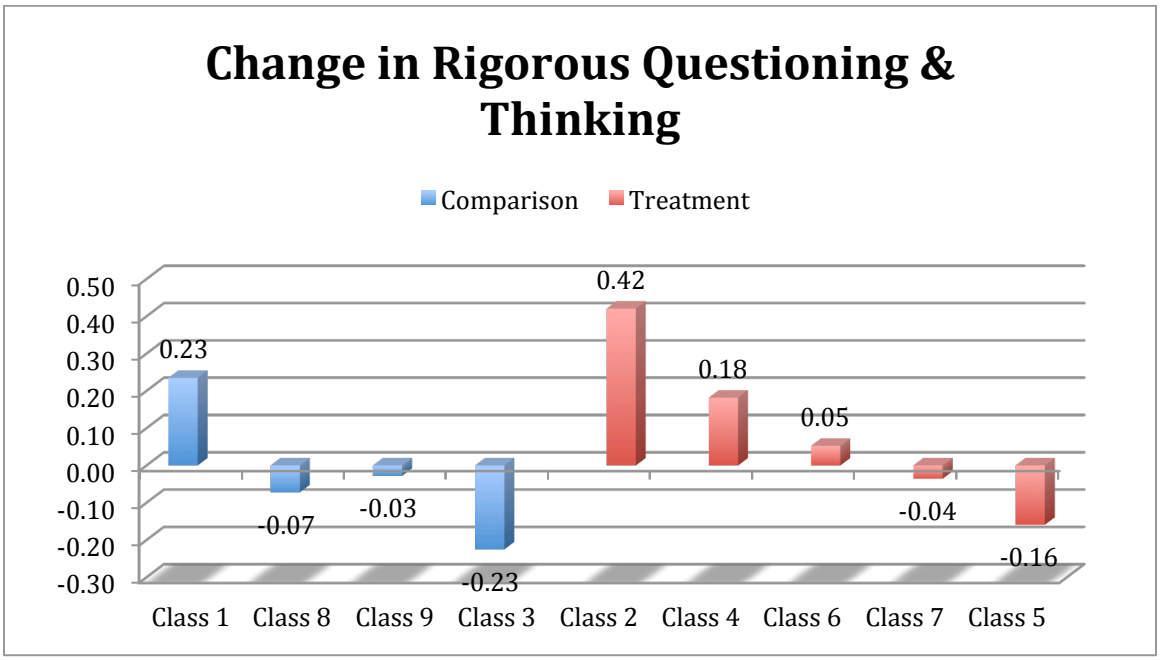

Figure 7. Change in "Rigorous Questioning and Thinking" latent variable for classrooms in treatment and comparison groups.

Observational data provided greater insight into the "rigorous questioning and thinking" construct. In general, we noticed that teachers who challenged students to anchor their claims with textual evidence offered more intellectually demanding experiences for students than teachers who accepted answers based on experience, feelings, or anecdotal evidence. Additionally, we noticed the following teacher practices to be positively associated with (what we observed to be) rigorous thinking: independent work; explicit directions for reading activities and note-taking; class time balanced between reading, writing, and discussion; expectations for annotating written texts; and the use of classic texts and literature. For example, in one classroom we observed, the teacher directed students to silently read an article about the Marshmallow Test, a psychological experiment conducted at Stanford in the 1960s, to discuss the concept of self-regulation and restraint. Then, students applied this knowledge in an evidence-based five-paragraph essay on the

question, "How does greed corrupt?" Students were asked to use evidence from the Marshmallow Test article and "The Odyssey." Before writing the essay, the teacher used the "cold-calling" technique to ask students to answer questions to summarize a chapter of The Odyssey in which the Sirens tempt the sailors of Odysseus's ship. Students were expected to copy the notes of this summary and use it for their independent writing assignment.

We noticed lower levels of intellectual rigor (observationally) in classrooms that employed the following techniques: teacher read-alouds, using photos for "pre-reading," prolonged free-writing, You Tube videos, lecture-based lessons, unaccountable group work, and public student behavior correction.

Shared Classroom Experience

"Shared classroom experience," is a student engagement construct that refers to the degree to which the classroom operates as a collaborative community. Figure 3 (see page 14) shows how "shared classroom experience" is a primarily meant to be a measure of classroom culture, which is built through teacher-student and student-student interactions. The survey items (1-5 scale) for the "shared classroom experience" category were as follows:

1. My teacher gives us time to share our ideas.
2. On most days, my teacher just talks while students listen.[6]
3. In this class, my classmates help me learn.

A t-test confirmed that that the average change in "shared classroom experience" was statistically significant in both our comparison (p=.001) and our treatment group (p=0.00). The net change in the average change in the "shared classroom" category was .36 for our comparison

[6] Note: Questions written in the negative were reverse-coded in the raw data; higher engagement scores are associated with greater student engagement.

group and .18 for our treatment group. As shown in Figure 8 below, all four classes in our comparison group showed positive average growth for "shared classroom;" four out of five classes in our treatment group showed positive growth.

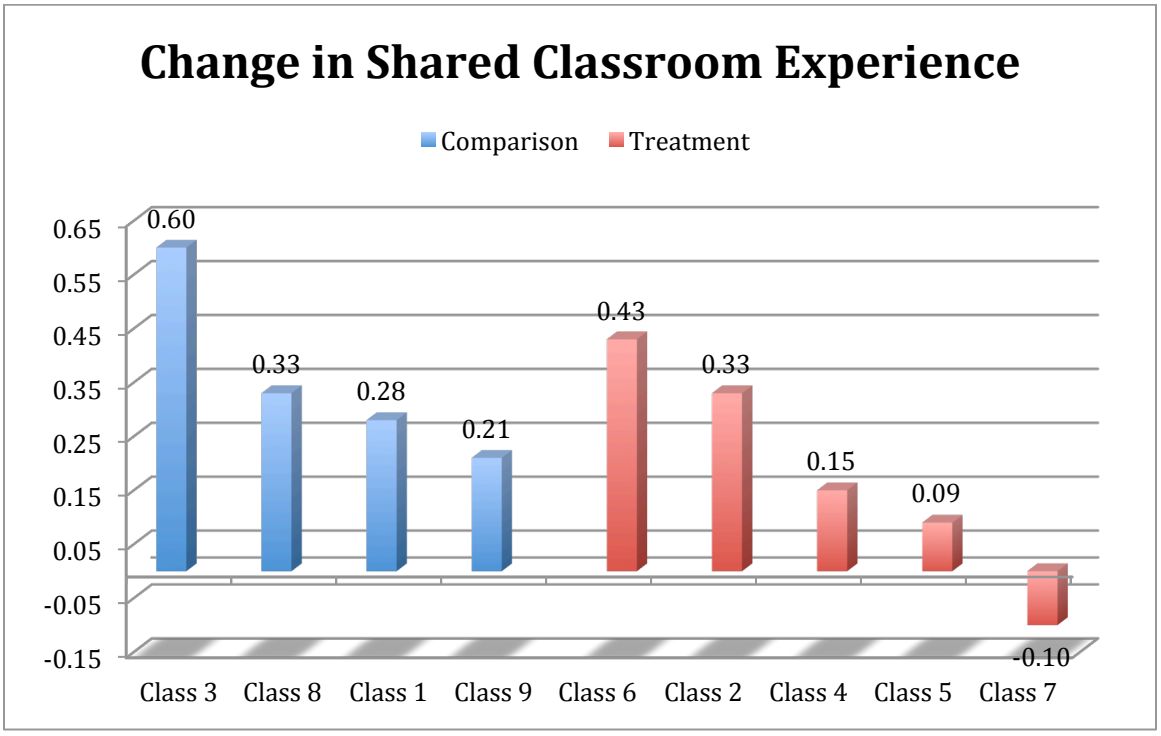

Figure 8. Change in "shared classroom experience" latent variable for classrooms in treatment and comparison groups.

Evidence of the CommonLit approach being used to enhance a shared classroom experience can be found in the transcript of teacher 2, who implemented CommonLit four times over four weeks in a unit anchored by *An American Plague*. Guiding his teaching with the question, "How does fear drive action?" this teacher led his students in deep inquiry and analysis of Theodore Roethke's poem, "The Storm," an article on the Salem Witch Trials, a news article about "Stop and Frisk," and an article about Edvard Munch's well-known painting, "The Scream." Throughout the four-week unit, we observed teacher 2 leading classroom discussions anchored by the question, "How does fear drive action?" Throughout the class period, the teacher employed best practices for creating a shared classroom environment: collecting exemplar student

commentary on the board, referring to students' claims, and pushing students to think critically about their responses to one another. The students in this transcript excerpt had just finished reading an article about the Salem Witch Trials:

> **Teacher:** Ok, so how does fear drive action? We talked about Trayvon Martin yesterday. Let's hear from some people.
>
> **Student:** I think the government in fear because that's why they accused witches and tried to arrest them, but it's clearly showing it's wrong if only 3 were convicted about of 150 people.
>
> **Teacher:** Well, you said a lot of things there. He said the government was afraid (writes on board). Now, does the government have emotions?
>
> **Student:** Fear turns innocent kind people into monsters.
>
> **Teacher:** Wait, whoa, that's awesome. (Writes) Fear drives innocent, kind people into monsters…so that's very interesting. Let me come back to Dominic.[7] He said the government was afraid. Can a government be afraid? What do you think?
>
> (Students muttering, talking: I think a government can be afraid. They can be afraid of being attacked...)
>
> **Teacher:** Let's hear from somebody. (Calls on student)
>
> **Student:** The government is afraid that the witches will overpower them. They are afraid the citizens were…(inaudible).
>
> **Teacher:** So, Sarah[8] brings up an issue of power, which is related to what Dominic[9] said. Let me ask what you think, can the government feel love, or shame?
>
> **Class response:** Yes. Yes. (calling out)
>
> **Teacher:** Nathan?[10]
>
> **Student:** Well, actually, I think the government needed a scapegoat to blame incidents on, and that was the witches.

[7] A pseudonym
[8] A pseudonym
[9] A pseudonym
[10] A pseudonym

Teacher: Interesting. So, Nathan said that the government needed to blame someone. So maybe fear drove the government to fabricate something.

Student: So, I agree with Sarah[11], because the government like, had more power than the citizens…and the government was afraid of that. In a democracy, like, we tell the government what we want. Maybe the government is afraid of these witches.

Teacher: So lets go back to this idea, fear makes innocent people monsters. What made him say that?

Student: I think he meant, like, if you're afraid of something, you should take it out, like get rid of it.

Teacher: Wow, ok (writes answer on board). Do you think that people during this time were afraid of witches? Do you think this could happen today? Do you think that innocent people are killed today during modern times?

Class: Yes, Yes, Yes. (calling out)

Student: I think this means say someone threatened you, and then you'll have to protect yourself, like, "I have to get you before you get me"

Teacher: Hmmm. (Writes on the board) Interesting. What Daniel[12] said is really interesting.

Student: (calls out) It's a little thing called revenge!

Teacher: If you say, I have to get you before you get me. Do you *know* that the other side is out to get you? Or, are you just paranoid?

Student: If you get a bad vibe from someone, you might get intimidated by them. So fear comes from feeling intimidated.

Student: Sometimes, you fear the unknown.

Teacher: Right (writing). If you don't know something, then you kind of assume the worst (writes on board) and sometimes it drives people to do something that they wouldn't ordinarily do.

Student: That means that if you're an innocent person, then sometimes you're like a monster to them, even when you're not trying to be.

---

[11] A pseudonym
[12] A pseudonym

Self-Confidence

"Self-confidence," is a student engagement construct that refers to the degree to which value themselves and feel valued by others in the class. Figure 3 (see page 14) shows how "self-confidence" is a primarily meant to be a measure of student-level factors that influence student engagement, which is influenced by teacher, classroom, and curricular-level factors. The survey items (1-5 scale) for the "self-confidence" category were as follows:

1. I am becoming a better reader.
2. Even if the work in this class is hard, I can learn it.
3. Sometimes, my teacher gives us texts that are too hard.[13]
4. I feel like a smarter person because of the things I learn in this class.
5. I don't speak up sometimes because I'm afraid of being wrong.[14]

A t-test showed that the average changes in the "self-confidence" category were statistically significant in our treatment group (p=0.00), but not for our comparison group (p=0.98). While we can make an assumption about CommonLit's effect in this category for the general population, we can not assume the trends we observe in our comparison group exist in the population.

The net change for the average change in "self-confidence" was .10 for our comparison group, and .15 for our treatment group. As Figure 9 reveals, three out of five classes in our treatment group showed average growth in this category beyond 0.10; one out of four classes in our comparison group showed average growth beyond this level.

---

[13] Note: Questions written in the negative were reverse-coded in the raw data; higher engagement scores are associated with greater student engagement.
[14] Note: Questions written in the negative were reverse-coded in the raw data; higher engagement scores are associated with greater student engagement.

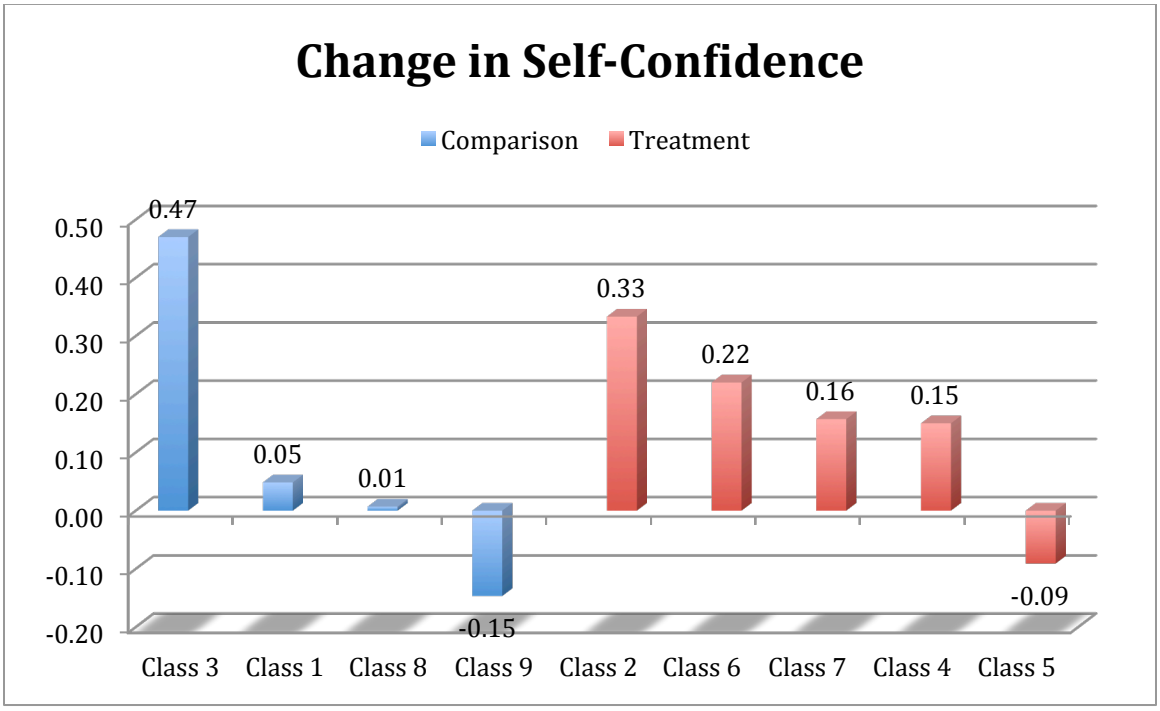

Figure 9. Change in "self-confidence" latent variable for classrooms in treatment and comparison groups.

Since student self-confidence and self-efficacy cannot be observed directly, we focused our observations on survey item 5, "I don't speak up sometimes because I'm afraid of being wrong." Our theory is that when students feel that they are contributing something of academic value to the class, their own sense of expertise and self-worth will improve. We were particularly interested in examining moments when students felt compelled to go "against the grain" of the class opinion, which we infer is the function of a student's self-confidence and self-efficacy. In the following transcript excerpt of class 4, we noticed the interrelatedness of the curricular and student domains (see Figure 3, page 14). The text in this transcript, "My Name" by Sandra Cisneros from *The House of Mango Street*, invited diverging opinions because the question, "Does the speaker like her name or not?" was not explicit or obvious in the text. Students were encouraged to form original opinions and support those opinions with evidence from the text. In the following transcript, the teacher begins with a read aloud of "My Name":

Teacher: Okay, I can see that you're eager to talk about this. Please follow along as I read it out loud. "In English, my name means hope, in Spanish my name means too many letters…Esperanza. My grandmother, I would have liked to have known her…She looked out the window her whole life…I wonder if she was sorry because she couldn't be all the things she wanted to be. I inherited her name, but I don't want to inherit her place by the window…Yes, something like Zeze the X will do." (Continues reading) Okay, what do you believe that she feels?

(Hands go up)

Student 1: She doesn't like it.

Teacher: So do we agree?

Students: YES! (aloud)

Student 2: (quietly) Well, not everyone.

Teacher: (Not hearing student 2) So, tell me one thing you read that made you say that.

Student 3: She said her name was like, muddy, like (reads), "The number 9."

Teacher: What do you think she means by that?

Student 4: Like, um, I think she was trying to say like, um, it's like very weird and disgusting.

Teacher: Why else doesn't she like her name?

Student 5: Because in Spanish it means sadness.

Teacher: Okay, and if something's sad, she doesn't really like it that much. What else?

Student 5: She said she would (reads) "like to baptize herself."

Teacher: Under a new name. What does that mean?

Student 5: She wants to change her name because a lot of people make fun of her name in school.

Student 6: Aw man, I was going to say that!

Students: (calling out) Man! Aw man!

Student 5: (calling out) Well, I said it first.

Teacher: Okay, what else haven't we said?

Student 7: She said that her name is ugly.

Teacher: Okay, she's comparing it to her sister. Her sister gets a nickname and she doesn't. What else? Why else doesn't she like her name?

Student 2: I disagree. I think she likes it. It was the same name of her grandmother and she inherited it. And she said that in English, her name means hope.

Teacher: Okay, there are two sides to this. Trey[15] agrees that she doesn't like it, but that's not the whole story. Good.

The Joy of Reading

"Joy of reading," is a student engagement construct that refers to the degree to which students enjoy reading and value the work in the class. Figure 3 (see page 14) shows how "joy of reading" is a general principle of student engagement that is influenced by factors in all four domains: curriculum, teacher, classroom, and student. The survey items (1-5 scale) for the "joy of reading" category were as follows:

1. I enjoy reading in this class.
2. When I read, it's fun to try to figure out what the author really means.
3. I like to take my time to figure out what the author really means.
4. I don't see the point of reading.[16]
5. When I read, I have ideas that are important to my life.

A t-test confirmed that in our sample, average change in the "joy of reading" variable was statistically significant for both the treatment (p= .01), and comparison group (p =.05). Our treatment saw a net change of .148; the net change for our comparison group was .086. As Figure 10 shows, for this latent variable, three out of the five treatment group classes had positive average

[15] A pseudonym

[16] Note: Questions written in the negative were reverse-coded in the raw data; higher engagement scores are associated with greater student engagement.

growth, compared to two out of four for the comparison group. The change in the "joy of reading" for class 7 in our treatment group was one and half times greater than the highest average change for the comparison group.

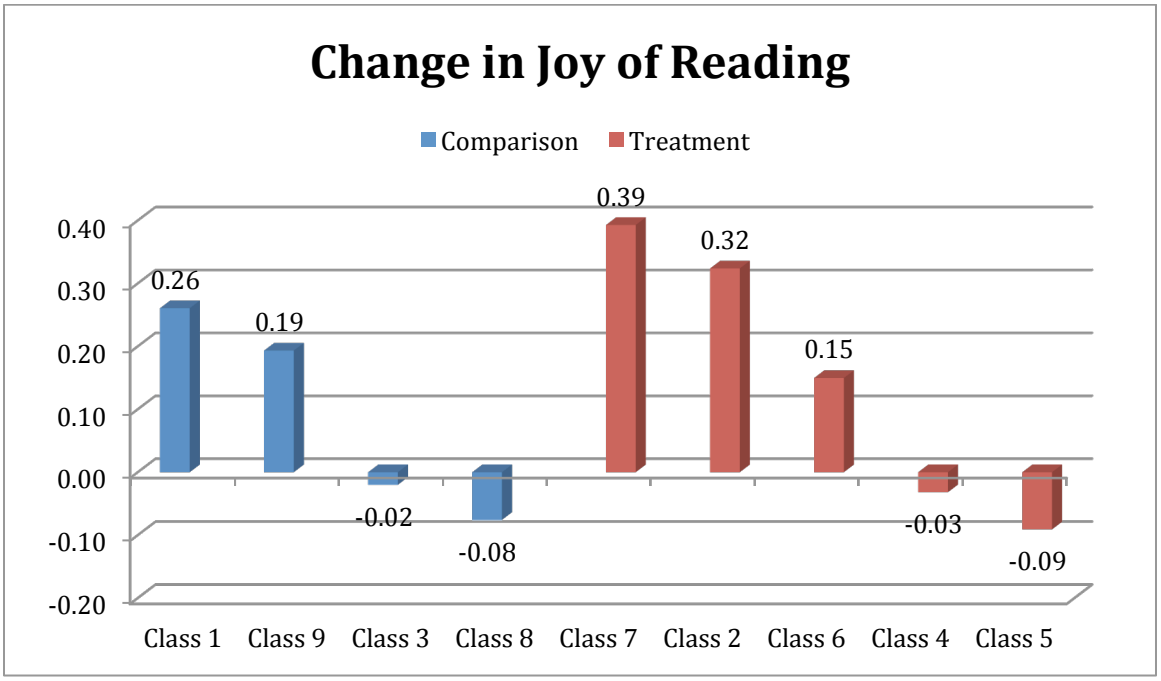

Figure 10. Change in "Joy of Reading" latent variable for classrooms in treatment and comparison groups.

In the following treatment classroom transcript, we notice evidence of four engagement constructs: relevance, rigorous questioning and thinking, shared classroom experience, and self-confidence. Our theory is that all of these four engagement constructs affect a variable we call "the joy of reading." For example, using a "kid-friendly" Fourteenth Amendment court case about a young American Muslim girl who was not allowed to participate on the school's soccer team, this teacher conducted a mock trial and pushed students to work together to draw a conclusion based on the presented evidence as members of the "jury." Our field notes capture the student response:

> **Teacher:** Okay! (Knocks with fist on desk) Order in the court! (Students laugh) So, we're going to need all jury to be very attentive so that you can decide if Scora's Fourteenth

Amendment right was not protected by this situation. So, we're going to call Scora to the stand and her lawyer.

(Students act as lawyers and witnesses, read written testimonials, and follow cross-examination procedures.)

Teacher: Okay, so no further questions. Now I will turn it to the jury. You have to decide if the school system is guilty or not. Is the school system violating her right? If you have questions, you will be allowed a couple of questions. So, jury get together and figure out what you decide.

How come the jury isn't getting together? You need to sit near each other to talk.

(Students move; teacher follows the jury to answer questions.)

Small Jury Group:

Teacher: So, they didn't let her play on the soccer team because they said it was dangerous for her to wear the clothes required by her religion. So based on that, do you think the school system is guilty?

Student: But, like if she plays better with the clothes, then she should be able to wear it.

Teacher: Okay, good point, but we have to agree. And one person has to be the speaker and tell what you thought and why you thought it. (Teacher leaves group)

(Students talk to one another for 3-4 minutes.)

Full Class:

Teacher: Okay, so we're going to get back together. Order in the court! (Students laugh)

Teacher: So, jury, did you make a decision?

Students: We say that the school people are guilty and I say that because, Scora, if she could play better with the skirt, that means that she should be able to play. Because in the beginning the principal said that she could play with the clothes.

Teacher: So, jury did you decide what should happen from here?

Student: Yeah, we decided that she should be allowed to play on the soccer team.

Teacher: Okay, so thank you to both sides. Oliver[17], I think you should be a lawyer!

[17] As pseudonym

Student: Yeah, my mom said that.

(Students get back to seats)

Teacher: So, later today, everyone is going to get the chance to write about this case because I want to be able to hear everyone's thoughts. So, today everyone is going to write about whether or not the coach violated the Fourteenth Amendment. I want you to write about it because I thought this was a really interesting case, and I want to be able to hear from everyone.

(Bell rings; end of transcript)

Limitations

The primary limitation of this study is the small sample size, which made it difficult to determine whether our results can be assumed for a whole population of students. We also recognize that there may also be bias in our sample because we were unable to achieve random assignment to true treatment and control conditions. Teachers who opted into our study may exhibit a characteristic that influences student engagement in reading. In addition, due to factors related to school scheduling and state testing, our four-week implementation period was short, and this may have limited the potential for faithful implementation of the intervention.

Conclusion

On average for our sample, being in the treatment group was associated with a 0.12 increment change in overall student engagement, controlling for engagement pre-scores. This result is statistically significant at the 0.05 level and explains approximately 53% of the variation in average student engagement post-scores. The treatment group outperformed the comparison group in every latent variable category in terms of overall mean changes in engagement, however, not all of these results were statistically significant, and some of the differences were quite small. Through classroom visits, we observed teachers in our treatment group employing some research-based best practices to promote rigorous cognition and student-centric discussion to promote

engagement. We noticed that one particular teacher (class 2) who implemented the intervention with high fidelity (defined in terms of frequency of use) tended to have higher mean engagement changes than other teachers. However, we recognize that this observed trend may not be true for a whole population of teachers. Nevertheless, we conclude that the CommonLit intervention merits further research to study its effectiveness as a tool to make adolescents more engaged in reading.

Acknowledgements

The author wishes to thank Harvard School of Education professor, Dr. Ronald Ferguson, for his ongoing support, encouragement, and inspiration as the faculty advisor for this study. Acknowledgements are also due to members of the pilot study team: Jessica Yarmosky, Taylor Percival, and Mike Dabrieo. The author also wishes to thank the teachers and administrators at the following Boston Public Schools for their cooperation: William B. Rogers Middle School, Denver McCormack K-8 School, Roxbury Preparatory Charter School, Jackson-Mann K-8 School, and Maurice J. Tobin K-8 School. The author also wishes to thank several Ph.D. candidates, Julie Russ-Harris, Kathleen Lynch, and Johanna Barmore, for their advice and expertise. Additional thanks goes to Joshua Biber and the middle school teaching cohort at Teach for America Boston for providing feedback on our professional development in its early stages. Finally, acknowledgements are also due to CommonLit volunteers, Lea Marmora, Sarah Mielbye, Lindsay Denman, and Audrey Dau for their ongoing dedication to the project.

Resources

Applebee, A. N., Langer, J. A., Nystrand, M., & Gamoran, A. (2003). Discussion-based approaches to developing understanding: Classroom instruction and student performance in middle and high school English. American Educational Research Journal, 40(3), 685-730.

Bill & Melinda Gates Foundation. (2010). *Learning about Teaching: Initial Findings from the Measures of Effective Teaching Project*. Met Project Research Paper.

Brophy, J. (1987). Synthesis of Research on Strategies for Motivating Students to Learn. Educational leadership, 45(2), 40-48.

Connell, J. P., & Wellborn, J. G. (1991). Competence, autonomy, and relatedness: A motivational analysis of self-system processes.

Eccles, J. S., Midgley, C., Wigfield, A., Buchanan, C. M., Reuman, D., Flanagan, C., & Mac Iver, D. (1993). Development during adolescence: the impact of stage-environment fit on young adolescents' experiences in schools and in families. American psychologist, 48(2), 90.

Ferguson, R. F. (2012). Can student surveys measure teaching quality?. *Phi Delta Kappan*, *94*(3), 24-28.

Finn, J. D., & Rock, D. A. (1997). Academic success among students at risk for school failure. Journal of applied psychology, 82(2), 221.

Guthrie, J. T., Meter, P., McCann, A. D., Wigfield, A., Bennett, L., Poundstone, C. C., ... & Mitchell, A. M. (1996). Growth of literacy engagement: Changes in motivations and strategies during concept-oriented reading instruction. *Reading Research Quarterly*, *31*(3), 306-332.

Guthrie, J. T., Wigfield, A., Barbosa, P., Perencevich, K. C., Taboada, A., Davis, M. H., ... & Tonks, S. (2004). Increasing Reading Comprehension and Engagement Through Concept-Oriented Reading Instruction. Journal of Educational Psychology, 96(3), 403.

Murnane, R. J., Sawhill, I., & Snow, C. (2012). Journal Issue: Literacy Challenges for the Twenty-First Century Volume 22 Number 2 Fall 2012.Journal Issue: Literacy Challenges for the Twenty-First Century Volume, 22(2).

National Center for Education Statistics. (2013). A First Look: 2013 Mathematics and Reading. National Assessment of Educational Progress at Grades 4 and 8. Retrieved from http://nces.ed.gov/nationsreportcard/subject/publications/main2013/pdf/2014451.pdf

Newmann, F. M. (1992). Student Engagement and Achievement in American Secondary Schools. Teachers College Press, 1234 Amsterdam Avenue, New York, NY 10027 (paperback: ISBN-0-8077-3182-X, $17.95; hardcover: ISBN-0-8077-3183-8, $38).

Nystrand, M., & Gamoran, A. (1991). Instructional discourse, student engagement, and literature achievement. Research in the Teaching of English,25(3), 261-290

Appendix

Themes and Existential Questions

America

What makes America unique?

How has the vision of America changed over time?

Fear & Paranoia

What are the causes and consequences of fear?

How does fear drive action?

Friendship & Loyalty

What makes a true friend?

Is it ever okay to betray a friend?

Growing Up

At what point are you grown up?

Should we treasure or value our childhood?

Identity

Where does identity come from?

Can you change your identity?

Justice, Freedom, & Equality

What should your rights be?

Which is more important: freedom or security?

Love

What is love?

How are people changed by love?

Morality & Evil

Why do good people do bad things?

How have people justified evil over time?

Peer Pressure

Why do people follow the crowd?

What are the effects of following the crowd?

Power & Greed

Can money buy happiness?

How does power corrupt?

Prejudice & Discrimination

How does prejudice emerge in a person or in a society?

What are the effects of prejudice & discrimination?

Resilience

How do people overcome adversity?

Is it ever okay to quit?

Social Change & Revolution

Can a single person change the course of history?

Why do people resist change?

Technology, Progress & Industry

What are the costs and benefits of progress?

Is technology good for humanity?

Table 2. Descriptive statistics summarizing change in latent variables for comparison (*n*=110) and treatment groups (*n*=112).

Variable	Mean	SD	Min	Max
Average Engagement Change (Comparison)	0.078	.432	-1.00	1.11
Average Engagement Change (Treatment)	0.198	.407	-0.89	1.80
Joy of Reading Change (Comparison)	.141	.768	-2.2	2.4
Joy of Reading Change (Treatment)	.156	.665	-2.0	2.0
Relevance Change (Comparison)	0.10	.821	-2.0	2.0
Relevance Change (Treatment)	0.17	.887	-2.50	2.5
Deep Thinking Change (Comparison)	0.02	.669	-2.0	2.0
Deep Thinking Change (Treatment)	0.12	.615	-1.5	1.75
Shared Class Change (Comparison)	0.294	.759	-1.33	2.67
Shared Class Change (Treatment)	0.317	.767	-1.67	3.00
Confidence Change (Comparison)	-.001	.608	-1.60	1.40
Confidence Change (Treatment)	0.20	.710	-1.20	4.60

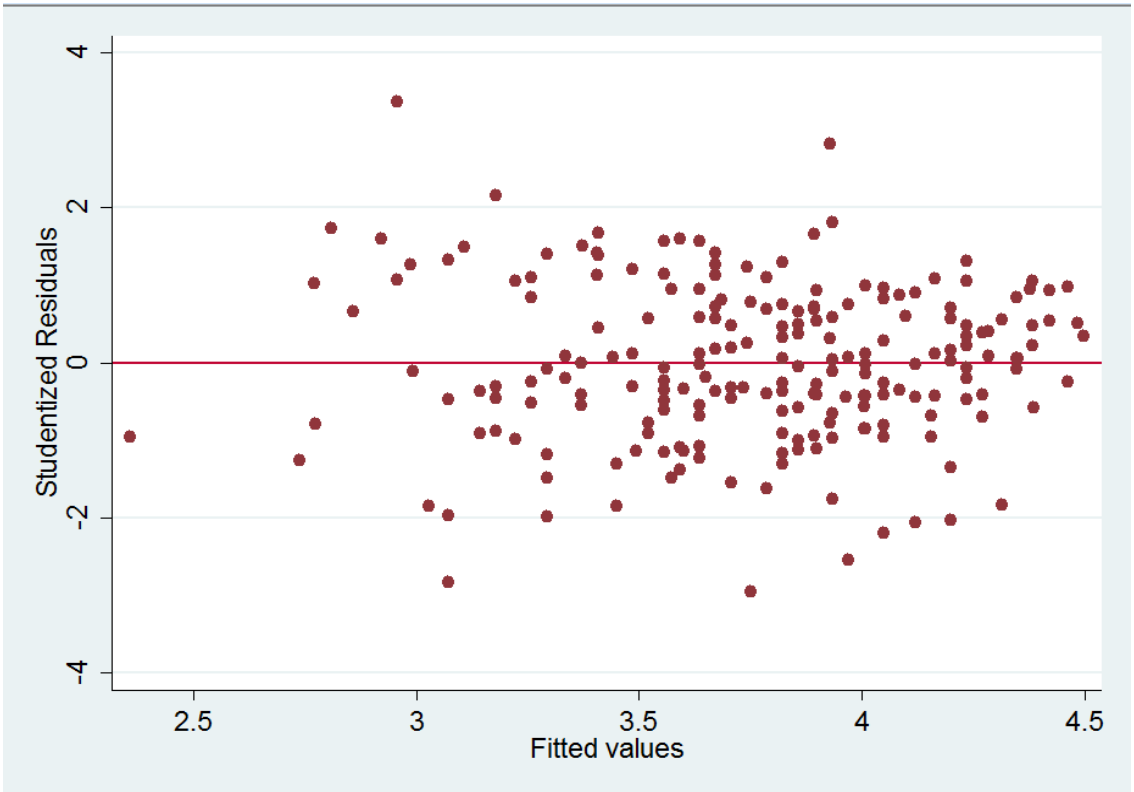

Figure 4. Residual Scatterplot of Model 2