Talking to Learn: Investigating the Relationship Between Classroom Discussion and Persuasive Writing

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters.

<table>
<thead>
<tr>
<th>Citable link</th>
<th><a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:37935844">http://nrs.harvard.edu/urn-3:HUL.InstRepos:37935844</a></th>
</tr>
</thead>
<tbody>
<tr>
<td>Terms of Use</td>
<td>This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA">http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA</a></td>
</tr>
</tbody>
</table>
Talking to learn:  
Investigating the relationship between classroom discussion and persuasive writing

Shireen Al-Adeimi

Catherine Snow  
Cathy O’Connor  
Luke Miratrix

A Thesis Presented to the Faculty  
of the Graduate School of Education of Harvard University  
in Partial Fulfillment of the Requirements  
for the Degree of Doctor of Education

2018
To my parents,
Mohamed Al-Adeimi and Huda Hussein

واخفص لَهُمَا جَنَاحَ الدُّلُوَّ من الرُّحُمَةَ وَقَلْ رَبّ ارْحَمُوهُمَا كَما رَبِّي صُغِّيَّرًا
Acknowledgements

I am grateful to my family, friends, mentors, and colleagues, whose support and encouragement has enabled me to complete this final milestone in my formal educational journey.

Thank you to my committee, Catherine Snow, Cathy O’Connor, and Luke Miratrix, for your mentorship, thoughtful insights, and meticulous feedback. Working with you and learning from you has been an invaluable experience for which I will always be grateful.

Thank you to my mentor Mary Schleppegrell, who first introduced me to the world of educational research, and whose encouragement led me to embark on this doctoral journey. Thank you also to my HGSE mentors Paola Uccelli and Pamela Mason for guiding me with your wisdom, encouragement, and support.

I am grateful to have been part of two research labs while at HGSE – Snowcats and Language for Learning. It was a privilege and honor to work alongside and learn from colleagues in both labs.

Thank you to my friend and statistics partner at HGSE, Wenjuan Qin. I am grateful for your friendship and support throughout the years, and I look forward to continued collaborations as we embark on our professional lives.

Thank you to the friends I’ve made in the Boston area, especially Aysha Ismail and Omair Saadat, who entered my life as neighbors but are now family.

Thank you to my sister, Yumna Al-Adeimi, for enriching my life with love and happiness. Your presence in my life is a gift for which I am truly thankful.
I am deeply indebted to my parents, Mohamed Al-Adeimi and Huda Hussein, who left their home, families, and careers behind to give me and my sister a better future. I am eternally grateful for your love, support, and sacrifice. Thank you also to my mother, who first taught me how to write and who contributed to this dissertation by helping me score students’ writing.

Finally, words cannot truly express my gratitude to my husband, Mahdi Ghassemi. Thank you for your love, patience, support, guidance, and unwavering belief that I could overcome any obstacle in my way.
# Table of Contents

Abstract .................................................................................................................................v

Chapter 1: Introduction .........................................................................................................1

Chapter 2: Study 1 .............................................................................................................17

Chapter 3: Study 2 .............................................................................................................62

Chapter 4: Implications for Classroom Practice ...........................................................118

Chapter 5: Conclusions and Future Directions ............................................................137

Bibliography .......................................................................................................................154
Thesis Abstract

Dialogic discourse emphasizes making meaning through discussion, and stands in contrast to recitation, where the teacher initiates a question that is followed by a student response and is then evaluated by the teacher (known as IRE). Though dialogic discussion is a desired medium through which students can develop various skills fundamental to their academic success such as reading comprehension and vocabulary development, it remains a rare occurrence in today’s classrooms. Moreover, the relationship between dialogic talk and outcomes such as writing quality is less understood. This dissertation explores the predictive association between features of whole class discussion and students’ writing quality in classrooms implementing debate-centered curricula. Findings from two empirical studies using the Low-inference Discourse Observation tool (LIDO) are highlighted. The LIDO is a unique instrument that captures various research-based indicators of discussion produced by students and teachers in whole class settings. The first paper analyzes teacher and student talk using audio recordings from 42 classrooms in grades four through seven, as well as corresponding student essays (n=471). Multilevel modeling shows that teacher-produced dialogic talk positively predicts students’ persuasive writing, while teacher- and student-produced low dialogic talk negatively predicts their writing. Contrary to hypothesized relationships, however, student talk coded as more dialogic using the LIDO was negatively predictive of persuasive writing. In this
study, participation rate was also positively associated with students’ writing. The second paper analyzes video recorded student data from one eighth grade classroom (n=15), where students discuss four civics topics over eight class sessions. Findings are consistent with the previous study, as dialogic student measures were negatively associated with persuasive essays. The fourth chapter examines implications for classroom practice gleaned from these results, such as emphasizing contestable questions over semi-open or quiz-like questions. The thesis concludes with suggested approaches to move the field of discussion research forward.
Chapter 1: Thesis Introduction

A common intuition among educators is that students learn best when spoken with, rather than spoken to (Blachowicz & Fisher, 2011). Even in ancient times, Socrates cautioned that depending solely on books limits learners’ ability to develop critical thought as the learner does not have an opportunity to interact with the author. That is, the reader is not able to debate with the author, challenge or question the thoughts presented in writing (Phaedrus, 274c-276d). Today, we have evidence that dialogic interactions in classrooms are associated with various positive learning outcomes such as comprehension of content (Ketch, 2005), and development of collaborative reasoning (Mercer, Wegerif, & Dawes, 1999). To what degree it is related to writing, however, is less understood.

Though writing research is not as well-funded as reading research (Juzwik et al., 2005), writing is an important learning outcome as it is a mechanism through which comprehension is revealed (Serravallo, 2012), and is in and of itself a process that is important for both academic and non-academic pursuits. Despite its significance for academic success, however, most fourth and eighth grade students in the United States are not proficient writers (National Center for Education, 2009). In order to improve students’ writing outcomes, classroom factors that may play a role in preparing students for writing are important to analyze. One such classroom factor is discussion that occurs in whole class settings.
In any given classroom, students talk with one another and with the teacher in various contexts and for several purposes. In small groups, students may discuss content for the purpose of understanding subject material through talk. Students also speak socially with their teachers, and more often, with each other. Teachers use talk to manage behavior and relay procedural instructions to students. More importantly, however, teachers also use talk to convey knowledge and lead discussions either one-on-one with a student, in small groups, or in whole class settings. Thus, talk among students and teachers is used for various purposes, most important of which is using talk to teach or to learn. This dissertation examines a specific type of classroom talk, dialogic discussion, and its relationship to an outcome students are expected to master by the end of high school, persuasive writing (Hillocks, 2002).

**Capturing Classroom Talk**

The empirical studies presented in this dissertation explore discussions in grades four through eight in a total of 43 classrooms located in the Northeastern United States, as well as the accompanying persuasive essays students produced after each discussion. Data used in this dissertation are drawn from the Catalyzing Comprehension through Discussion and Debate (CCDD) study, which investigates the role of a discussion-based curriculum in improving academic outcomes for students in grades four through eight. The curriculum developed to achieve this goal, Word Generation, is a cross-disciplinary supplementary program developed in partnership with the Strategic Education Research Partnership in
order to help improve students’ academic language, argumentation skills, perspective taking skills, and other academic outcomes that together target students’ deep reading comprehension (Snow, Lawrence, & White, 2009).

Word Generation units begin with a reading passage that outlines a controversial dilemma and highlights related academic vocabulary. Throughout the unit, students learn more about the topic by reading about various perspectives on the issue and engaging in activities, including a classroom debate, that further their understanding of the issue in various contexts. At the end of the unit, students produce a written persuasive essay in which they present and defend their own perspectives on the same topic of discussion.

In order to operationalize the classroom talk at the heart of the Word Generation intervention, an instrument that would help capture the various types of content-related student and teacher talk during whole class settings was developed (O’Connor & LaRusso, 2014; Elizabeth & O’Connor, 2014).

The Low Inference Discourse Observation tool (LIDO) is a novel instrument used to capture classroom talk and was developed for the purpose of coding whole class interactions by both students and teachers (O’Connor & LaRusso, 2014; Elizabeth & O’Connor, 2014). In the following section, the LIDO is introduced and its coding principles are described and illustrated with examples.

Drawing on previous work on accountable talk by Michaels, O’Connor and Resnick (2008), as well as measures of classroom talk identified by researchers over the years (e.g. Nystrand, Wu, Gamoran, Zeiser, & Long, 2003; Snow,
Lawrence, & White, 2009; Soter et al., 2008; Wells & Arauz, 2006; Mercer, Wegerif, & Dawes, 1999), the Low-Inference Discourse Observation tool seeks to operationalize classroom talk at all grade levels. It captures information about discourse practices that can then inform researchers’ and practitioners’ understanding of how to promote effective classroom discussions (Elizabeth & O’Connor, 2014).

The LIDO is composed of five sections: Parts A through D of the LIDO are informally called the “paper LIDO” as they involve on-site observation and coding of classroom and discussion information, while Part E involves off-site coding from audiotapes. Part A of the LIDO contains descriptive information about the classroom such as grade level, topic, treatment type, the duration of a lesson, and number of students present at the beginning and end of the lesson. In addition, in Part A the observer evaluates students’ participation by indicating the percentage of students who were actively involved in the discussion. The participation information provided in Part A of the LIDO served as a predictor variable in Study 1 in order to investigate the relationship between classroom participation rate and persuasive writing quality.

The LIDO also captures information about the types of classroom activities and the format in which they were conducted (e.g. whole class, small group work, or pair work) (Part B). Also, the LIDO includes a diagram of the classroom that depicts where students were seated in relation to the audio equipment (Part C), and the research assistant’s assessment of engagement during small group and pair
work (Part D) (Elizabeth & O’Connor, 2014). Information from Parts B, C, and D were not used in this dissertation.

First and foremost, as the name suggests, the LIDO is a low-inference instrument. Observational tools that rely on complex inferences to make judgments about teacher and student intentions, using categories such as “encouragement” or “transactivity” (Berkowitz & Gibbs, 1983), are difficult to use and make establishing reliability a difficult task (Elizabeth & O’Connor, 2014). Conversely, zero-inference approaches that use measures such as words per utterance, student to teacher word ratios, and the number of transition words used during talk do not provide insight into the ways in which classroom talk advances during discussion (Elizabeth & O’Connor, 2014), nor do they help capture the various functions of student and teacher talk. As a low-inference measure, the LIDO can be used across grade levels to provide information about the degree to which teachers and students display the use of basic discourse moves. The LIDO provides information that can play a role in making inferences about instructional quality, presence of discussion, and productivity of conversations (Elizabeth & O’Connor, 2014).

**LIDO categories.** Part E of the LIDO classifies both student and teacher talk turns during whole-class discussion into various categories (Figure 1).

---

1 Part E of the LIDO involves off-site coding using audio or video recordings of classroom discussions. In this dissertation, all teacher-and student-discussion measures were extracted from Part E, which is henceforth simply referred to as “LIDO”.
Student turns are categorized into one of the following: 1) direct, student-to-student talk about content; 2) indirect student talk (students refer to each other’s ideas); 3) providing claims and presenting evidence or reasoning to support one’s claim; 4) providing extended responses (longer than a single clause); and 5) providing minimal responses that are a single clause or less.

Separately, each teacher turn is classified as one of the following six categories: 1) prompting student-to-student interactions about content; 2) following up with a particular student by asking for clarification, examples, or evidence; 3) using active listening to keep a student talking; 4) asking an open-ended, contestable question; 5) asking a semi-open question; and 6) asking a quiz-like question (Figure 1).

**Capturing Student vs. Teacher Talk.** There are important differences between teacher-produced and student-produced discussion features captured by the LIDO. Namely, teacher utterances are categorized into questions that are hypothesized to build students’ critical thinking skills (open-ended, contestable questions), and those that are not conducive to building these skills (quiz-like questions) (Alexander, 2008; Cazden, 1988). On the other hand, student utterances captured by the LIDO cannot be presumed to also indicate quality, as students are much less likely than teachers to produce on-topic utterances that also improve their thinking skills. Thus, while some student utterances captured by the LIDO may be dialogic in form (e.g. direct student to student talk), these utterances
cannot be assumed to be of high quality. For these reasons, it can be assumed that teacher scores captured by the LIDO are also indicative of quality of discussion, whereas student scores only reflect the form, rather than quality, of student discussions.

**LIDO validation.** Using classroom talk coded in three treatment groups (control classrooms not part of the Word Generation program, intervention classrooms, and classrooms where teachers who had previously used Word Generation were engaged in non-Word Generation discussions), the LIDO was validated against comparable components of the Classroom Assessment Scoring System (CLASS) for Upper Elementary (Pianta, Hamre, Haynes, Mintz, & LaParo, 2006). These sections are Quality of Feedback, Content Understanding, Instructional Dialogue, Instructional Learning Formats, Analysis and Inquiry, and Regard for Adolescent Perspectives. LIDO scores from both student and teacher talk were highly correlated with all six measures from the CLASS (O’Connor, LaRusso, & Harbaugh, 2016).

**LIDO coding.** The LIDO was used to code over 500 classroom discussions using audio recordings from the three aforementioned conditions. In order to accomplish this, 16 research assistants were recruited and trained to use the instrument to code talk that was produced by teachers and students during whole class discussions. Each discussion, therefore, was coded twice using separate indicators of talk – one focused entirely on teacher talk, and another capturing
student talk only. The following coding principles were established when coding the large set of discussion data collected over three years.

**Coding principle #1.** Each student or teacher turn is coded minute-by-minute using audio (and less frequently, video) data. When coding teacher turns, codes were assigned based on the specific part of a turn that ultimately prompts a student response, which is usually the last prompt in a teacher’s turn. To illustrate, in the following turn, the teacher asks multiple questions, including a contestable question (*Why do you think one is better than the other?*) and a semi-open question (*What do you remember about the government of Athens?*). The codable portion of this utterance, however, is “Raise your hand…,” a closed or quiz-like question, as that is the talk move that resulted in a response (students raising their hands) (Figure 1).

Teacher: What do you remember about the government of Athens? Were they better than Sparta? Why do you think one is better than the other? **Raise your hand if you agree that Athens was better.**

**Coding principle #2.** As the LIDO is a low-inference tool, coders were instructed not to make judgments or high-level inferences about what was said or to whom it was said, without concrete evidence. For example, when using the LIDO to code for student talk, it can be difficult to determine whether students engaged in talk about content are *speaking to one another directly* or whether they are *referring to each other’s ideas* but talking to the teacher or other students. As such, coders were instructed to mark an utterance as *student to student talk about*
content, the highest dialogic student talk move, only if the student used another student’s name in direct address, or a 2nd person pronoun in the utterance. If there was sufficient evidence to suggest that students were speaking to one another about content without the use of names or pronouns, the utterance was coded as student refers to another student’s ideas (Figure 1).

Coding principle #3. Both student and teacher discourse practices were coded based on a hierarchy of talk moves. When listening to classroom talk, research assistants were trained to code from top to bottom (Figure 1), while only assigning each utterance one code. For example, when coding for teacher talk, coders must first determine if a teacher’s utterance concerned content (as opposed to procedural or tangential talk), as only content-related utterances are codable. Next, coders determine if the codable utterance is prompting students to respond to one another, the highest dialogic category (Figure 1). In those cases, the utterance was assigned that code; if not, the coder assessed whether the utterance was a follow-up with a particular student.

Follow-ups with students could receive one of two LIDO codes: prompting a student to expand, explain, or clarify his/her response or using active listening techniques to get a student to continue speaking (Figure 1). If the utterance does not belong to any of the previous three categories of teacher talk but is a content-related question, then coders determined whether the utterance was one of the following question types: a) open, truly contestable question; b) semi-open question; or c) quiz-like question. Thus, each utterance was assessed from the top
of the pyramid to the bottom and only received one code (Figure 1). Student talk was coded in a similar way, with more dialogic student moves prioritized over those that are less dialogic (Figure 1).

In addition to the main coding principles outlined above, coders were provided with “rules” documents explaining each code in the student and teacher LIDO using various examples, as well as a flowchart (Table 1) that helps explain how to make decisions about how an utterance should be coded. Coders also went through several phases before they coded individually: first, they were trained by the research team over a period of two days, each focusing on either student or teacher coding. After each training day, coders were assessed and given feedback by the research team based on their responses. Coders were then paired together, and each pair underwent several rounds of training, where each person in the pair first coded the audio recording separately, then coders compared and discussed their codes to produce a “consensus” document. Using my own coding, I evaluated coders’ consensus files and provided feedback when there were disagreements between expert and novice coding. Following a number of rounds, coders were then assigned to independently code audio files using the LIDO.

***Insert Table 1 Here***

**Thesis Motivation and Organization**

The thesis begins with an empirical study that aims to explore the relationship between classroom-level discussion variables and student-level persuasive essays. Using a multilevel model, whole class discussions were coded
using the LIDO for both teacher and student talk, both of which were then divided into talk moves that are considered more dialogic and those that are considered less dialogic. Overall participation rate and its relationship to essays was also investigated.

To select the final sample, after all data in the CCDD study were coded using the LIDO, I identified as many recorded classroom discussions as possible for which there were completed essays. This proved to be a challenging task. Though there were 12 units in 4th and 5th grade and 24 units in 6th through 8th grade in the Word Generation program, with accompanying essays at the end of each unit, there were only two recorded whole classroom discussions per year in each classroom. Also, as with any program, fidelity issues meant that not all workbooks were completed – sometimes topics were discussed (and recorded) in whole class settings but post-discussion essays were not completed, and at other times, only a few students completed post-discussion essays. As such, in Study 1, only classrooms with four or more completed essays per classroom for which audio data were also available were used, yielding a total of 42 classrooms.

Next, the second empirical paper explored the relationship between student talk and writing using video data of one eighth grade classroom over eight sessions. Student talk was first coded using the LIDO, followed by analysis of whether individual students’ talk moves and participation levels predicted their post-discussion written essays, which were scored using two approaches. Findings concerning the role of high student dialogic discussion in Study 2 were consistent
with findings at the classroom level (Study 1). Each of the empirical papers are written for a particular journal in mind: Study 1 will be submitted to the *American Educational Research Journal* (AERJ), while Study 2 will be submitted to *Discourse Processes*.

In the third paper, I discuss the findings in the context of classroom practice, and offer instructional recommendations that may help teachers strengthen their discourse practices in hopes of improving persuasive writing. The dissertation ends with a concluding chapter that summarizes results from both empirical studies and discusses future research recommendations. Finally, the chapter suggests research approaches aimed at further understanding the relationship between classroom discourse and student outcomes by incorporating tools that enable the collection and management of large-scale classroom and student data.
References


Association, Washington, DC.


Tables and Figures

Table 1

Flowchart of Coding for Teacher Talk using the LIDO

1. Is the T’s codable utterance about the contents of instruction?
   Yes ↓  No → Not Codable

2. Is the Teacher’s codable utterance explicitly prompting students to respond to the contribution of another student? (e.g. do you agree or disagree with Jason? Who wants to respond to that? Can anyone add on to what she said?)
   No ↓  Yes → Prompts student interaction

3. Is the Teacher’s codable utterance addressed to the immediately previous student contributor?
   No ↓  Yes → T2  
   Does the Teacher’s utterance get the previous speaker to continue by repeating or rephrasing that student’s contribution (e.g. with a rising intonation, I see…) or by using other “active listening” techniques (e.g. say that again, okay)?
   No ↓  Yes → T3  
   Is the T’s utterance a completely new, unrelated question?
   Yes → T4, T5, T6 as applicable

4. Is the Teacher’s utterance a prompt that contains an open-ended query with unconstrained and contestable responses?
   No ↓  Yes → T4

5. Is the Teacher’s utterance a prompt that contains a semi-open question that is not very contestable but has some latitude for explanation?
   No ↓  Yes → T5

6. Is the Teacher’s utterance a prompt that contains a closed or display question?
   No ↓  Yes → T6

Go back and check again

---

2 This flowchart was developed by the LIDO team (led by Cathy O’Connor) and was used for training research assistants.
Figure 1. Student and teacher discussion talk moves captured by the Low Inference Discourse Observation Tool (LIDO).
Chapter 2: Study 1

Does discussion improve writing? A multilevel analysis of the relationship between classroom talk and persuasive writing in the middle grades

Introduction and Literature Review

Writing is an essential, yet difficult component of literacy. It is a cognitive, social, and linguistic task that requires writers to manage several demands at once, as writers seek to communicate their thoughts to a non-present audience (Gee, 2007). Students’ academic success is in part dependent upon their success in mastering academic writing, and in upper-elementary and middle school, students encounter a new and more challenging form – persuasive writing.

To persuade an audience, writers are required to take on a position, then present evidence and reasoning to support their position, while also considering counterarguments and offering rebuttals (Toulmin, 2003) – a process that places more cognitive and linguistic demands than narrative writing (Nippold, 2000; Salahu-Din, Persky & Miller, 2008). This genre is emphasized in the Common Core State Standards across subject areas (National Governors Association, 2010) and students are expected to master it by the end of high school (Hillocks, 2002; National Governors Association, 2010), in preparation for college entrance exams which often test students’ written skills through persuasive writing tasks (College Board, 2012). Beyond the classroom setting, persuasive writing also prepares students for the work environment and for tasks encountered in society at large (Nippold, 2000).
Writing can pose a challenge for students as it is a productive, rather than a receptive task (Kress, 1994). Additionally, many students struggle with proficiency in the persuasive writing genre (National Commission on Writing, 2006), perhaps because the demands of writing may not be taught explicitly. Not only is writing difficult to produce, but this complex task is oftentimes difficult to instruct, partly due to the lack of comprehensive research on writing instruction as well as the disconnect between research and classroom practice (Myhill and Fisher, 2010). As such, much is yet to be learned about classroom factors that can promote persuasive writing, a task that this paper aims to explore. One such classroom factor is talk.

Talk among students and between students and teachers is a perceptible feature of any given classroom. However, though language is the currency for learning, the type of language used matters, as not all forms of talk promote learning and thinking (Michaels, O’Connor, & Resnick, 2008; Cazden, 1988; Tharp & Gallimore, 1988). Bakhtin (1984) distinguishes between monologic talk, which positions the teacher as the possessor of knowledge, and dialogic talk, where truth is not “found inside the head of an individual person” but is instead co-constructed through the process of dialogue (Bakhtin, 1984, p.110). Classroom talk is considered dialogic when it supports learning and thinking, and “connotes social relationships of equal status, intellectual openness, and possibilities for critique and creative thought” (O’Connor & Michaels, 2007, p.277). Thus, in
dialogic discussions, both students and teachers construct meaning through dialogue.

Whether referred to as *accountable talk* (Michaels, O’Connor, & Resnick, 2008), *instructional conversations* (Cazden, 1988), *academically productive talk* (Chapin & O’Connor, 2007; Chapin, O’Connor, & Anderson, 2003) or *dialogic discussions* (Tharp & Gallimore, 1988), these forms of talk are illustrative of the Bakhtinian notion of dialogic discourse (Bakhtin, 1984), Vygotsky’s social constructivist theory (Vygotsky, 1978), Dewey’s call for inquiry-oriented learning (Dewey, 1938), and even Socrates’ emphasis on uncovering meaning through dialogue. They also stand in stark contrast to *recitation*, a common approach in which communication is controlled by the teacher, who often engages students in IRE sequences – that is, a teacher-initiated question, followed by a student’s response, and the teacher’s evaluation (Mehan, 1979).

Though common in Western education and certainly useful for certain purposes such as assessing students’ factual knowledge, IRE sequences have been heavily criticized for their inability to involve students in learning that promotes their thinking (Alexander, 2008; Cazden, 1988; 2001; Nystrand, 1997). It is argued that these linear, “pseudo-discussions” (Zwiers, 2014, p.123) contrast with discussions occurring in settings outside the classroom, where talk is circular, exploratory, and involves back-and-forth exchanges where speakers question, disagree, and clarify (Gilles, 1993).
Classroom talk is important for students in all subject areas, not only those that are traditionally more conducive to dialogue (e.g. Social Studies). Across subject areas, mastery of content involves academically productive conversations around content (Moje, 2007; Langer, 2010). Teachers’ prompts may facilitate more monologic talk in a given classroom if test-like, “known information questions” (Mehan, 1998) are posed to students. For instance, a prompt such as, “What was the Pharaoh’s name?” would elicit a minimal student response such as “Imhotep,” which, if followed with a teacher evaluation such as, “That’s right” would be an instance of the aforementioned IRE sequences. If quiz-like questions are at one end of the classroom discourse spectrum, “authentic questions” (Harvey & Goudvis, 2007) are at the other.

Unlike known questions, authentic questions are not asked to elicit correct responses and aim instead to explore students’ thinking and reasoning. These types of questions are most indicative of dialogic discourse, and teachers can encourage them by prompting students to engage with and build upon each other’s contributions. For example, a teacher prompt that asks, “Who agrees with Sarra and why?” would encourage a more open-ended, dialogic student response such as one that begins with, “I agree with Sarra” and continues with an explanation to substantiate the claim.

Though dialogic discussion is a desired medium through which students can develop various skills fundamental to their academic success, it is a rare occurrence in U.S. classrooms (Nystrand et al., 1997; Applebee, Langer, Nystrand,
In a study of 450 eighth and ninth grade classrooms, Nystrand and Gamoran (1991) found that 85 percent of instructional time involved seatwork, recitation, and lecture. In eighth grade, they found evidence of true, dialogic discourse in just 52 seconds per class hour, and only 15 seconds per hour in ninth grade. More recently, Nystrand, Wu, Gamoran, Zeiser, & Long (2003) noted in a study of over 200 classrooms in the United States that what most teachers identified as discussion was in fact “question-answer discussion,” or the aforementioned recitation sequences. Ketch (2005) also notes that adolescents are not provided with sufficient opportunities to engage in conversation that builds critical thinking skills. Teachers also report lack of time to initiate and facilitate dialogic discussions given high-stakes tests and demands such as the Common Core State Standards (Juzwick, Borsheim-Black, Caughlan & Heintz, 2013). Thus, in a classroom climate with a relative absence of dialogic discussions, students may not be fully prepared for discussions in less protected settings.

Dialogic talk provides students with opportunities to internalize and appropriate thinking processes and linguistic terms used during discussion (Bakhtin & Holquist, 1981; Vygotsky, 1978). Beyond theorization about its utility, however, empirical studies have shown that whole class dialogic discussion is related to content comprehension (Ketch, 2005), and development of deep reasoning and critical thinking skills (Michaels, O’Connor, & Resnick, 2008), all of which are cornerstones for success in school and beyond. However, there
remains a question about whether such appropriation manifests itself in students’ writing.

Few studies have empirically explored whether dialogic approaches contribute to students’ argumentation skills in writing; however, the evidence is accumulating. For example, Kuhn, Hemberger, and Khait (2016) have shown that *dialogic communication* (small group discussions, followed by electronic dialogues between student pairs, and culminating in a whole class discussion) in middle-school classrooms improved students’ writing quality. Persuasive essays collected from 38 students over two years were found to contain significantly more arguments from both sides of the issue, an evaluation of the opposing position, and the use of *however* to connect opposing views, when compared to pre-intervention essays. The authors contend that more evidence is needed before they can “prove that the dialogic aspect of [the] intervention contributed to students’ achievement” (p.115) but they draw their conclusions based on the transfer of students’ oral statements (e.g. “Others might say...”) to their writing. The same dialogic approach utilized in the previous study was also found to contribute positively to students’ argumentative discourse (Crowell & Kuhn, 2014) and students’ comprehension of argument (Kuhn, Zillmer, Crowell, & Zavala, 2013). Additionally, Kuhn and Crowell (2011) found that over time, argumentative reasoning skills were enhanced for students who took part in dialogic communication prior to writing on the same topic.
These findings indicate a link between dialogic discourse and persuasive writing on the same topic. This study aims to explore this relationship more explicitly, as specific features of dialogic discussion will be examined. Furthermore, whereas the aforementioned studies employ a unique discussion structure that may be difficult to implement in regular classrooms, this study will concentrate on whole class, teacher-facilitated discussions only, a structure that is familiar to students and teachers in K-12 classrooms. The hypothesis is that if students engage in discussions about a given topic prior to producing persuasive essays where they take a position on the topic, perhaps they will be able to create argument schemas that will then transfer from verbal argumentation to persuasive writing (Reznitskaya et al., 2001).

This study aims to empirically investigate the predictive association between various features of whole class dialogic discussion and students’ writing quality. Examining how classroom-level data predict students’ individual academic outcomes, this paper analyzes not only teacher-initiated discussion, but will also examine student talk during whole class discussion. Specifically, the following research question will be examined: When accounting for students’ prior academic language skills, race, gender, socio-economic status, special education status, and English learner status, are students’ persuasive essay scores predicted by:
a) teacher- and student- produced discussion features indicative of maximal dialogic discussion;

b) teacher- and student-produced discussion features indicative of minimal dialogic discussion;

c) overall student participation rate?

**Method**

**Research Setting**

All data analyzed in this study were gathered as part of the Catalyzing Comprehension through Discussion and Debate (CCDD) study, a large-scale intervention involving urban elementary and middle schools (grades 4-7) in the Northeastern United States. The CCDD study extended the original Word Generation program, a supplementary, cross-disciplinary curriculum developed in partnership with the Strategic Education Research Partnership that aims to improve students’ academic language and argumentation skills (Snow, Lawrence, & White, 2009).

The original Word Generation program consisted of weeklong units with lesson sequences that introduced middle school students to social or civic dilemmas through short activities, along with five academic words integrated into these activities. Throughout the week, students read materials related to a given topic, developed positions on the topic, identified reasons and evidence for their positions, and engaged in debates about the particular issue. At the end of a week of daily lessons focused on the same topic, students were asked to produce a
persuasive essay in response to a prompt that asked them to take a stance on the given issue. They were also strongly encouraged to incorporate the taught academic vocabulary into their discussions and written essays (Snow, Lawrence, & White, 2009).

Data analyzed in this paper are from the enhanced Word Generation program. First, Word Generation Elementary (for fourth and fifth grade) extended the program to upper elementary grades, with 10-day long units involving cross-content-area activities that replicated the structure of the original Word Generation program. Furthermore, six-week long units in science and social studies were developed for middle-school students (Science Generation and Social Studies Generation). Thus, sixth and seventh grade students in the enhanced Word Generation program participated in 12 of the original Word Generation weekly lessons, in addition to 12 weeks of science and social studies units (with each lesson lasting approximately 45-minutes per day).

All Word Generation topics examined in this study are presented in Table 1 below. Essay and discussion topics include debates on whether everyone should learn a second language (grade four), why people fight (grade five), whether laws should regulate our use of paper or plastic (grade six), and whether physician-assisted suicide should be allowed (grade seven). It is important to note that fidelity of implementation varied across schools; thus, essays per classroom ranged from minimal (4 essays) to full participation (22 essays). Out of the 25 topics in the sample, the topic “Why do we buy what we buy?” produced 45
essays (from three classrooms), while “Pyramids and other monumental structures: Great achievements or a waste of Egypt’s surplus?” and “How do we test that idea?” produced the fewest essays (with 4 essays each) (Table 1).

***Insert Table 1 Here***

Participants and Data

A total of 42 audiotapes of classroom discussions from two districts were analyzed, along with their corresponding essays (n=471). As the most dominant racial group was Black/African American (41.7%), all other racial groups are discussed in reference to Black students. A high percentage of students in this sample (86.5%) were eligible for free or reduced lunch. Other student demographic data are presented in Table 2. While all students benefit from dialogic classroom practices, students from low socioeconomic backgrounds (i.e. the majority of students in this sample) are especially positioned to benefit more from classroom practices centered on discussion (Kong & Pearson, 2003). As such, it is especially important to understand how participants who have lacked opportunities that lead to success on academic measures can benefit from dialogic practices.

***Insert Table 2 Here***

---

3 There were 369 unique students, 25 different topics, and 24 different teachers in this sample
Measures

**Outcome variable.** Students’ post-discussion persuasive essays were scored holistically from 1-4, with 4 representing essays that included a claim(s), evidence or reasoning to support the claim, acknowledgement of counter-claims, and a stated conclusion. Incomplete essays or those which were written about an irrelevant topic were excluded. Essays with a score of 1 often included an unsupported claim or a series of claims, while essays with a score of 2 often included a claim and reason(s)/evidence to support the claim. A score of 3 indicated presence of three out of the four categories (claim with evidence and a conclusion, or claims with evidence and counterclaims.) High inter-rater reliability between two scorers was achieved using 20% of essays in the sample (n=95), resulting in a Cohen’s kappa of 0.96.

**Classroom discussion measure (predictor).** All audio recordings of classroom discussions were coded for discussion features by a team of trained research assistants using the Low-Inference Discourse Observation tool (LIDO) (O’Connor & LaRusso, 2014) (Figure 1). The LIDO captures both teacher and student discourse features and was developed through an iterative process that started from understandings about dialogic discourse from work by Michaels, O’Connor, and Resnick (2008) on accountable talk (see also O’Connor & Michaels 2007), as well as measures of classroom talk identified by researchers studying K-12 classroom discourse (Mercer, Wegerif, & Dawes, 1999; Nystrand et al., 2003).
The LIDO (Figure 1) was validated using the Classroom Assessment Scoring System (CLASS) for Upper Elementary (Pianta, Hamre, Haynes, Mintz, & La Paro, 2006), a well-known observational instrument that assesses classroom interactions. Because not all elements of the CLASS were relevant for LIDO, the following six CLASS scales were used: Quality of Feedback, Content Understanding, Instructional Dialogue, Instructional Learning Formats, Analysis and Inquiry, and Regard for Adolescent Perspectives. CLASS scores were then correlated with the Teacher and Student Logit scores calculated from the LIDO using 512 treatment and control files, and all correlations were found to be statistically significant (p<.001) (O’Connor, LaRusso & Harbaugh, 2016).

Tharp and Gallimore (1991) argue that classroom talk should be used to explore students’ ideas, rather than simply require students to regurgitate answers already known to the teacher through the use of quiz-like, closed questions. As shown in Figure 1 above, the LIDO distinguishes among three types of teacher questions: those that promote discussion and critical thinking (teacher poses contestable question), those that prompt a known answer (teacher poses quiz-like question), and questions that are not truly contestable but have a broader set of responses than closed questions (teacher poses semi-open question). Additionally, the LIDO includes categories that indicate whether student interaction is prompted (prompts student interaction), as well as a teacher’s use of clarification and active listening with a particular student (prompts for clarification and uses active
listening) (Figure 1). Thus, the LIDO captures forms of teacher talk that both explore students’ ideas and those that test knowledge, both of which are present in any given classroom discussion, though the latter is typically predominant (Nystrand & Gamoran, 1991).

Teacher discussion features at the top of the pyramid in Figure 1 are hypothesized to be more indicative of dialogic discussion as they are intended to prompt students to think more critically and engage with others on a given topic. A teacher may prompt for students to engage with one another by asking questions such as, “Who wants to agree or disagree, or add on to that?” Also, a teacher may ask for clarification or explanation by asking a student, “How did you come to that conclusion?” or “What’s your evidence?” On the other hand, teacher-produced discussion features at the bottom of the pyramid are hypothesized to be less indicative of dialogic discourse, as they are less likely to engage students in extended discussion. For example, a teacher may pose a quiz-like question such as, “What is the opposite of negative six?” This question elicits a one-word response evaluated with respect to a correct answer, and therefore cannot be considered dialogic (Figure 1).

Student-produced discussion features at the top of the pyramid are hypothesized to be more indicative of dialogic discussion, as they show that students are engaged with others and with the subject matter at a level that would allow them to deepen their understanding on a given topic (Figure 1). A student may directly engage with another student by stating, “I agree with what you said
because…” Additionally, a student may refer to another student’s ideas by stating, “I disagree with Sarra. I think the rulers were actually really harsh.” On the other hand, student discussion features at the bottom of the pyramid represent student input that does not take place in the context of an extended, student-focused dialogue (Figure 1). At their extreme, these responses include minimal responses such as “Yes/No” or responses in minimal clauses such as “He was a bad ruler.”

***Insert Figure 2 Here***

Presence of teacher- and student-produced discussion features was coded minute-by-minute by trained coders. Using the LIDO (Figure 1), teacher and student discussion features were broken up into two categories to capture both maximal and minimal levels of teacher-produced dialogic discussion (MaxTD and MinTD) and two measures of student-produced discussion features (MaxSD and MinSD) (Figure 2). To account for variability in classroom discussion time, features in each category were summed and divided by classroom discussion time.

**Participation rate (predictor).** An additional component of this study, student participation rate, was captured by research assistants present during in-class observations of discussions. Low levels of participation (0-25%) were assigned a score of 1; participation rates of 25-50% were assigned a score of 2, participation rates of 50-75% were given a score of 3; and finally, participation rates of over 75% were assigned a score of 4.

**Academic language (control).** Prior to discussions, students were given a validated measure of overall academic language skills, the Core Academic
Language Skills Instrument (CALS-I) (Uccelli, Barr, Dobbs, Galloway, Meneses, & Sanchez, 2015). The language skills captured by the CALS-I are defined as “knowledge and deployment of a repertoire of language forms and functions that co-occur with oral and written school learning tasks across disciplines” (Uccelli et al., 2015, p.1079). Specifically, the instrument assesses students on six skills: unpacking complex words (e.g. morphological skills), comprehending complex sentences (understanding complex syntax), connecting ideas (knowing how to use connectives and discourse-markers), tracking themes (anaphoric resolution), organizing argumentative texts, and awareness of academic register, which was measured in two ways (Uccelli et al., 2015).

**Demographic variables (controls).** Other student-level variables are students’ socioeconomic status (SES) as indicated by receipt of free or reduced-price lunch, students’ gender (Female), race (Black, Latino, White, and Other, with Black as a reference category), English language learner status (ELL), and special education status (students with an individualized education plan).

**Fitted model.** Given that certain topics are likely to elicit more discussions than others, which could then impact initial essay score (intercept), all models fitted in this study have a random intercept by topic. Intercepts were also allowed

---

4 Anaphors refer to words or phrases that refer to a previously stated idea or person (e.g. ‘she’ refers to ‘Sarra’). Anaphors in academic texts can be conceptual (e.g. ‘this’ refers to “The discovery of insulin.”)

5 Students were asked to identify the academic register (i.e. academic vs. colloquial definition) and were also asked to produce the academic register (i.e. they were asked to write definitions of four words for a dictionary for adults.)
to vary randomly by teacher so as to account for any possible teacher effects.

Additionally, I fitted a random slope for participation rate by topic in order to account for the possibility that different topics might elicit different levels of participation, which may then also influence the relationship between participation and essay score (slope). The following is a 2-level multilevel model:

**Level-1 model (Individual-level):**

\[ E_{\text{Score}}_{ij} = \beta_{0j} + \beta_1 AL_{ij} + \beta_2 SES_{ij} + \beta_3 ELL_{ij} + \beta_4 SPED_{ij} + \beta_5 White_{ij} + \beta_6 Latino_{ij} + \beta_7 0th\_Race_{ij} + \beta_8 Female_{ij} + \epsilon_{ij} \]

\[ \epsilon_{ij} \sim N(0, \sigma^2_\epsilon) \]

**Level-2 model (Group-level):**

\[ \beta_{0j} = \gamma_{00} + \gamma_{01} MaxTD_j + \gamma_{02} MaxSD_j + \gamma_{03} MinD_j + \gamma_{04} PAR_j + u_{0j} \]

\[ u_{0j} \sim N(0, \sigma^2_u) \]

Where:

- \( E_{\text{Score}} \) is the predicted essay score of student \( i \) in classroom \( j \)\(^7\)
- \( \beta_{0j} \) denotes a random intercept for each classroom and is further described by the level-2 model\(^8\)
- \( \beta_1 - \beta_8 \) are individual-level controls (369 observations)
- \( \gamma_{00} \) is the overall intercept across all groups
- \( \gamma_{01} - \gamma_{04} \) are the classroom predictors (42 observations):
  - \( MaxTD \) is Maximal Teacher Dialogic Discussion
  - \( MaxSD \) is Maximal Student Dialogic Discussion
  - \( MinD \) combines both Minimal Teacher Dialogic Discussion (\( MinTD \)) and Minimal Student Dialogic Discussion (\( MinSD \)) due to collinearity
  - \( PAR \) is overall student participation rate
- Errors are assumed to be normally distributed at both the student and classroom levels and are denoted by \( \epsilon_{ij} \) and \( u_{0j} \) respectively

\(^6\) Although there are 42 classrooms in this sample, there were 24 unique teachers teaching those lessons. Additionally, of the 471 student essays, 369 were produced by unique students.

\(^7\) Student outcomes are influenced by both student-level predictors (e.g. \( AL_{ij} \)) and classroom-level predictors (e.g. \( MaxTD_j \)).

\(^8\) Coefficients describing the magnitude and polarity of the relationships between predictors and essay scores are represented by \( \beta \) coefficients for student-level predictors (e.g. \( \beta_2 SES_{ij} \)) and \( \gamma \) coefficients for classroom-level predictors (e.g. \( \gamma_{01} MaxTD_j \)).
Results

Summary statistics for student- and classroom-level variables are shown in Table 3. Over 75% of students received a score of 2 (n=241) or 3 (n=122) out of 4 on their essays, while the rest either scored a perfect score (n=65) or the lowest score (n=43) (Figure 3). The mean essay score was 2.4 out of 4 ($SD = 0.84$), or 60%.

***Insert Figure 3 Here***

Measures of minimal dialogic discussion were more prevalent, on average, than those representing maximal dialogic discussion (Table 3). Also, unlike measures of minimal dialogic discussion, some classrooms did not contain any observations of maximal teacher or student dialogic discussion. That is, in some classrooms, there were no instances of the teacher asking students to engage with one another, prompting a student to clarify, using active listening, and asking a contestable question. Similarly, in some classrooms, students did not engage with each other directly or indirectly, and they did not state and support a claim. In those classrooms, teachers only asked semi-open and quiz-like questions, and students only responded with extended or minimal responses.

There was high student participation (75-100%) in 40.5% of classrooms (n=17), medium participation (50-75%) in 33.3% of classrooms (n=14), 25-50% in 21.4% of classrooms (n=9), and very low (0-25%) in only two of the 42
classrooms (4.8%). Whole class discussion time ranged from 14 to 76 minutes, with a mean of 34 minutes ($SD = 11$).

***Insert Table 3 Here***

**Correlations**

Next, correlations among variables of interest were explored (Figure 4).

***Insert Figure 4 Here***

**Correlations between outcome and main predictors.** As shown in Figure 4, the strongest correlation between the predictors and essay score is observed with participation rate (PAR), where there is a statistically significant, positive relationship ($r = 0.28, p < .001$). Also, statistically significant negative correlations were found between both measures of minimal dialogic discussion: minimal teacher dialogic discussion (MinTD) ($r = -0.20, p < .001$) and minimal student dialogic discussion (MinSD) ($r = -0.14, p < .01$). The direction and magnitude of these correlations are an early indication of the hypothesized relationship between minimal dialogic discussion and essay scores. However, there was no correlation between essay scores and measures of maximal dialogic discussion.

**Correlations between outcome and controls.** The strongest correlation is observed with prior academic language skills (AL), where there is a positive, statistically significant correlation ($r = 0.30, p < .001$). Other control variables that statistically significantly correlated with the outcome are special education
status (SPED) \((r = -0.11, p < .05)\) and socioeconomic status (SES) \((r = -0.11, p < .05)\).

**Correlations among variables.** The correlation matrix indicates a statistically significant, strong, positive correlation between the two measures of minimal teacher and student dialogic discussion (MinTD and MinSD) \((r = .88, p < .001)\). As this suggests collinearity, both variables were combined to produce overall minimal dialogic discussion (MinD), which will henceforth be used in this study (see summary statistics in Table 3). Although the two measures of maximal dialogic discussion (MaxTD and MaxSD) are positively correlated with one another \((r = 0.44, p < .001)\), the correlation is not high enough to indicate collinearity and the variables were thus left as separate measures of maximal dialogic discussion. There is also a strong, positive relationship between maximal student dialogic discussion (MaxSD) and participation rate (PAR) \((r = 0.27, p < .001)\).

**Multilevel Modeling Results**

**Predictors.** A multilevel model with all predictors and controls was fitted to answer the research question outlined above (Table 4, Model 3). All measures of dialogic discussion were regressed on essay scores using a random intercept (by topic, teacher, and student), random slope (by participation rate and topic) multilevel model to examine which measures of discussion predict essay scores. Beginning with measures of maximal dialogic discussion, the hypothesized
relationship between maximal teacher dialogic discussion and essay scores was confirmed, as it was found to positively predict essay scores when accounting for other measures of discussion, participation rate, prior academic language, and students’ demographic features ($\beta_{MaxTD} = 0.53, p < .01$). Though findings are per minute, we will discuss these findings per 10 minutes of discussion for ease of interpretation. Thus, on average, for every additional maximal teacher dialogic discussion move per 10 minutes of discussion time, we observe an increase in essay scores by 1.33%.

Consistent with the hypothesis that minimal dialogic discussion would be negatively related to writing, results show that (teacher-and student-produced) overall minimal dialogic discussion was associated with a small, but statistically significant decrement in essay scores, when accounting for other measures of discussion, student participation rate, prior academic language scores, and demographic variables ($\beta = -0.11, p < .01$). Thus, on average, every additional occurrence of minimal dialogic discussion moves per 10 minutes of discussion was associated with a slight decrease of 0.28% in essay scores.

While the findings so far confirmed the outlined hypotheses, the relationship between maximal student dialogic discussion (MaxSD) and essay scores did not align with the hypothesis that higher dialogic talk would be associated with higher quality essays. In this study, maximal student dialogic

---

9 As essays are scored out of 4, $0.53 \div 4 = 0.133$ (13.3% increase per minute or 1.3% increase per 10 minutes of discussion time.)
discussion was found to *negatively* predict student essays when accounting for other discussion features, participation rate, prior academic language, and demographic variables ($\beta_{\text{MaxSD}} = -0.47, p < .01$). Thus, on average, each additional occurrence of MaxSD per 10 minutes of discussion time was associated with a decrease of 1.2% in essay scores.

Given the statistically significant correlation between MaxTD and MaxSD ($r = 0.44, p < .001$), two additional multilevel models were fitted to test whether results still held when only one of these two predictors were present. As shown in Table 4 (Model 1), maximal teacher dialogic discussion was still a significant predictor of essay scores ($\beta_{\text{MaxTD}} = 0.42, p < .05$). On the other hand, in the absence of MaxTD, maximal student dialogic discussion was no longer a significant predictor of essay scores, though the sign was still negative ($\beta_{\text{MaxSD}} = 0.16, p = 0.32; CI[-0.49,0.22]$) (Model 2). However, given that all discussion variables co-occur in classroom discussions, the final model chosen was the one containing all student- and teacher-produced discussion predictors (Model 3). The statistically significant negative relationship found between MaxSD and essay scores in the final model is no doubt counter-intuitive; however, it may be explained by various factors. For instance, this predictor does not distinguish between classrooms in which higher dialogic input was the result of few students’ interactions with each other and the teacher, or whether it was resulting from interaction among most students. Furthermore, while teacher
discussion features captured by the LIDO can be an indication of quality, those produced by students are indicative of dialogic form, not quality. An additional consideration is that the LIDO only captured student discussions that occurred in whole class settings; discussions in small groups, on the other hand, were not coded using the LIDO. These factors are further elaborated upon in the Discussion section.

Finally, even when the relationship between participation rate and essay scores was allowed to vary by topic, a statistically significant relationship between participation rate and essay scores was found ($\beta_{PAR} = 0.23, p < .05$). Thus, on average, every one-unit increase in overall student participation rate (e.g. from 25-50% to 50-75%) is associated with an increment in essay scores of 5.8%. In other words, students who were in classrooms characterized by high participation rates produced essays that were scored significantly higher, regardless of topic.

**Controls.** The only statistically significant relationship between control variables and the outcome was found between prior academic language skills (AL) and essay score ($\beta_{AL} = 0.18, p < .001$). Thus, socioeconomic status, gender, race, special education status, and English learner status were not statistically significantly associated with essay scores.

***Insert Table 4 Here***
Discussion

Classrooms are complex, dynamic environments that can be difficult to study. While they share many similarities, salient variations such as classroom size, and less tangible variations such as classroom culture exist. This study sought to explore one such factor that, although common among all classrooms in this sample, was implemented differently from one classroom to another. In the 42 classrooms analyzed in this paper, the utilization of classroom talk varied in that some classrooms were set up in ways that promoted dialogic interactions, while others were not. Even when discourse was dialogic, it was not conducted uniformly across all classrooms, as teachers varied widely in how they facilitated classroom talk and which formats they used (e.g. fishbowl discussion vs. all-class). Nonetheless, the study aimed to control as many variables as possible in order to test the question, *What is the relationship between classroom discussion and writing?*

In addition to classroom differences, there were various individual differences among students. While most students in this sample came from low-income households, students differed in their race/ethnicity, gender, special education status, English learner status, and prior academic language skills. Though other differences certainly exist among students, these variables were accounted for so as to understand the role of discussion in writing, regardless of the racial, socioeconomic status, gender, and academic backgrounds of students.
To our knowledge, this is the first empirical investigation of the simultaneous occurrence of teacher and student talk in a whole class setting, as well as its relationship to persuasive writing. We present a new approach to documenting whole class discussion (the Low-inference Discourse Observation tool), and control for demographic variables as well as students’ prior academic language skills (using the CALS-I). The outcome in this study was also scored using a simple rubric that teachers can easily utilize to assess the quality of their students’ persuasive writing. Thus, unlike other studies in this field, we examine an approach already occurring in K-12 classrooms everywhere (whole class discussions) and discuss concrete ways to use it to improve students’ writing outcomes.

The findings support the theoretical assumptions that place semi-open and quiz-like teacher questions as well as elaborated and minimal student responses at the lower end of the dialogic spectrum, and assumes that these discussion features are not as conducive to higher academic outcomes (in this case, persuasive writing). Similarly, the findings partially support the hypothesis that higher dialogic discussion features promote student writing, although in this analysis, only those features produced by teachers positively predict writing (Table 4). Also, student participation positively predicted essay scores, even when taking into account the topics discussed. Though the study does not allow for causal inferences, these results highlight possible pedagogical approaches (e.g.}
prioritizing open-ended, contestable questions) that may be beneficial for promoting students’ writing outcomes.

It is also worth noting that the positive effects of maximal teacher dialogic discussion mattered more for essay scores than the negative effects of minimal dialogic discussion. To illustrate, Figure 5 below examines the expected change in essay scores in four classrooms observed in this sample.

***Insert Figure 5 Here***

As illustrated in Figure 5, in ten minutes of discussion time, when 1.71 occurrences of discussion features indicative of dialogic discussion are observed, the associated gain in essay scores is 2.3%. To observe a 2.1% decrease in essay scores, however, 7.44 occurrences of discussion features indicative of low dialogic discussion need to be observed. Similarly, 11.71 occurrences of MaxTD are associated with essay gains of 15.8% in essay scores, but for a similar decrement in essay scores to occur, we need to observe nearly 60 occurrences of minimal dialogic discussion per 10 minutes of discussion time (Figure 5). Thus, a small amount of dialogic discourse is associated with positive essay scores, while much higher instances of IRE-type discourse is needed to negatively impact essay scores in the same way.

If found to be causal, this finding is encouraging as teachers can be encouraged to add to their repertoire of questions or discussion moves during a whole class discussion. For example, one component of teacher talk captured by the Low-Inference Discourse Observation tool is teacher questions: a) truly open,
contestable questions; b) semi-open questions that provide more latitude but are not truly contestable in nature; and c) closed, quiz-like questions whose answers are already known to the teacher. The first type of teacher questions, contestable questions, are highly dialogic in that they prompt a wide range of responses – that is, there are no “correct” responses to a dialogic question. These types of questions have been shown to be positively associated with essay scores (as part of maximal teacher dialogic discussion). One mechanism through which this may occur is that dialogic questions, when asked in a whole class setting, invite students to explore various ideas and perspectives on a given topic. For example, when a teacher asks, “Why do you think the Pharaohs were good leaders?” or “Why do you think everyone should learn a second language?” students will explore multiple views and opinions.

Contestable questions also require students to use evidence and/or reasoning to support their claims, as the very nature of these questions does not accept a response that is not supported by evidence or reasoning. That is, it is not sufficient to answer a question like “Why do you think everyone should learn a second language” by saying, “Yes, I agree that learning a second language is important.” In order to fully answer the question, students must support their responses with textual or other forms of evidence and reasoning.

Additionally, when asked in whole class settings, contestable questions expose students to various viewpoints that they may have otherwise not encountered or considered. When writing persuasively, exposure to different
perspectives may in turn allow students to defend their positions more strongly, as they can draw upon multiple viewpoints which they can then either support or refute using evidence and reasoning they may have practiced during whole class discussion. It can be argued that even students who are not active participants may benefit from being part of classrooms that prioritize contestable questions, as their simple presence during a dialogic discussion can help them gather evidence and viewpoints needed for persuasive writing.

In comparison to contestable questions and quiz-like questions, semi-open questions offer more latitude in that they generally deal with processes and procedures. They do, however, still contain a set number of responses, and are therefore not truly contestable as they aim to arrive at a “correct” or agreed-upon response. Like quiz-like questions, semi-open questions were found to be negatively associated with essay scores. This is perhaps because the scoring system used for essays in this paper did not value additional claims or additional pieces of evidence; that is, essays with one claim and one supporting bit of evidence gained the same score as essays with multiple claims and multiple supporting bits of evidence or reasoning. Semi-open questions may elicit elaborated and minimal responses that students can in turn draw upon when writing, but they may not necessarily help students incorporate (and defend or undermine) other viewpoints and perspectives. Thus, it could be hypothesized that one mechanism for improving persuasive writing would be for teachers to adjust their instructional practices to include more contestable questions and fewer semi-
open or quiz-like questions, they may help students acquire the skills necessary to improve their persuasive writing.

This paper also shed light upon another classroom variable, student participation rate during discussion, and its relationship with post-discussion writing. The statistically significant, positive association found in this study highlights that simply being part of classrooms in which many students participate in talk can positively relate to students’ persuasive writing. Importantly, this finding was independent of topic, all teacher- and student-produced predictors of discussion, student demographic variables, and their prior academic language skills. Thus, by simply encouraging more students to participate during a discussion, teachers may be able to improve students’ post-discussion persuasive writing.

One finding that did not align with the hypothesis was the negative relationship between maximal student dialogic discussion and writing. The student discussion features coded as dialogic using the Low Inference Discourse Observation tool are direct and indirect student-to-student talk about content, as well as stating and supporting one’s own claims with evidence or reasoning. Like maximal teacher dialogic discussion features, these student discussion features were hypothesized to positively predict writing. Though puzzling at first glance, this finding can be explained by various coding-related issues as well as differences between student- and teacher-produced utterances captured by the LIDO.
First, coders did not distinguish between classrooms where a maximal student dialogic feature (and resulting MaxSD score) was a result of interaction among several (or all) students in a classroom, and one that featured one or very few student interlocutors. Thus, if a teacher followed up with one student for several turns to push for reasoning or evidence, that classroom would have a high MaxSD score. This score may be the same as one produced in a classroom where multiple students referred to each other’s ideas during the discussion, or one where the teacher consistently followed up with several students to encourage them to explain or clarify their thinking. This could be further investigated in future studies.

This finding can also be related to what the LIDO does and does not reflect when capturing student and teacher talk. The LIDO captures participation and dialogicity, both of which are prerequisites for high quality discussions. On the other hand, the LIDO is not intended to capture content-related measures of quality. However, given that teachers are much more likely than students to make more content-rich and topic-relevant utterances, their high dialogic scores will also indicate good quality. On the other hand, we cannot assume that all student utterances are also an indication of quality, even if their form indicates potential for dialogicity. This necessitates qualitatively examining student discussions.

Another key difference between teacher- and student-produced utterances is that the LIDO distinguishes among types of teacher questions that are more likely to help build students’ critical thinking (contestable questions), and those that are
less likely to produce the desired outcome of building students’ critical thinking and reasoning skills (quiz-like questions) (Alexander, 2008; Cazden, 1988). This is further indication that unlike scores produced by students, high dialogic scores produced by teachers are also reflective of high quality discussions.

Finally, the LIDO does not capture an important aspect of student discussion – those that occur in small peer groups. Thus, teacher utterances captured by the LIDO are more representative of the teacher’s participation during a given lesson than student utterances, as the latter only includes what students said during whole-class interactions. Capturing student discourse in both small group settings and whole is necessary to fully understand the relationship between student discussion and outcomes such as writing.

Limitations and Future Directions

Several limitations exist in this study. Given that this study employs data from the intervention group alone (due to the unavailability of post-discussion essays in control classrooms), teachers are likely aware of the Word Generation program’s overall goals of fostering comprehension through discussion and debate. Thus, it would be worth comparing these results to classrooms that are not implementing a discussion-based curriculum in order to analyze whether results found in this study are replicated in classrooms that are more representative of average (non-discussion based) classrooms. It is important to note, however, that even in this sample, there was wide variation in what teachers considered
‘discussion,’ as some classrooms were characterized by high levels of dialogic talk while others were characterized by high levels of non- or minimally dialogic teacher talk.

Other limitations pertain to the use of one data-point as a proxy for students’ participation and one for their writing quality. Students may be more engaged on a given day and less engaged on another; as such, one observation may not be reflective of their typical participation levels. Furthermore, using one data-point as an indication for a teacher’s dialogic discussion score is not ideal, as an audio-recorded discussion on any given day may not reflect a typical classroom discussion. Longitudinal studies that explore various discussion factors, including classroom talk and participation levels over time, may help solidify our understanding of the relationship between talk and writing.

Additionally, students’ essays were first drafts and therefore may not be indicative of their overall writing quality (this may explain the relatively low mean essay score of 60%). Allotting sufficient time for writing, as well as navigating students through the writing process of drafting, editing, and re-writing, can help students develop essays that are higher quality (i.e. more persuasive).

To further explore the negative relationship between measures of high student dialogic discussion and writing, a follow-up study that uses video data to understand the breadth and depth of each student’s input during discussion and examine its relationship to the student’s post-discussion essay was conducted. By considering the proportion of talk of each student relatively to his or her
classmates, the second study sought to investigate how proportion of talk (rather than simple, overall participation), as well as dialogic and less dialogic student discussion features, are related to writing outcomes.

Additionally, while the follow-up study maintained the scoring scheme presented in this paper, it also included an additional scoring method that specifically explores students’ argument types. Using Kuhn and Crowell’s (2011) coding scheme will help expand our understanding of whether and how various features of student discussions during whole class discussion are related to their post-discussion essays.

**Conclusions**

Today’s learning environments are characterized by an increased utilization of online platforms, either as a replacement to traditional classrooms or in tandem with existing ones. Thus, in an era where education has found a home online, through lectures, massive open online courses (MOOCS), and other platforms, discussion may become the most distinguishing aspect of traditional classrooms. Furthermore, with the increased emphasis on discussion and writing in all subject areas (National Governors Association, 2010), it follows that the relationship between these cornerstones of academic success should be more carefully examined. We hope that future empirical work will further investigate both student and teacher talk in various whole class settings to help uncover the best ways of improving students’ literacy outcomes.
References


Gilles, C. (1993). *We make an idea: Cycles of meaning in literature discussion*
groups. In K. Pierce & C. Gilles (Eds.), *Cycles of meaning* (pp. 199-217). Portsmouth, NH: Heinemann.


Ure, J. (1971). “Lexical density and register differentiation.” In *Applications and


### Tables and Figures

**Table 1**

*Topics by Grade Level (471 Essays Produced in 42 Classrooms Discussing 25 Topics)*

<table>
<thead>
<tr>
<th>Grade</th>
<th>Topic</th>
<th>Essays</th>
<th>Classes</th>
<th>Prompt</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>Purchasing</td>
<td>45</td>
<td>3</td>
<td>Why do we buy what we buy?</td>
</tr>
<tr>
<td></td>
<td>Language</td>
<td>36</td>
<td>2</td>
<td>Should everyone learn a second language?</td>
</tr>
<tr>
<td></td>
<td>Origins</td>
<td>35</td>
<td>2</td>
<td>Who cares where we’re from?</td>
</tr>
<tr>
<td></td>
<td>Clothing</td>
<td>31</td>
<td>4</td>
<td>Why do we wear what we wear?</td>
</tr>
<tr>
<td></td>
<td>Nutrition</td>
<td>24</td>
<td>2</td>
<td>Who should decide what we eat?</td>
</tr>
<tr>
<td></td>
<td>Uniforms</td>
<td>22</td>
<td>1</td>
<td>Should students be required to wear uniforms?</td>
</tr>
<tr>
<td></td>
<td>Fairness</td>
<td>20</td>
<td>1</td>
<td>What is fair?</td>
</tr>
<tr>
<td></td>
<td>Behavior</td>
<td>9</td>
<td>1</td>
<td>Should students share responsibility for each other’s behavior in school?</td>
</tr>
<tr>
<td>5</td>
<td>Fighting</td>
<td>36</td>
<td>3</td>
<td>Why do we fight?</td>
</tr>
<tr>
<td></td>
<td>Privacy</td>
<td>35</td>
<td>2</td>
<td>Do we need to give up our privacy to protect our communities?</td>
</tr>
<tr>
<td></td>
<td>Inclusion</td>
<td>33</td>
<td>5</td>
<td>Should everyone be included?</td>
</tr>
<tr>
<td></td>
<td>Impact</td>
<td>31</td>
<td>2</td>
<td>How can the words and ideas of one person impact a community?</td>
</tr>
<tr>
<td></td>
<td>Caring</td>
<td>12</td>
<td>2</td>
<td>Why should I care?</td>
</tr>
<tr>
<td></td>
<td>Divisions</td>
<td>10</td>
<td>1</td>
<td>What divides us and how can we resolve our differences?</td>
</tr>
<tr>
<td></td>
<td>Happiness</td>
<td>10</td>
<td>1</td>
<td>Why do communities have different ideas about what brings happiness?</td>
</tr>
<tr>
<td></td>
<td>Power Laws</td>
<td>6</td>
<td>1</td>
<td>The power of Power</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5</td>
<td>1</td>
<td>Do we need laws to regulate our behavior?</td>
</tr>
<tr>
<td>6</td>
<td>Plastic</td>
<td>9</td>
<td>1</td>
<td>Should our use of paper or plastic be an individual choice or be regulated by the government?</td>
</tr>
<tr>
<td></td>
<td>Steroids</td>
<td>6</td>
<td>1</td>
<td>Steroids: Substance abuse or an innocent boost for athletes?</td>
</tr>
<tr>
<td></td>
<td>Ideas</td>
<td>4</td>
<td>1</td>
<td>How do we test that idea?</td>
</tr>
<tr>
<td>7</td>
<td>School</td>
<td>17</td>
<td>1</td>
<td>Is an extended school day the right choice for U.S. students?</td>
</tr>
<tr>
<td></td>
<td>Suicide</td>
<td>13</td>
<td>1</td>
<td>Should doctors be allowed to assist seriously ill patients with suicide?</td>
</tr>
<tr>
<td></td>
<td>Units</td>
<td>12</td>
<td>1</td>
<td>So what exactly is a unit?</td>
</tr>
<tr>
<td></td>
<td>Leaders</td>
<td>6</td>
<td>1</td>
<td>The Pharaohs of ancient Egypt: Protectors of order or sources of oppression?</td>
</tr>
<tr>
<td></td>
<td>Spenders</td>
<td>4</td>
<td>1</td>
<td>Pyramids and other monumental structures: Great achievements or a waste of Egypt’s surplus?</td>
</tr>
</tbody>
</table>
Table 2

*Student Demographic Data (N_{total} = 369)*

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eligible for free or reduced lunch</td>
<td>319</td>
<td>86.50</td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>154</td>
<td>41.73</td>
</tr>
<tr>
<td>White</td>
<td>106</td>
<td>28.73</td>
</tr>
<tr>
<td>Latino</td>
<td>89</td>
<td>24.12</td>
</tr>
<tr>
<td>Other</td>
<td>20</td>
<td>5.42</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>200</td>
<td>54.20</td>
</tr>
<tr>
<td>Male</td>
<td>169</td>
<td>45.80</td>
</tr>
<tr>
<td>English language learners (ELL)</td>
<td>18</td>
<td>4.88</td>
</tr>
<tr>
<td>Students with Individualized Education Plans (IEPs)</td>
<td>37</td>
<td>10.03</td>
</tr>
</tbody>
</table>

Table 3

*Summary Statistics for Student-Level and Classroom-Level Variables*

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>St. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Student-level variables (N = 471)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Essay Scores (Outcome)</td>
<td>2.40</td>
<td>0.84</td>
<td>1.00</td>
<td>4.00</td>
</tr>
<tr>
<td>Academic Language (AL)</td>
<td>-0.04</td>
<td>1.00</td>
<td>-2.40</td>
<td>4.30</td>
</tr>
<tr>
<td><strong>Classroom-level variables (N = 42)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discussion Features (per minute)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maximal Teacher Dialogic Discussion (MaxTD)</td>
<td>0.45</td>
<td>0.33</td>
<td>0.00</td>
<td>1.20</td>
</tr>
<tr>
<td>Maximal Student Dialogic Discussion (MaxSD)</td>
<td>0.37</td>
<td>0.55</td>
<td>0.00</td>
<td>2.20</td>
</tr>
<tr>
<td>Minimal Teacher Dialogic Discussion (MinTD)</td>
<td>1.00</td>
<td>0.66</td>
<td>0.08</td>
<td>2.70</td>
</tr>
<tr>
<td>Minimal Student Dialogic Discussion (MinSD)</td>
<td>1.30</td>
<td>0.76</td>
<td>0.04</td>
<td>3.30</td>
</tr>
<tr>
<td>Overall Minimal Dialogic Discussion (MinD)</td>
<td>2.30</td>
<td>1.40</td>
<td>0.19</td>
<td>5.90</td>
</tr>
<tr>
<td>Other classroom variables</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participation Rate (PAR)</td>
<td>3.10</td>
<td>0.91</td>
<td>1.00</td>
<td>4.00</td>
</tr>
<tr>
<td>Discussion Time (in minutes)</td>
<td>34</td>
<td>11</td>
<td>14</td>
<td>76</td>
</tr>
<tr>
<td>Variable</td>
<td>Model 1</td>
<td>Model 2</td>
<td>Model 3</td>
<td></td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td></td>
</tr>
<tr>
<td><strong>Dependent Variable: Essays</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Fixed Effects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>2.03 ***</td>
<td>1.95 **</td>
<td>1.84 ***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.26)</td>
<td>(0.44)</td>
<td>(0.26)</td>
<td></td>
</tr>
<tr>
<td>Maximal Teacher Dialogic Discussion (MaxTD)</td>
<td>0.42 *</td>
<td>0.53 **</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.17)</td>
<td>(0.17)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maximal Student Dialogic Discussion (MaxSD)</td>
<td>-0.16</td>
<td>-0.47 **</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.15)</td>
<td>(0.16)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minimal Dialogic Discussion (MinD)</td>
<td>-0.13 *</td>
<td>-0.11 *</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td>(0.05)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participation Rate (PAR)</td>
<td>0.14</td>
<td>0.23 *</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.09)</td>
<td>(0.09)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Academic Language (AL)</td>
<td>0.17 **</td>
<td>0.18 **</td>
<td>0.18 ***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
<td>(0.04)</td>
<td>(0.04)</td>
<td></td>
</tr>
<tr>
<td>Socioeconomic Status (SES)</td>
<td>0.00</td>
<td>-0.01</td>
<td>-0.01</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.09)</td>
<td>(0.09)</td>
<td>(0.09)</td>
<td></td>
</tr>
<tr>
<td>English Learner Status (ELL)</td>
<td>-0.18</td>
<td>-0.17</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.14)</td>
<td>(0.14)</td>
<td>(0.14)</td>
<td></td>
</tr>
<tr>
<td>Special Education Status (SPED)</td>
<td>-0.08</td>
<td>-0.09</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.11)</td>
<td>(0.11)</td>
<td>(0.11)</td>
<td></td>
</tr>
<tr>
<td>Race: White</td>
<td>-0.07</td>
<td>-0.07</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.08)</td>
<td>(0.08)</td>
<td>(0.08)</td>
<td></td>
</tr>
<tr>
<td>Race: Latino</td>
<td>0.03</td>
<td>0.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.09)</td>
<td>(0.09)</td>
<td>(0.09)</td>
<td></td>
</tr>
<tr>
<td>Race: Other</td>
<td>-0.06</td>
<td>-0.06</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.15)</td>
<td>(0.15)</td>
<td>(0.15)</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>0.09</td>
<td>0.09</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.06)</td>
<td>(0.06)</td>
<td>(0.06)</td>
<td></td>
</tr>
<tr>
<td><strong>Random Effects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>471</td>
<td>471</td>
<td>471</td>
<td></td>
</tr>
<tr>
<td>Groups:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Students</td>
<td>369</td>
<td>369</td>
<td>369</td>
<td></td>
</tr>
<tr>
<td>Topics</td>
<td>25</td>
<td>25</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>Teachers</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>Students (Intercept) Std.Dev.</td>
<td>0.23</td>
<td>0.23</td>
<td>0.23</td>
<td></td>
</tr>
<tr>
<td>Topic (Intercept) Std.Dev.</td>
<td>0.27</td>
<td>0.33</td>
<td>1.41</td>
<td></td>
</tr>
<tr>
<td>Participation (Slope) Std.Dev.</td>
<td>0.20</td>
<td>0.25</td>
<td>0.48</td>
<td></td>
</tr>
<tr>
<td>Teachers (Intercept) Std.Dev.</td>
<td>0.27</td>
<td>0.24</td>
<td>0.15</td>
<td></td>
</tr>
<tr>
<td>Residual Std.Dev.</td>
<td>0.55</td>
<td>0.55</td>
<td>0.54</td>
<td></td>
</tr>
<tr>
<td>AIC</td>
<td>977.3</td>
<td>975.4</td>
<td>974.6</td>
<td></td>
</tr>
<tr>
<td>DIC</td>
<td>866</td>
<td>865</td>
<td>857</td>
<td></td>
</tr>
<tr>
<td>Deviance</td>
<td>903.7</td>
<td>902.3</td>
<td>896.9</td>
<td></td>
</tr>
</tbody>
</table>

*p<.05 **p<.01 ***p<.001
Figure 1. Student and teacher discussion talk moves captured by the Low Inference Discourse Observation Tool (LIDO).
Figure 2. Maximal and minimal teacher- and student-produced dialogic scores as captured by the LIDO.
Figure 3. Essay scores ranged from 1 through 4, with over half of the students in the sample attaining a score of 2, followed by 3, 4, and 1 out of 4.
Figure 4. Correlation matrix of essay scores, predictors (participation rate, maximal teacher and student dialogic discussion, minimal teacher and student dialogic discussion), and controls (pre-discussion academic language scores, special education status, ELL status, SES, and gender).
Figure 5. The predicted difference in essay scores given observed MaxTD and MinD scores (per 10 minutes of discussion time) in four classrooms.
Chapter 3: Study 2

Exploring the relationship between individual student talk and writing in an eighth grade classroom

Introduction and Literature Review

Navigating society without having achieved fundamental skills such as proficiency in reading and writing is a challenging endeavor for any person living in today’s literacy-rich environment. Yet, students in the United States continue to struggle with achieving proficiency in reading and writing. Data from the 2015 National Assessment of Educational Progress (NAEP) showed that only 36% of fourth graders and 34% of eighth grade students in the United States scored at or above proficiency levels in reading (U.S. Department of Education, 2015). Unlike reading and mathematics, fourth grade writing is not assessed regularly by NAEP; however, eighth and twelfth grade data from 2011 show that only one quarter of students about to enter high school are proficient writers, and at graduation, still only one quarter are proficient writers (U.S. Department of Education, 2011).

When students’ writing was assessed in 2007, their scores varied widely from state to state, with only 15% of eighth graders in Mississippi scoring at or above proficiency levels, compared to 56% of eighth graders in New Jersey (U.S. Department of Education, 2007). These statistics are alarming, as they reveal that between 44% and 85% of eighth grade students are entering high school without having reached proficiency levels (let alone mastery) in academic writing.
Furthermore, as enrollment of diverse learners such as emergent bilinguals continues to increase in today’s public schools (U.S. Department of Education, 2015), teachers face the challenging task of ensuring both native and non-native speakers of English achieve at least basic proficiency levels in reading and writing. Additionally, if students are exiting middle school without mastering skills necessary for basic academic work in high school, they face the risk of continuing to struggle in meeting the more demanding academic skills expected of them in high school and beyond.

**Higher Grades, Different Challenges**

In elementary school, students are predominantly exposed to narrative texts and are expected to master narrative writing; this changes in middle and high school, where there is a perceptible shift in expectations, as students now encounter expository and informative texts and are expected to gain proficiency in the persuasive writing genre (National Governors Association, 2010; Hillocks, 2002). In argumentative or persuasive writing, authors draw on evidence and reasoning to convince their audience of a particular position (Salahu-Din, Persky & Miller, 2008). This genre requires writers to take a position, support it with arguments to which their audience can relate, support these arguments with evidence and reasons while also considering counterarguments and appropriate

10 Also referred to as English language learners (ELLs)
rebuttals, all with the goal of persuading the reader using logical arguments and conclusions (Toulmin, 2003; Hillocks, 2002).

The National Assessment of Educational Progress tests of fourth, eighth, and twelfth grade students’ narrative, informative, and persuasive writing reflect the increased demand on persuasive writing over the course of schooling. For example, while 25% of fourth grade prompts were persuasive in nature (i.e. they required students to persuade friends/an editor, refute arguments or take a position in a debate), 30% of prompts for eighth grade students and 40% of prompts for twelfth graders were persuasive in nature (U.S. Department of Education, 1998).

Increasingly, educational standards in the United States have placed more emphasis on students’ persuasive writing as an academic outcome (National Governors, 2010). For example, the Common Core State Standards require eighth grade students, across subject areas, to support their written claims with “clear reasons and relevant evidence” (National Governors, 2010). Additionally, students are expected to not only present claims, but also “distinguish the claim(s) from alternate or opposing claims;” that is, students’ written persuasive texts are expected to include clearly stated claims that are supported with logical evidence and reasoning, as well as the articulation of counter-claims (National Governors, 2010). Students’ skills in mastering the persuasive writing genre, however, do not reflect these shifting expectations.
Writing vs. Reading?

Writing is not only a cognitive task, but it is also a linguistic and social task, as writers are expected to communicate their thoughts and ideas to an audience (Gee, 2007). Furthermore, writing can reveal what students have learned; that is, it makes comprehension visible as it provides a window into students’ reading process and comprehension (Serravallo, 2012). Yet, despite its importance and the increasing demands on students’ writing proficiency, writing research continues to take a back seat to reading research. Compared to reading research, writing research is underfunded, descriptive rather than experimental, and mostly conducted in post-secondary settings (Juzwik et al., 2006). Also, writing is sometimes conceived as a process different from reading, a view that can limit educators’ understanding of the role of writing in promoting reading and comprehension outcomes.

In a meta-analysis of writing instruction and its impact on reading, Graham and Hebert (2011) found that a) writing instruction improves reading skills; b) writing about materials that were previously read enhances reading comprehension; and c) increasing time allotted for writing improves reading comprehension. The meta-analysis also found that high quality writing instruction can improve students’ reading fluency and word-solving skills (Graham and Hebert, 2011).
Additionally, Duke, Pearson, Strachan, and Billman (2011) included *engaging students in discussion* and *integrating reading and writing* in their ten essential elements of effective reading comprehension. Thus, a successful literacy program recognizes that writing and reading are mutually dependent and should therefore be tackled together to improve students’ literacy outcomes. One such program that addresses these goals is the Word Generation Program, which not only incorporates reading and writing throughout the curriculum, but also engages students in meaningful discussion (Snow, Lawrence, & White, 2009; Lawrence, Crosson, Paré-Blagoev, & Snow, 2015). Such classroom discussions, however, and their relationship to various student outcomes, need to be better understood.

**Classroom Discussion**

From Socrates in ancient times to modern-day educators, people have postulated that dialogue is key to building critical thinking. Technology has enabled today’s learning environments to move beyond a traditional setting and shift toward online learning, either through massive open online courses (MOOCs) such as EdX and Coursera or through online platforms such as Canvas that allow for greater or exclusive course interaction online. Even within grades k-12, students can now access educational resources and lessons digitally that were previously only offered in traditional classroom settings (e.g. Khan Academy). Furthermore, teachers now have access to online tools that enable them to conduct various portions of classroom activities online, outside the confines of a traditional classroom.
When weighing the differences between online learning environments and traditional classrooms, discussion surfaces as a component that is manifested very differently, depending on whether it is conducted online or in person. For example, discussions occurring online may not occur in real-time (that is, students have the ability to type their response to a prompt, to which other students, at a later time, can read and respond). Even if students engage in real-time talk via online chats, these differ significantly from in-person interactions as they require students to first formulate their thoughts into written responses. By contrast, in-person discussions can take many forms, and capture students’ real-time reactions to statements made by others. They occur in person, face-to-face, and often with the facilitation of a teacher. They can take place between student pairs, among small groups of students with or without teacher input, or at the whole class level; regardless of their form, however, discussions may become the most distinguishing aspect of traditional classrooms.

Not only do classrooms provide an opportunity for students to discuss ideas face-to-face, they exemplify one of the few remaining places where controversial or moral questions can be discussed (Hess, 2009). Additionally, with the facilitation of a teacher using dialogic approaches, public school classrooms provide students opportunities to hone their argumentation skills while discussing controversial topics in civil and democratic ways (Juzwick, Borsheim-Black, Caughlan & Heintz, 2013). Students can also benefit from exposure to perspectives and experiences vastly different from their own, in a way that is
difficult to do outside of the classroom environment. Thus, there are many reasons to have classroom discussions, but what exactly counts as a discussion?

Scholars have distinguished between recitation-like discussions (Mehan, 1979), still a defining feature of western education, and discussions that engage students in ways that help them develop their thinking (i.e. academically productive talk, or dialogic discussions) (Michaels, O’Connor, & Resnick, 2008). Years of research in this field have shown that these forms of discussion help develop students’ vocabulary (Stahl and Clark, 1987), improve text comprehension (Murphy, Wilkinson, Soter, Hennessey, & Alexander, 2009), and improve students’ development of deep reasoning and critical thinking skills (Michaels, O’Connor, & Resnick, 2008).

Additionally, a series of studies by Deanna Kuhn and colleagues has shown that dialogic communication helped improve students’ argumentative reasoning skills over time (Kuhn & Crowell, 2011), was positively related to students’ comprehension of argument (Kuhn, Zillmer, Crowell, & Zavala, 2013), improved students’ argumentative discourse (Crowell & Kuhn, 2014), and improved students’ writing quality (Kuhn, Hemberger, and Khait, 2016). Kuhn and colleagues had students discuss topics in pairs, followed by small group discussions electronically, and ending with a whole class discussion. Collectively, these studies show that much is to be gained from dialogic discourse, not only for students’ oral language skills, but also for their critical thinking and outcomes such as reading comprehension and writing. The question that remains, however,
is whether whole class classroom discussions improve students’ writing quality, and if so, which elements of discourse produced by students help (or hinder) their post-discussion writing.

**Student Participation**

Another important classroom feature that may be related to post-discussion outcomes is student participation. In Study 1, participation was measured using in-class observers’ indication of total percentage of students who orally participated during a given discussion. The relationship between participation and writing was found to be statistically significant, in that more participation at the classroom level was associated with higher essay scores at the classroom level, regardless of topic. However, it is worth considering whether active participation is the only form of participation that can yield positive outcomes.

To that end, though her study did not include student outcomes, Clark (2015) questions whether a student’s mere presence during rich discussions, even without active participation, may be sufficient to produce higher academic outcomes. A study by Stahl and Clark (1987) sought to answer this question by assigning students to three conditions: students who contributed to whole class discussions after being told they would (Group A), students who were told they would be called on but were not (Group B), and students who were listeners rather than contributors (Group C). Immediate post-discussion vocabulary scores showed that the listener group (Group C) did as well as students in Groups A and B, but not as well as those groups on the delayed science vocabulary posttest. Though
their study examined the effects of discussion on vocabulary learning, it highlights the importance of distinguishing between listeners and active participants in discussions.

Recent work by O’Connor, Michaels, Chapin, and Harbaugh (2017) examined sixth grade students’ participation in two classroom conditions: one characterized by academically productive talk (or dialogic discussion), and another “direct instruction” condition with no discussion. They found that although there was a significant relationship between students’ scores and condition, with higher scores associated with the dialogic discussion condition, there was no relationship between vocal participation and students’ scores. Another, large-scale study by Kelly (2008) found that even middle school students who did not actively participate in classroom discussions benefited from being in classrooms that implemented dialogic discourse. Collectively, these studies highlight the need to consider the effects of participation level on student outcomes, whether that is displayed through active oral participation in whole class settings or otherwise.

Study Aims and Research Questions

This study draws upon Word Generation data to understand the role of student discussion in students’ persuasive writing development. Expanding on findings from the previous study (Study 1), I sought to understand whether individual students’ contributions during class discussion can predict their post-discussion writing scores. Specifically, the present study focuses on the following research questions by examining student-level scores only:
After accounting for students’ prior standardized English Language Arts test scores, the total number of words per essay, gender, and socio-economic status:

1. Does the type of student discussion (highly dialogic vs. less dialogic) predict their post-discussion persuasive writing scores?
2. Does the relative frequency of student contributions during discussions predict their post-discussion persuasive essay scores?

**Method**

**Research Setting**

Data analyzed in this study are drawn from the Catalyzing Comprehension through Discussion and Debate (CCDD) intervention study, which involved improving students’ literacy outcomes through the Word Generation Program (Snow, Lawrence, & White, 2009). Word Generation is a supplementary curriculum that aims to enhance fourth through eighth grade students’ academic language, complex reasoning, perspective taking, and other skills that together produce deep reading comprehension, through activities structured around the discussion of social dilemmas. This curriculum integrates reading, writing, and discussion, as students read materials that present the issue and various positions on the topic and engage them in writing activities that help them articulate and strengthen their positions. Students then engage in discussion and debate on the topic, and end the unit with a persuasive essay where they are asked to state and
defend their position using evidence and reasoning (Snow, Lawrence, & White, 2009).

Data from this study were collected in the final wave of the study, between January and June 2014. The focus of this study is one eighth grade classroom, where students discussed four Social Studies topics over a period of eight classroom sessions. As seen in Table 1, students discussed four civics-oriented topics. These controversial topics ranged from whether the NFL should require the Washington Redskins to change their name due to objections on the grounds that the name is an ethnic slur against Native Americans, to whether the United States has an obligation to address its history of mistreatment against groups like African Americans through mechanisms such as compensation. At the end of each unit, students answered the writing prompts shown in Table 1. All student essays were first drafts.

***Insert Table 1 Here***

Data and Participants

A total of eight video-taped classroom discussions of four topics and their corresponding essays (n=60) were analyzed. The first topic, Should the NFL require the Washington Redskins to change their name? was discussed for a total of 1 hour and 55 minutes\(^1\) over two class periods, 51 minutes of which were coded

\(^{11}\) Throughout the study, time is rounded to the nearest minute.
using the Low Inference Discourse Observation tool (LIDO). The second topic, *When is a crime not a crime?*, was discussed for 3 hours and 15 minutes over three class sessions, yielding a total codable time of 140 minutes. The topic *Where is the justice in our justice system?* was discussed over one class session for a total of 1 hour and 21 minutes, with 52 minutes of codable time. Finally, *How do we right the wrongs of the past?* was discussed over two class sessions that were a total of 59 minutes, 41 minutes of which were coded using the LIDO (Table 1).

The public school from which these data were drawn is located in a semi-urban area of the Northeastern United States, and serves students in grades eight through twelve. At the time of data collection, 77.9% of the school’s population were White, and the remaining students were either Hispanic/Latino (8%), Black/African American (5.5%), or belonging to other ethnic groups (8.6%). These demographics are different from the state’s data, where there was a higher percentage of Latino (17%) and Black/African American students (8.7%), as well as a lower percentage of White students (64.9%) enrolled in that academic year. The student-to-teacher ratio in this school (11.9 to 1) was also lower than the state’s average of 13.2 students to one teacher (Massachusetts Department of Elementary and Secondary Education, 2014).

There were fifteen students in this classroom: seven females and eight males. Similar to the demographics reflected in their school, the majority of

---

12 As the LIDO focuses on whole class discussion only, other activities such as turn and talks or small group work were not coded. Also, text reading was not coded using the LIDO.
students (80%) were White (n=12), while three students were non-White (one student was Latino, one student was Black/African American, and one student was identified as Native American/Pacific Islander/Hawaiian.) None of the students in this sample were identified as English language learners, and one student had an individualized educational plan. Only three out of the fifteen students (20%) were qualified to receive free or reduced lunch, indicating that most students in this sample did not come from low-income backgrounds.

**Measures**

**Holistic essay score (outcome #1).** The outcome variable of interest, students’ post discussion persuasive essay scores, was scored in two ways. The first is a holistic measure which accounts for students’ presentation of a claim or a series of claims, evidence and reasoning they use to support their claims(s), their inclusion of counter-claims or different positions, and a stated conclusion. Essays were scored from 1 through 4 (there were no essays that were scored 0, as all students completed their essays and, at the very least, provided one claim and one piece of supporting evidence.) This scoring method was used in the previous study, where a high reliability between two scorers was achieved using 20% of essays, yielding a Cohen’s kappa of 0.96.

**Argumentation features (outcome #2).** Unlike the previous study, these essays were also scored using the coding scheme developed by Deanna Kuhn and colleagues (Kuhn & Crowell, 2011; Kuhn et al., 2013). First, students’ essays were divided into idea units, with each idea unit representing “an assertion with
any accompanying justification” (Kuhn et al., 2013, p. 464.) After segmenting students’ essays, each idea unit was identified as belonging to one of four categories.

*No-argument* idea units contained only claims, without any accompanying reasons or evidence to support that claim. For example, “I believe the Washington Redskins shouldn’t change their name” constitutes a *no-argument* idea unit, as it contains a claim that is not supported by evidence or reasoning. *Own-side arguments* contained positive attributes of one’s own position. For example, one student asserted that, “As Bob Costas says, the term is used to honor Native Americans.” This constitutes an *own-side argument* as the student supports his/her claim by using textual evidence that highlights a positive attribute of the student’s favored position (in this case, that the NFL should not require the Washington Redskins to change their name.) *Dual-perspective arguments* are ones that support the author’s position by highlighting negative attributes of the opposing position. To illustrate, the following idea unit is classified as a *dual-perspective* argument as it highlights a negative attribute of the opposing passion: “It would cost too much money to do. So no, the NFL should not require the team to change their name.”

The fourth and last type of idea unit is an *integrative perspective*, which is when the author either acknowledges a negative attribute of his/her own position, or a positive attribute of the opposing position. For example, the statement, “I know that some people think it’s a derogatory word, but the majority of people don’t
think this anymore” is considered *integrative* because it acknowledges a positive attribute of the opposing position.

The total of each of the four argument types (*no argument, own-side arguments, dual-perspective arguments, and integrative-perspective arguments*) was divided by the number of *idea units* in each essay. Thus, the proportion of each argument type relative to the total idea units in each essay is used in the analysis.

**Classroom discussion (predictors).** All video tapes were coded for student discussion using the Low Inference Discourse Observation Tool (LIDO) (O’Connor & LaRusso, 2014; Elizabeth & O’Connor, 2014). Developed through an iterative process based on research on accountable talk (Michaels, O’Connor, & Resnick, 2008) as well as measures of classroom talk (Mercer, Wegerif, & Dawes, 1999; Nystrand, Wu, Gamoran, Zeiser, & Long, 2003; Michaels & O’Connor, 2007), the LIDO captures both teacher and student discussion features during whole class discussion. In this study, however, only students’ input was coded.

Unlike the previous study, where student discourse was measured at the classroom level by aggregating students’ overall responses during whole class discussion, this paper examines each student’s talk individually. This was made possible by working with video, rather than audio data, which allowed for the identification of each student and his/her contributions during discussions.

I began this process by making a classroom map of each video (some students moved during the course of the eight class sessions). Relying mostly on
the teacher’s use of students’ names when calling them to respond, I learned all students’ names and verified them using the study’s code key. Listening to each individual carefully, I coded each instance of student talk using the Low Inference Discourse Observation tool (described below). After scoring all students’ essays (which were de-identified) using the two coding schemes described above, I matched students’ IDs with their first names such that all discussion scores aligned with students’ writing scores. Students’ names were then removed from the dataset.

Students’ discussion inputs, as outlined by the LIDO, range from features that are more dialogic, to those that are less likely to prompt dialogic discussion. Unlike monologic talk, dialogic talk allows students to internalize thinking processes and linguistic terms used during discussion (Bakhtin & Holquist, 1981; Vygotsky, 1978), through discussion between students and their teacher as well as among students themselves.

The LIDO captures five types of student talk; namely, 1) students’ direct talk; 2) students referring to each other’s ideas; 3) students presenting and supporting claims with evidence or reasoning; 4) students engaging in turns that are more than a simple clause but do not fit into the previous three categories of talk; and 5) students engaging in responses that are a single clause or less.

In this study, I divided student talk into three categories: Student-to-Student (S2S) refers to direct and indirect student talk (the first two types of student talk). Claim with Evidence (CE) refers to students presenting and supporting their claims
(the third category), and *Minimal Talk* (MIN) refers to student talk that is either one clause or longer, but does not belong to the previous categories. The first two types of talk, S2S, are hypothesized to be most indicative of dialogic discussion, whereas the last two (MIN) are hypothesized to be less likely to prompt truly dialogic interactions.

**Student to student talk (S2S).** The following exchanges between two students discussing the first topic, *Should the NFL require the Washington Redskins to change their name*, are coded as S2S:

Student 1: I’ll respond to the “society has moved forward part.” Society has moved forward to get rid of Oscar the Grouch from Sesame Street just ‘cause it’s moving forward doesn’t mean it’s not moving forward too much. It’s not meant as an insult so we’re not trying to insult you.

Student 2: That’s not good enough for me. I don’t want to just hear about Oscar the Grouch. The term was also introduced in the 1800s to offend Native Americans.

Student 1: So was Yankees. Yankees is a good team and everyone loves the name now.

Student 1: But what if someone is still offended by the name? What if they don’t take it as honor they take it as an insult from others?

This exchange between these two students lasts for a total of seven turns, with no interruption by the teacher or other students. Though these two students were exchanging ideas directly with one another, the prevalence of S2S is relatively low across discussions in this sample compared to Minimal Student Talk (MIN).

**Claim with evidence (CE).** Student turns are coded as Claim with Evidence when they state a claim and support their own claim with evidence or reasoning.
For example, in the same discussion about the Washington Redskins, the teacher asks, “Who thinks it’s implicit?” in reference to a stereotype. One student first agrees that the example provided by the teacher is an example of an implicit stereotype, then supports that position by saying, “because they’re not saying it to anyone – they’re just thinking it to themselves.” Later in the discussion, the teacher asks, “Are you biased?”, and a student responds by saying, “No – no, you’re not biased because you’re supporting the team. It’s the same as a Jets fan walking around Foxborough and at a Jets game and you’re wearing a Patriots shirt and you get offended because you like the Patriots.” In both cases, students support their own claims by drawing on personal experience or knowledge. Unlike Study 1, where CE was part of the Maximal Student Dialogic Discussion measure, it was isolated in this study in order to better understand its unique contribution to students’ writing.

Though S2S and CE are more indicative of dialogic interactions, together, they were found to negatively predict students’ essays in Study 1. I postulated that this finding was due to the imprecise coding of these utterances, in that coding did not account for the difference between classrooms where a high maximal student dialogic score was a result of interactions among multiple students, or those between just two students. This finding motivated the present study, as I sought to understand the relationship between student discussion and persuasive writing by examining individual, rather than classroom, student interactions. Thus, despite
the previous finding, both S2S and CE are hypothesized to positively predict student essays.

**Minimal student talk (MIN).** The third, and most frequently occurring type of student turns is Minimal Student Talk, where students provide answers that are either one clause or longer (and that do not belong to the previous, more dialogic categories). For example, student responses such as, “I don’t know,” or “Yes, I agree but it’s not always the case,” are examples of MIN, as coded using the LIDO. Unlike the exchanges in the first two categories (S2S and CE), these exchanges are not as dialogic and resemble the typical format of talk in classrooms, the IRE sequences. IRE exchanges are less likely to be dialogic, as they are often “quiz-like” in that they often require a student to produce a “correct” response. As such, MIN talk is hypothesized to negatively predict essay scores. To investigate the connection between student talk and writing, these three indicators of discussion are used as predictors in this paper.

Each of the three categories of student talk (*Student-to-Student*, *Claim with Evidence*, and *Minimal Talk*) was summed then divided by the total number of minutes in each discussion to produce predictors that reflect students’ input per minute of discussion time. To illustrate, if a student produced seven utterances coded as S2S over the three days of discussion around the topic of *When is a crime*

---

13 Teacher-initiated questions that elicit a student response and end with a teacher evaluation are referred to as IRE sequences (Mehan, 1979).
not a crime?, that score was divided by 140 minutes to produce an S2S per minute score of 0.05.

As there were students who missed one class each (n=4), these students’ scores were calculated based on the total discussion time in classes for which they were present. For example, one student attended all classes except the last session of the Crime topic, which had a total of 20 minutes of (LIDO-codable) discussion time. That student’s S2S score was divided by 120 (instead of 140) to reflect her proportion of Student-to-Student interactions.

In this paper, there is a low number of participants (15 students) and a relatively high number of outcomes (60 essays produced over four topics). To address this issue, each of the 15 students was treated as independent across four topics. However, given that independence cannot be assumed from one topic to the next, each student’s discussion score was normalized with respect to that student’s own average over time in order to minimize student effects. To illustrate, each student’s S2S per minute score was divided by the mean S2S score over all eight time points for that student, yielding a score that is reflective of student-to-student utterances per minute given the individual’s mean S2S over time. This is similar to a z-score, except that the value produced here is in reference to an individual student’s mean over time, rather than all students’ mean scores over time. This process of group mean centering tests whether students produced better essays after discussions in which they themselves had produced more dialogic talk than
was their norm. This process of comparing students to themselves resulted in 60 discussion features as predictors (15 per topic), and 60 essay scores as outcomes.

**Classroom participation (predictor).** To answer the second research question, the proportion of each students’ talk relative to his/her peers was used as a predictor. Findings from Study 1 suggest that classroom participation rate (measured as the total percentage of class participation on a given day), positively predicted students’ essay scores, even when controlling for the effects of topic on participation rate. However, unlike the previous study where participation rate was calculated using classroom observers’ estimate of the percentage of students who participated during each discussion, this variable (ptalk) was calculated more precisely by first counting the total number of codable student utterances, then dividing each student’s total talk moves by the total utterances produced during whole class talk. I hypothesize that proportion of talk will positively predict essay scores in this study.

**Standardized test (control).** In order to account for students’ academic performance, students’ Massachusetts Comprehensive Assessment System (MCAS) scores in English Language Arts from the previous year’s spring semester were used as a control variable.

**Total words written (control).** The total number of words written per essay (TW) is used as a control variable in this sample. This measure, though simplistic, has been used in curriculum-based measures of writing and has been
shown to correlate with students’ scores on written standardized assessments (e.g. Deno, Marston & Mirkin, 1982). All words written by students were counted. Misspelled words were corrected, but additional words were not added in cases where students omitted a word.

**Demographic variables (controls).** Other controls used in this study included students’ gender (FEM) and socioeconomic status (SES) as measured by being eligible for free or reduced school lunches.

**Fitted Models**

To address whether different forms of student input during whole class discussion predict their holistic essay scores (RQ1), four models are fitted (Table 3). Given the structure of the data, student effects are assumed to be correlated with the outcomes; thus, all models fitted in this study are student-level fixed effects multilevel models. The first three models test each of the three predictors separately, while the fourth and final model (Model 4), tests the relationship between all three discussion predictors and students’ writing:

**Level-1 model (discussion-level):**

\[
E\text{Score}_{ij} = \beta_{0j} + \beta_1S2S_{ij} + \beta_2CE_{ij} + \beta_3MIN_{ij} + \epsilon_{ij}
\]

\[
\epsilon_{ij} \sim N(0, \sigma^2_{\epsilon})
\]

**Level-2 model (student-level):**

\[
\beta_{0j} = \gamma_0 + \gamma_{01}TW_j + \gamma_{02}MCAS_j + \gamma_{03}SES_j + \gamma_{04}FEM_j + u_{0j}
\]

\[
u_{0j} \sim N(0, \sigma^2_{u})
\]

**Where:**

- *E*\text{Score} is the predicted essay score given discussions \(i\) by individual students \(j\) and, like the previous study, were scored holistically from 1-4
- \(\beta_{0j}\) denotes a random intercept for each student and is further described by the level-2 model
\[ \beta_1 - \beta_3 \] are discussion-level predictors given student’s mean scores over time (60 observations):
- Student-to-Student talk per minute (S2S); Claim with Evidence per minute (CE); Minimal Student talk per minute (MIN)
- \( \gamma_{00} \) is the overall intercept across all groups
- \( \gamma_{01} - \gamma_{04} \) are the student-level controls (15 observations):
  - \( TW \) is total words written per essay
  - \( MCAS \) is students’ prior standardized test scores
  - \( SES \) is socioeconomic status, defined as eligibility for free/reduced lunch
  - \( FEM \) is gender, with female as the reference category
- Errors are assumed to be normally distributed at both the discussion and student levels and are denoted by \( \epsilon_{ij} \) and \( u_{0j} \) respectively

Next, a series of student-level fixed effects multilevel models with different outcomes were fitted to address the first research question using the writing outcomes produced by scoring the essays using the coding scheme developed by Kuhn and Crowell (2011) (Table 4). Namely, the following outcomes were examined: total idea units (Table 4 Model 5), proportion of no-arguments (Model 6), proportion of own-side arguments (Model 7), proportion of dual-perspective arguments (Model 8), and proportion of integrative-perspective arguments (Model 10).

To address the second research question, a series of student-level fixed effects models examining whether the proportion of student talk predicted any of the writing outcomes were fitted (holistic essay scores, as well as the writing outcomes calculated using Kuhn and colleagues’ coding scheme: idea units, no arguments, own-side arguments, dual-perspective arguments, and integrative-perspective arguments) (Table 5, Models 10-15).
Results

Summary statistics for all variables are shown in Table 2. Students’ essay scores (scored holistically from one through four) ranged from a minimum of 2 to a maximum of 4, with a mean of 2.6 ($SD = 0.67$), or 65%. As these were first draft essays, the relatively low score is not unexpected. Out of the 60 essays produced by students across four topics, 53.3% received a score of 2 (n=32), 36.7% received 3 out of 4 (n=22), and 10% of essays (n=6) scored a perfect score of 4 (Figure 1.)

As for students’ argument types (scored using Kuhn and colleagues’ coding scheme), students’ total idea units ranged from 3 to 16, with a mean of 6.7 and a standard deviation of 3.0. The proportion of no-arguments (NOARG) ranged from 0 to 1, with a mean of 0.35 and a standard deviation of 0.18. The proportion of the remaining argument types (OWN, DUAL and INT) are all strongly right-skewed (Table 2).

Also, dual-perspective arguments and integrative-perspective arguments were predominantly absent from students’ writing compared to no-argument and own-side arguments. Of the sixty essays, three essays did not contain any no-argument ideas, while only two essays did not contain any own-side arguments. On the other hand, 78.3% of essays (n=47) contained no dual-perspective arguments, and 88.3% of essays (n=53) contained no integrative-perspective arguments.
Student-to-student utterances per minute relative to students’ own average ranged from 0 to 4, with a mean of 0.80 and a standard deviation of 1.10. This predictor was much less frequent than the other two, with 53.3% of individual students’ discussions (n=32) containing no S2S discussion features. Claim with Evidence per minute relative to students’ average ranged from 0 to 4 ($M = 0.93, SD = 1.20$), and was also less frequent than MIN, with 51.7% of individual student discussions (n=31) containing no CE. On the other hand, only 10 student discussions (6.7%) contained no Minimal Talk (MIN), and students’ MIN per minute relative to their own average ranged from 0 to 4 ($M = 1.00, SD = 0.73$.) (Table 2).

The second predictor, the proportion of student talk, ranged from 0 to 0.37, with a mean of 0.07 and a standard deviation of 0.08. That is, students’ proportion of talk ranged from 0% of total talk time (11 students) to 37% of total talk (1 student). This variable was highly right-skewed, with 46 observations (76.7%) falling at 10% of talk or lower, and the remaining 14 observations (23.3%) falling between 11% and 37% of total talk. When the student with the highest proportion of talk ($ptalk$) was removed from the dataset, the range lowered from a maximum of 37% to a maximum of 20%, with 82.1% of observations (n=46) falling between 0 and 10%, while the remaining 17.9% (n=10) fell between 11 and 20%. To

\[\text{Note: each person contributes 4 data points out of 60}\]
address the skewness, ptalk was log transformed; however, it remained right skewed even after the transformation.

Finally, the control variable total words written (TW) ranged from 33 to 310, with a mean of 133 and a standard deviation of 64 total words per essay (Table 2), and students’ prior MCAS scores in English Language Arts ranged from 224 to 264 ($M = 246, SD = 10$) (Table 2).

***Insert Table 2 Here***

Correlations

Next, correlations among predictors and the outcomes were examined (Figure 2).

***Insert Figure 2 Here***

**Correlations between holistic essay scores and main predictors.** As seen in Figure 2, the only predictor that was statistically significantly correlated with holistic essay scores was S2S, where a negative correlation was observed ($r = -0.33, p < .05$). Claim and Evidence per minute, relative to student’s own average, was not significantly correlated with essay scores ($r = -0.24$) and neither were student’s minimal responses per minute relative to their own average ($r = 0.21$). It is worth noting, however, that CE has a negative sign, whereas MIN has a positive sign (Figure 2). So far, these results contradict the hypothesis that higher dialogic discussion features would be positively associated with essay scores, while lower-
dialogic discussion features would be negatively associated with student’s overall essay scores.

**Correlations among argument types (outcome #2) and main predictors.**

As essays were scored two different ways, correlations among the main predictors and argument types are also examined. First, there was a statistically significant, positive correlation between MIN and total number of idea units ($r = .27, p < .05$). No other predictors were associated with ideas, though higher dialogic features (CE and S2S) were also negatively correlated (though not significantly). Similarly, there were no statistically significant correlations among any of the three discussion predictors and the four argument types: proportion of no-arguments, proportion of own-side arguments, proportion of dual-perspective arguments, and proportion of integrated-perspective arguments. This is not surprising, as the majority of essays did not contain any dual-perspective or integrative-perspective arguments, and own-side arguments were highly variable among students (Table 2).

**Correlations among outcomes and controls.** Due to their low frequency (n=15) compared to other observations in this study (n=60), students’ standardized prior Language Arts scores, socioeconomic status, and gender were not included in the correlation matrix. Total words written (TW), was strongly positively correlated with holistic essays scores ($r = .53, p < .001$), but not correlated with
any other argument types. However, there was a strong, positive correlation between total words written and total idea units ($r = .87, p < 001$).

**Correlations among variables.** Other noteworthy correlations among variables were found between idea units and holistic essay scores ($r = .51, p < .001$), which is not surprising given that the same relationship was found between essay scores and total words written, a variable that is highly correlated with idea units. Also, no arguments were significantly negatively correlated with own-side arguments ($r = -.52, p < .001$) and dual-perspective arguments ($r = -.29, p < .05$). Similarly, own-side arguments were statistically significantly correlated with dual-perspective arguments ($r = -.39, p < .01$) and integrative-perspective arguments ($r = -.33, p < .05$).

**Model Results**

**RQ1 (outcome #1).** To answer the first research question about the predictive relationship between discussion and persuasive essay scores (scored holistically), a series of multilevel student level fixed effects models were fitted, starting with the most dialogic discussion feature, student-to-student utterances per minute, given each student’s average S2S score over time (Table 3). In Model 1, S2S significantly predicted essay scores when controlling for total words written (TW), students’ standardized test scores, and their demographic variables.

However, contrary to the hypothesized prediction but reflective of the correlational relationship, S2S negatively predicted essay scores ($\beta = -0.19, p < $
.01). Total words written (divided by 100) was strongly associated with essay scores ($\beta = 0.54, p < .001$).

Model 2 examined the relationship between Claim with Evidence and essay scores (Table 3). This relationship was not statistically significant when controlling for total words written, prior MCAS scores, and students’ demographic variables ($\beta = -0.05, p = .48$). Total words written was once again statistically significantly predictive of essay scores ($\beta = 0.51, p < .001$).

The third model examined the relationship between students’ Minimal Talk (MIN), which was found to not predict essay scores ($\beta = 0.09, p = .41$). Once again, total words written was positively predictive of essay scores ($\beta = 0.51, p < .001$).

The final model (Model 4) combines all three discussion variables; the relationships are the same as the previous models, where only S2S is negatively predictive of essay scores ($\beta = -0.18, p < .05$), and total words written is the only covariate positively predictive of essay scores ($\beta = 0.51, p < .001$). The relationship between student-to-student talk and essay scores did not align with the stated hypothesis, as this predictor was found to negatively predict student essay outcomes when accounting for other discussion features and control variables. This model predicts that on average, each additional occurrence of student-to-
student talk per minute, relative to that student’s average S2S over time, is associated with a decrease of 4.75% in essay scores.\textsuperscript{15}

***Insert Table 3 Here***

**RQ1 (outcome #2).** Next, a series of student-level fixed effects models, each with a different outcome variable, were fitted to examine the relationship among features of the second outcome, argument types, and discussion predictors (Table 4). The first model in Table 4 examines the predictive relationship between total idea units and the student discussion predictors. As shown in Model 5, none of the predictors are statistically significantly predictive of total idea units. Similar to prior model results, total words written (TW) is a statistically significant predictor of total idea units, when accounting for all discussion variables and other controls ($\beta = 3.80, p < .001$). Students’ prior standardized English Language Arts test scores, a control variable, is approaching significance ($\beta = 0.54, p = 0.10$) (Table 4, Model 5).

Next, relationships among discussion predictors and the proportion of no-arguments were examined (Table 4, Model 6). There were no statistically significant associations among any predictors and this outcome, nor among control variables and the outcome (Table 4). Similarly, in Models 7 and 8, where the outcomes were proportion of own-side arguments and dual-perspective arguments, respectively, there were no statistically significant predictors of those outcomes.

\textsuperscript{15} As essays were scored out of 4, the predicted estimate of S2S (-0.19) was divided by 4 then multiplied by 100 to arrive at 4.75%.
Finally, in Model 9, S2S per minute (given a student’s average), is a statistically significant predictor of students’ integrative-perspective arguments, when controlling for other discussion variables, students’ total words written, and demographic variables ($\beta = -0.02, p < .05$). Despite this statistically significant result, Models 7 through 9 (Table 4) should be taken with caution, as they relied on information from very few observations given their low frequency in students’ essays (see Table 2).

***Insert Table 4 Here***

**RQ 2.** The second research question aims to investigate the relationship between essay scores and participation. Here, participation was calculated as the proportion of students’ talk relative to their classmates in each of the four topics (ptalk). This variable remained right-skewed even after a log-transformation was performed; thus, it was used in the analysis without the transformation.

A series of student-level fixed effects multilevel models were fitted, each examining the relationship between proportion of talk and a different writing outcome (holistic essay score, total number of ideas, no-arguments, own-side arguments, dual-perspective arguments, and integrative-perspective arguments). As seen in Models 10 through 15, the proportion of student talk relative to their classmates was not statistically significantly predictive of any of the writing outcomes (Table 5).

***Insert Table 5 Here***
However, when the student with the highest proportion of talk relative to other classmates was removed from the analysis (Student A in Figure 4 below), the predictive association between proportion of talk and holistic essay scores approached significance, and the sign of the coefficient changed from positive (Model 10) to negative ($\beta = -2.64, p = 0.08$). This indicates that the high degree of variability among students’ talk proportions poses a problem for interpreting results associated with proportion of talk (ptalk).

**Qualitative Analysis of Students’ Discussion**

Examining the data longitudinally can help us understand whether students demonstrated any improvement in their writing or discussion skills over time, and whether there were differences by topic. Though on average, students’ S2S scores (per minute of discussion time) increased over time, there is a sharp increase when discussing *Where is the justice in our justice system?*, followed by a decline in the unit, *How do we right the wrongs of the past?* (though it is higher than the topics *When is a crime not a crime?* and *Should the NFL require the Washington Redskins to change their name?*) (Figure 3). Similarly, on average, students start with high minimal talk (MIN) during the NFL discussion, and continue to use fewer MIN utterances, with the sharpest decline occurring during the third unit (*Justice*). Unlike Student-to-Student talk and Minimal Talk, Claim with Evidence remains fairly constant throughout the four units. Furthermore, essay scores were higher, on average, during the second unit (*Crime*), compared to the other topics.

***Insert Figure 3 Here***
These observations suggest that topics may be an influential variable, as certain topics may elicit more interest, and therefore higher engagement, than other topics. In this case, *Where is the justice in our justice system?* was also the one topic which was only discussed for one class session, compared to the other topics that were discussed over two or three class periods. Similarly, *When is a crime not a crime?* was discussed the longest (three class sessions), which could help explain the high essay scores students achieved, on average, after discussing this topic.

Next, individual differences among students’ discussion and essay scores were examined by analyzing talk moves and associated essay scores of the student with the highest proportion of talk (Student A) and a student with relatively low classroom participation rate (Student B). Student A’s utterances during discussion (Figure 4) resemble the mean utterances and essay scores over time/ across topic (Figure 3), in that A’s Minimal Talk per minute is closely associated with essay scores. On the other hand, Claim with Evidence remains fairly low (and consistent) across topics, and Student-to-Student talk shows an upward incline. However, not all students fit this pattern.

***Insert Figure 4 Here***

When compared to Student B (Figure 5), we see that like Student A (and the average scores), there was an incline in essay scores on the second unit, *Crime*. However, unlike Student A, Student B’s Minimal Talk decreased over time, while
his/her other talk moves, Student-to-Student and Claim with Evidence, slightly increased over time.

***Insert Figure 5 Here***

Quality over quantity? As the LIDO is not a measure of student discussion quality, these findings necessitate further investigation into the type of discourse utterances that were assumed to be dialogic in nature. The data presented in Table 6 are from the third unit, Where is the justice in our justice system? where students produced the highest average of S2S occurrences per minute (Figure 3). The transcript in Table 6 includes discussion that is mostly driven by Student C and Student A (who had the highest overall proportion of talk, and whose discussion and essay scores are shown in Figure 4). Students B (Figure 5), D, and E are also engaged in this three-minute discussion that produces a total of 25 student-to-student (S2S) utterances. The discussion begins when Student A asks the teacher if he could respond to Student C’s previous statement.

***Insert Table 6 Here***

In this section, students are discussing the different purposes of punishment (rehabilitation, deterrence, retribution, incapacitation, and restitution.) In small groups, students first presented arguments supporting their favored positions, and were then given time to respond to a different group’s position. After they wrote down their thoughts, Student A asked to respond to another group’s position. Throughout the three minutes, students C, D, and later E, challenge Student A on his contradictory assertion that a “criminal is a criminal” yet “people can change.”
As the discussion progresses, A finally decides to “[throw] in the white flag,” as Student C says, and declares that he is no longer interested in continuing this discussion (he repeats, “I’m done.”) Despite evidence from Student C, D, and E, Student A remains firm in his conviction and does not recognize the contradictory nature of his statement.

The teacher takes on an interesting role in that she at times attempts to clarify Student A’s statements and address his misconceptions, while encouraging him to continue speaking by saying he is “making a good point” just before turning to another student who takes the conversation in another direction.

While this discussion is certainly dialogic in form, in that there is a clear back-and-forth discussion among students around content, it does not seem to help Student A further his understanding through talk, and all students in this case hold on to their positions without attempting to build upon each other’s perspectives. This highlights an important consideration when examining student discussion: as a low-inference instrument, the LIDO captures what appear to be dialogic talk among students in form, but this form of talk is not necessarily conducive to students’ learning in substance. Next, Student A’s post-discussion essay is examined.

***Insert Table 7 Here***

During this unit’s discussion, Student A produced a total of 12 student-to-student utterances, yielding an S2S score of 0.23 per minute of discussion time (Figure 4). Despite having a relatively high S2S score, his essay score did not
contain any counter-arguments, and received a score of 2 out of 4. Also, Student A’s position that all people can change is identical to all other students’ position, and so is the evidence he uses to support his claim. Even Student E, who had a very high S2S per minute score of 0.60, scored 2 out of 4 on this topic’s essay (compared to 3 out of 4 on all other essays). This brings into question whether these dialogic discussions are in fact helping students develop their thinking, or if they are simply a way for students to get their ideas heard.

Furthermore, students’ positions on the first unit (should the NFL require the Washington Redskins to change their name?) were identical, in that they all stated the team should not be required to change its name. Reasons given included that other teams are not required to change names deemed derogatory, the team and its players do not intend to offend Native Americans, and Native Americans should be “honored” by the name as the team is a successful team. Not a single student took a different position, raising questions about whether the classroom discussion was nothing more than an echo chamber.

Discussion

The findings presented in this paper and in Study 1 suggest that indicators of high student dialogic discussion (e.g. direct or indirect student to student interaction) are negatively associated with their post-discussion essay scores. To understand this finding, we must consider what the LIDO is intended to measure, and what it does not reflect. Specifically, while the LIDO captures forms of classroom discourse, it is not a measure of discussion quality. When a student
asserts a claim and provides evidence to support that claim (CE), the type of evidence or the quality of the evidence used to support one’s own argument is not assessed by the LIDO. Additionally, when students are engaged in direct or indirect discourse around content (S2S), those interactions are not coded for quality. Instead of discussion as a tool to improve perspective taking (and thereby strengthen one’s persuasive thinking and writing), discourse coded as S2S in this classroom was at times parallel, rather than integrative, in nature.

The discussion transcript provided in Table 6 shows that while the discussion was in fact dialogic in that it “connotes social relationships of equal status, intellectual openness, and possibilities for critique and creative thought” (O’Connor & Michaels, 2007, p.277), it did not seem to promote learning and thinking, another aspect of dialogic discussions (Michaels, O’Connor, & Resnick, 2008; Tharp & Gallimore, 1988). The discussion did not prompt Student A to reflect or build upon other students’ ideas, thereby gaining new perspectives and enhancing his learning and understanding of the topic. Other students too, were firm in their stances, and did not attempt to convey their ideas in a different way, prompting Student A to withdraw from the discussion all together.

Moreover, the LIDO captures student discourse that occurs in a whole class setting. Also, though the LIDO was not intended for use in small group interactions, even if exploring small group discussions was of interest, it was not possible due to the set-up of recording equipment in the classroom. Thus, whereas Kuhn and colleagues constructed and monitored small group discussions over
time, both during classroom talk and online (Kuhn & Crowell, 2011; Kuhn, Hemberger, & Khait, 2016), that was not possible in this study. Small group discussions allow students who participate very little or do not participate at all during whole class discussions to share their perspectives with peers. Thus, to fully understand the role of student discussion in promoting outcomes such as writing quality and reading comprehension, all student discussion, including those that occur in small group, are important to consider.

The finding that proportion of talk was not predictive of writing outcomes highlights the need to differentiate between participation and engagement, the latter of which was not captured in this study. For example, it is possible that students who participate infrequently (or do not participate at all) use listening skills that help develop their thinking (and thus their post-discussion writing). Conversely, it may be that frequent talkers do not incorporate other perspectives when writing, but use classroom talk to generate ideas. Student A (Figure 4) exemplifies the finding that Minimal Talk (the least dialogic feature) is positively predictive of total number of idea units in essays (Model 5).

An additional factor that may play a role in determining how classroom talk is used and whether it is predictive of student outcome is grade level. While most participants in Study 1 were in either fourth or fifth grade, the sample examined in this study was with an older group, eighth graders. The difference in age and maturity could play a role in how classroom discussions are carried out, including whether they are more or less dialogic, on average. Also, certain elements of
classroom talk captured by the LIDO could predict writing outcomes differently for students in different grades. Future investigations that account for grade-level differences could help further inform our understanding of the relationship between dialogic talk and writing.

Finally, the structure of discussions led by the teacher may have an impact on the forms of student talk. In this study, though all four topics were led by the same teacher, her approach to discussion differed from one unit to another. In some instances (e.g. *When is a crime not a crime?*), the teacher allowed students to work in small groups, electing a student representative to share the group’s positions. While there was an opportunity for asking questions, students did not engage in an open debate. By contrast, the third discussion, *Where is the justice in our justice system?*, students were given approximately 25 minutes of class time to work in small groups in preparation for their debates. After hearing other groups’ positions, students were asked to write down a counter argument that was to then be presented by a different group member. Discussion structure, coupled with time allotted for each discussion, should be examined separately in order to assess whether certain forms of discussion produce better outcomes than others.

**Limitations**

Unlike the first empirical study where audiotapes were used to analyze students’ discussion, the use of video recordings of classroom discussions allowed for the identification of specific students’ discussion moves, which were then linked to the persuasive essays students produced after each unit. However, there
are several limitations that should be considered when reflecting upon implications
drawn from this study, as well as future studies that emerge from this work.

Although the data presented here were longitudinal in nature, the small
number of participants (one classroom with 15 students) meant that statistical
techniques such as growth level modeling could not be performed. As such, these
observations were treated as different observations across topics (yielding 60
observations instead of 15), which introduced various limitations. For example,
this treatment of observations may invalidate the assumption of independence of
observations from one topic to another, as each student observation is represented
four times in the data. To address this issue, each discussion predictor was
calculated as an individual’s score divided by that individual’s average over time,
thus making each score reflective of potential student effects across the topics.
This, however, is not an ideal solution, and does not address all issues arising from
treating the variables in this manner.

Another issue arising from the small sample size is that students’
demographic information (socioeconomic status, gender, and prior standardized
scores) are at the student level; thus, with only 15, not 60 observations, they are
therefore not strong controls in this study. Though the inclusion of total words
written (TW) as a control can be seen as comparable to using students’ prior
standardized scores, the issue of limited demographic information was not
properly addressed.
Other limitations in this study include the high variability of features associated with the second outcome, argument type, across students. This is in part due to the difficulty in using the coding scheme to code first-draft essays, as students often changed their claim mid-essay and it was not always clear which perspectives they were supporting. Future studies should ideally have final-draft essays, where scoring reliability can be more easily established.

Other limitations include the different times allotted for each topic (NFL and Wrongs were discussed over two class sessions each, while Justice was only discussed over one class, and Crime was discussed over three class sessions). Given the high average essay scores after the discussion on When is a crime not a crime?, future investigations should endeavor to investigate performance on topics that were discussed for the same length of time or when discussion time is systematically varied.

**Conclusions and Future Directions**

Despite these limitations, examining student talk using video data from a classroom implementing the Word Generation curriculum allowed for a unique opportunity to study the association between discussion and writing in eighth grade. The most important finding is that this study highlighted the differences between using the Low Inference Discourse Observation tool to study teacher-discussion moves and using it to examine student-produced discussion. The data captured here reflect the range of student input, but further study is required for
examining the content-related quality of students’ interactions during whole class settings.

Future investigations examining longitudinal discussion data should consider examining topic differences as well as differences in discussion structure, as these can impact the types of talk students produce, as well as their preparation for writing. For example, though students in this study had only one teacher, discussion formats and time allotted for discussion of each topic were not uniform; thus, students’ preparation for writing could have varied across topics.

This study presented a novel approach to investigating whether students’ discussion can predict their writing scores, but there are also opportunities to design interventions that allow for the study of reading comprehension as an outcome. Future studies should also examine, more closely, the distinction between participation and engagement, as well as between frequency and quality of discussion. An important aspect of discussion not captured in this study is that which occurs during small peer groups; future investigations should aim to explore student talk in both small and large group settings in order to better understand its relationship to student outcomes. Also, a much larger longitudinal dataset would allow for latent growth modeling, a direction that future research could take. Additionally, future directions could include examining students’ talk at the language-level, using natural language processing techniques that will allow automated identification of dialogue patterns that are predictive of post-discussion literacy outcomes such as essay scores.
References


School Officers, Washington D.C.


## Tables and Figures

### Table 1

**Discussion Topics and Writing Prompts**

<table>
<thead>
<tr>
<th>Word Generation Topic</th>
<th>Writing Prompt</th>
<th>Class Sessions</th>
<th>Discussion Time (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NFL: Should the NFL require the Washington Redskins to change their name?</td>
<td>Should the NFL require the Washington Redskins to change their name?</td>
<td>2</td>
<td>51</td>
</tr>
<tr>
<td>Crime: When is a crime not a crime?</td>
<td>Is it possible for justice to be blind? Respond to President Obama’s speech and give him advice on what you feel could be done to make a more perfect union (a better country). Be sure you comment on his claim that “there is a history of racial disparities in the application of our criminal laws,” making clear whether you think this is important to consider or not.</td>
<td>3</td>
<td>140</td>
</tr>
<tr>
<td>Justice: Where is the justice in our justice system?</td>
<td>Do you think people can change? Should someone who has been incarcerated be permanently condemned as a criminal? What do you think Hopwood means when he says citizens would need to change their views about the criminal justice system and prisoners?</td>
<td>1</td>
<td>52</td>
</tr>
</tbody>
</table>
| Wrongs: How do we right the wrongs of the past? | Write a blog entry that you would want Dr. King to read updating him on how the United States has or has not changed in the last 50 years. Two possible options are:  
  - The United States has taken great steps to achieve racial equality  
  - The United States has not confronted its past and still suffers from institutional racism | 2              | 41                    |
Table 2

Summary Statistics of all Outcomes, Predictors, and Controls

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>St. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outcome #1: Holistic Essay Scores (EScore) (N=60)</td>
<td>2.60</td>
<td>0.67</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Outcome #2: Argument Types (N=60)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Idea Units (ideas)</td>
<td>6.7</td>
<td>3.0</td>
<td>3</td>
<td>16</td>
</tr>
<tr>
<td>No Argument % (NOARG)</td>
<td>0.35</td>
<td>0.81</td>
<td>0</td>
<td>1.00</td>
</tr>
<tr>
<td>Own-side % (OWN)</td>
<td>0.55</td>
<td>0.21</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>Dual-perspective % (DUAL)</td>
<td>0.06</td>
<td>0.12</td>
<td>0</td>
<td>0.57</td>
</tr>
<tr>
<td>Integrative-perspective % (INT)</td>
<td>0.03</td>
<td>0.08</td>
<td>0</td>
<td>0.40</td>
</tr>
<tr>
<td>Predictor #1: Discussion Features (N=60)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student-to-Student Talk (S2S)</td>
<td>0.80</td>
<td>1.10</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Claim with Evidence (CE)</td>
<td>0.93</td>
<td>1.20</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Minimal Student Talk (MIN)</td>
<td>1.00</td>
<td>0.73</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Predictor #2: Proportion of talk (ptalk) (N=60)</td>
<td>0.07</td>
<td>0.08</td>
<td>0</td>
<td>0.37</td>
</tr>
<tr>
<td>Control (N=15):</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prior MCAS ELA scores (MCAS)</td>
<td>246</td>
<td>10.0</td>
<td>224</td>
<td>264</td>
</tr>
<tr>
<td>Control (N=60):</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total words written (TW)</td>
<td>133</td>
<td>64</td>
<td>33</td>
<td>310</td>
</tr>
</tbody>
</table>
Table 3

*Student Level Fixed Effects Model Showing Relationships among Discussion Features and Persuasive Essay Scores. There are 60 observations from 15 students*

<table>
<thead>
<tr>
<th></th>
<th>Dependent Variable: Holistic Essay Scores</th>
<th>Estimate (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Model 1</td>
<td>Model 2</td>
</tr>
<tr>
<td>Intercept</td>
<td>0.91</td>
<td>1.82</td>
</tr>
<tr>
<td></td>
<td>(0.19)</td>
<td>(0.22)</td>
</tr>
<tr>
<td>Student-to-Student (S2S)</td>
<td>−0.19**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.07)</td>
<td></td>
</tr>
<tr>
<td>Claim with Evidence (CE)</td>
<td></td>
<td>−0.05</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.06)</td>
</tr>
<tr>
<td>Minimal Student Talk (MIN)</td>
<td></td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.11)</td>
</tr>
<tr>
<td>Total words written (TW)</td>
<td>0.54***</td>
<td>0.51***</td>
</tr>
<tr>
<td></td>
<td>(0.11)</td>
<td>(0.13)</td>
</tr>
<tr>
<td>Prior MCAS ELA scores (MCAS)</td>
<td>0.06</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>(0.10)</td>
<td>(0.11)</td>
</tr>
<tr>
<td>Socioeconomic Status (SES)</td>
<td>0.22</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td>(0.24)</td>
<td>(0.25)</td>
</tr>
<tr>
<td>Female (FEM)</td>
<td>0.12</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td>(0.16)</td>
<td>(0.17)</td>
</tr>
<tr>
<td>AIC</td>
<td>124.1</td>
<td>131.2</td>
</tr>
<tr>
<td>DIC</td>
<td>75</td>
<td>84</td>
</tr>
<tr>
<td>Deviance</td>
<td>91.7</td>
<td>99.6</td>
</tr>
</tbody>
</table>

*p < .05 **p < .01 ***p < .001
Table 4

**Student Fixed Effects Model Showing Relationships among Student Discussion Features and Argument Types (Different Outcome in Each Model)**

<table>
<thead>
<tr>
<th>Dependent Variables</th>
<th>Estimate (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Idea Units</td>
</tr>
<tr>
<td>Intercept</td>
<td>1.21*</td>
</tr>
<tr>
<td></td>
<td>(0.60)</td>
</tr>
<tr>
<td>Student-to-Student (S2S)</td>
<td>-0.04</td>
</tr>
<tr>
<td></td>
<td>(0.18)</td>
</tr>
<tr>
<td>Claim with Evidence (CE)</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>(0.16)</td>
</tr>
<tr>
<td>Minimal Student Talk (MIN)</td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td>(0.26)</td>
</tr>
<tr>
<td>Total words written (TW)</td>
<td>3.80</td>
</tr>
<tr>
<td></td>
<td>(0.32)</td>
</tr>
<tr>
<td>Prior MCAS ELA scores (MCAS)</td>
<td>0.54~</td>
</tr>
<tr>
<td></td>
<td>(0.30)</td>
</tr>
<tr>
<td>Socioeconomic Status (SES)</td>
<td>1.04</td>
</tr>
<tr>
<td></td>
<td>(0.70)</td>
</tr>
<tr>
<td>Female (FEM)</td>
<td>-0.13</td>
</tr>
<tr>
<td></td>
<td>(0.48)</td>
</tr>
</tbody>
</table>

*p < .05  **p < .01  ***p < .001  ~p < .1
Table 5

*Student Fixed Effects Model Showing Relationships among Proportion of Talk and all Writing Outcomes*

<table>
<thead>
<tr>
<th>Dependent Variables</th>
<th>Essay Scores</th>
<th>Idea Units</th>
<th>No Own-side Arguments</th>
<th>Dual-Arguments</th>
<th>Integrative-Arguments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Model 10</td>
<td>Model 11</td>
<td>Model 12</td>
<td>Model 13</td>
<td>Model 14</td>
</tr>
<tr>
<td>Intercept</td>
<td>1.70***</td>
<td>1.20*</td>
<td>0.40***</td>
<td>0.56***</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>(0.21)</td>
<td>(0.52)</td>
<td>(0.06)</td>
<td>(0.08)</td>
<td>(0.05)</td>
</tr>
<tr>
<td>Proportion of talk</td>
<td>0.54</td>
<td>2.78</td>
<td>−0.10</td>
<td>0.02</td>
<td>0.13</td>
</tr>
<tr>
<td>(ptalk)</td>
<td>(1.06)</td>
<td>(2.72)</td>
<td>(0.32)</td>
<td>(0.41)</td>
<td>(0.25)</td>
</tr>
<tr>
<td>Total words written</td>
<td>0.54***</td>
<td>3.88***</td>
<td>−0.01</td>
<td>−0.01</td>
<td>0.00</td>
</tr>
<tr>
<td>(TW)</td>
<td>(0.12)</td>
<td>(0.29)</td>
<td>(0.04)</td>
<td>(0.04)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>Prior MCAS ELA</td>
<td>0.06</td>
<td>0.45</td>
<td>−0.02</td>
<td>0.04</td>
<td>0.00</td>
</tr>
<tr>
<td>scores (MCAS)</td>
<td>(0.11)</td>
<td>(0.30)</td>
<td>(0.03)</td>
<td>(0.05)</td>
<td>(0.03)</td>
</tr>
<tr>
<td>Socioeconomic Status (SES)</td>
<td>0.16</td>
<td>0.91</td>
<td>−0.10</td>
<td>0.02</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>(0.26)</td>
<td>(0.69)</td>
<td>(0.08)</td>
<td>(0.11)</td>
<td>(0.07)</td>
</tr>
<tr>
<td>Female (FEM)</td>
<td>0.18</td>
<td>−0.07</td>
<td>−0.03</td>
<td>−0.01</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>(0.18)</td>
<td>(0.47)</td>
<td>(0.05)</td>
<td>(0.06)</td>
<td>(0.05)</td>
</tr>
</tbody>
</table>

*p < .05 **p < .01 ***p < .001
Table 6

Discussion among Students during Unit #3: “Where is the Justice in our Justice System?” with Student-to-Student Codes Italicized

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Utterance</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>The criminal isn’t afraid of what the punishment is because they’re gonna do it anyway ‘cause they’re a criminal and that’s what they do, is that they do crimes and they don’t care about the punishment.</td>
</tr>
<tr>
<td>Teacher</td>
<td>So does that support incapacitation?</td>
</tr>
<tr>
<td>C</td>
<td>Yeah, does it?</td>
</tr>
<tr>
<td>A</td>
<td>It does, it does.</td>
</tr>
<tr>
<td>C</td>
<td>How?</td>
</tr>
<tr>
<td>Teacher</td>
<td>So he’s saying that a criminal is a criminal is a criminal.</td>
</tr>
<tr>
<td>C</td>
<td>Well that’s kind of like changing yours that’s like disagreeing with yours because you’re saying that a person can change.</td>
</tr>
<tr>
<td>A</td>
<td>No</td>
</tr>
<tr>
<td>Teacher</td>
<td>No, he’s not saying that.</td>
</tr>
<tr>
<td>C</td>
<td>Yeah, he is.</td>
</tr>
<tr>
<td>Teacher</td>
<td>Incapacitation doesn’t say that. What does it say?</td>
</tr>
<tr>
<td>A</td>
<td>It says that… (flips through workbook)</td>
</tr>
<tr>
<td>Teacher</td>
<td>What’s the intent?</td>
</tr>
<tr>
<td>A</td>
<td>The intent, incapacitation, wait I have it right here. [Reads] “It’s a punishment that prevents an offender from committing a crime again by controlling his or her action.”</td>
</tr>
<tr>
<td>C</td>
<td>Why did you say it takes five years to change a person?</td>
</tr>
<tr>
<td>A</td>
<td>The purpose is, like, to wake the person up and say, ‘you’re not doing stuff that is right and that you should stop doing it.’</td>
</tr>
<tr>
<td>C</td>
<td>You just said it again, a criminal is a criminal so they won’t change.</td>
</tr>
<tr>
<td>A</td>
<td>No, I do not say that.</td>
</tr>
<tr>
<td>D</td>
<td>Yes, you did!</td>
</tr>
<tr>
<td>A</td>
<td>I did not say that.</td>
</tr>
</tbody>
</table>
D  You said a criminal is a criminal.
A  I said a criminal is a criminal and a criminal is gonna do the crime while he’s a criminal.
C  Yeah! You just said it again.
Teacher  So he said that they won’t be deterred by what?
A  By punishment. I never said that they can’t change.
C  You just said a criminal is a criminal.
A  Yeah, a criminal is a criminal.
D  Before they get help
A  Before they get help and then if they get help, they can change.
Teacher  But incapacitation isn’t giving them help.
A  It could be actually
Teacher  It’s protecting society and putting them away. It doesn’t say anything about help, does it?
A  No, it could help them though.
D  They’re sitting by themselves for like eight years…[unclear]
A  No they’re not, they’re not sitting by themselves.
Teacher  D is right here, right? She’s saying that incapacitation leaves it up to the criminal to make that decision.
D  What they wanna do
A  No, no
Teacher  Read the definition again.
A  I know what the definition is but they do [laughter] they do help you with…
Teacher  What do they help you with?
D  They help you to get your life turned around. They have like people in there that can help you if you need help.
E  Isn’t that [unclear]
Teacher  Go ahead…I’m just gonna let the flood gates open. Go ahead E.
If there was people in there helping them, that would be rehabilitation, not incapacitation.

Amy Winehouse wrote a song about this.

I’m done.

No, no, you’re not done.

I am done.

No

I am done.

A just threw in the white flag.

A is making a good point, though. Incapacitation, and the…right, they’re trying to…go ahead

<table>
<thead>
<tr>
<th>Topic: Where is the Justice in our Justice System?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prompt: Do you think people can change? Should someone who has been incarcerated be permanently condemned as a criminal? What do you think Hopwood means when he says citizens would need to change their views about the criminal justice system and prisoners?</td>
</tr>
<tr>
<td>I believe people can change. Shon Hopwood did so why can’t others? Shon even believed that people can change. He says it only takes five years to actually get it. People should only be permanently condemned for murders. Other crimes I believe people can change but if you’re a murderer then you are a threat to society and should be condemned.</td>
</tr>
</tbody>
</table>

Score: 2 (essay contains claims and evidence/reasoning, but no counter-evidence or conclusion).
Figure 1. Essay scores ranged from 2 to 4, with over half receiving a score of 2, followed by 3 and 4.

Figure 2. Correlation Matrix showing correlations among predictors and outcomes (N=60)
**Figure 3.** Mean discussion utterances per minute and essay scores over time/topic

**Figure 4.** Student with highest frequency of utterances per minute over time/topic
Figure 5. Student with relatively low utterances per minute over time/topic
Chapter 4: Implications for Classroom Practice

Introduction

Increasingly, students in U.S. public schools come from diverse linguistic and social backgrounds. Over 10% of the student population in seven states and the District of Columbia are emergent bilinguals, and over half of all students in public schools are Black, Latino, Asian, or belong to other minority ethnic groups; both percentages have increased over the past several years and are projected to continue growing (U.S. Department of Education, 2016). With this cultural and ethnic diversity comes linguistic diversity, and students, whether considered emerging bilinguals or not, speak a variety of home languages with Spanish being spoken by nearly 8% of all public K-12 students in the United States (U.S. Department of Education, 2016). While the increased proportion of diverse learners enriches classrooms and provides tremendous opportunities for educators, their diverse learning needs also pose new challenges for educators who may need to incorporate new ways of instruction to meet the academic needs of all students.

Data from the 2015 National Assessment of Educational Progress (NAEP) reveal that on average, only 36% of fourth grade students and 34% of eighth grade students read proficiently (Kena et al., 2016). Writing scores for fourth and eighth grade students were not much different, with only 33% of eighth graders achieving proficiency in reading and 31% in writing (Kena et al., 2016). Though these figures are low, they were even lower for students who were eligible for free or reduced lunch, non-white students, and emergent bilinguals (Kena et al., 2016). In
other words, our schools are increasingly diverse but they were not meeting the needs of most students, with our most vulnerable students lagging behind their peers on reading and writing measures. In today’s changing world where an undergraduate degree is the accepted norm for employers, these figures are simply not acceptable. Educators, both practitioners and researchers, have been working tirelessly, and their efforts have helped improve students’ literacy outcomes over the years (Kena et al., 2016). However, we have much more work to do to ensure that all our students, White or Black, rich or poor, native or non-native speakers of English, exit our educational system with the tools and skills needed to successfully navigate today’s society.

Writing and Classroom Talk

Writing is an important component of literacy that does not receive as much attention as reading. Not only is writing difficult to produce (National Commission on Writing, 2006), but this complex cognitive, linguistic, and social task is oftentimes difficult to instruct, partly due to the lack of extensive research on writing instruction as well as the disconnect between research and practice (Myhill & Fisher, 2010). This paper will focus on a genre of academic writing that students encounter increasingly at the middle and high school level, persuasive writing. The Common Core State Standards emphasize the need for students to develop proficiency in persuasive writing across subject areas (National Governors Association, 2010); however, it is up to classroom practitioners to
decide how to prepare students to write persuasively. As such, we have much to learn about classroom factors that can improve students’ writing.

A classroom feature that is widely prevalent in any classroom is talk. Talk among students and talk between students and the teacher, is vital to today’s learning environments. Students talk to each other about content in pairs, work in small groups to discuss ideas or learn content, or engage (as listeners or active participants) in whole classroom discussions facilitated by the teacher. But are all forms of talk academically productive? (Chapin & O’Connor, 2007; Chapin, O’Connor, & Anderson, 2003) That is, do all forms of classroom talk help students achieve learning outcomes like content comprehension and critical thinking? These are questions educators have sought to answer (e.g. Cazden, 1998; Tharp & Gallimore, 1988; Nystrand & Gamoran, 1991; Michaels, O’Connor, & Resnick, 2008; Murphy, Wilkinson, Soter, Hennessey, & Alexander, 2009, Tharp & Gallimore, 1988; Kuhn & Crowell, 2011; Kuhn, Hemberger, & Khait, 2016).

Educational research has shown that one form of classroom talk that has been associated with producing positive learning outcomes for students is dialogic talk; that is, talk that helps students uncover meaning through dialogue (Bakhtin, 1984; Tharp & Gallimore, 1988). In other words, students learn best when they’re spoken with, rather than spoken to (Blachowicz & Fisher, 2011). Specifically, dialogic talk can improve students’ comprehension (Murphy et al., 2009), reasoning skills (Michaels, O’Connor, & Resnick, 2008; Kuhn & Crowell, 2011), and writing quality (Kuhn, Hemberger, & Khait, 2016). However, despite
accumulating evidence that dialogic discussions help improve various academic outcomes such as oral language skills, critical thinking skills, and reading comprehension, dialogic talk remains rare and difficult to achieve (Nystrand & Gamoran, 1991), with teachers often misidentifying IRE sequences (or recitation) as dialogic discussion (Nystrand, Wu, Gamoran, Zeiser, & Long, 2003). To better understand how teachers can incorporate more dialogic discussions in their classrooms, it is important to first identify components of classroom talk that are more likely to produce dialogic talk and those that are less likely to do so.

**Measuring Classroom Talk**

The Low Inference Discourse Observation tool (LIDO) is an innovative instrument that captures classroom discourse that is produced by both the teacher and students. Developed through an iterative process that built upon the work on *accountable talk* by Michaels, O’Connor & Resnick (2008), as well as measures of classroom talk (Mercer, Wegerif, & Dawes, 1999; Nystrand et al., 2003), the LIDO captures talk produced during whole class discussion.

While the dialogicity captured by the LIDO is a prerequisite to high quality discussion, attention to content (e.g. the nature of claims, the appropriateness of evidence, the relevance of the talk to the topic, etc.) is also crucial. Thus, higher LIDO scores are necessary but not sufficient for ensuring quality of discussion. Therefore, though not a measure of discussion quality, the student- and teacher-produced discussion features captured by the LIDO can be divided into those that are more or less reflective of dialogic practice in the classroom (Figure 1).
Dialogic practices initiated by teachers include encouraging students to respond to one another, following up with a student by asking that student to explain or clarify his/her response or provide evidence to support his/her claim, and signaling active listening to keep students talking. For example, a teacher may encourage students to respond to one another by asking, “Who agrees or disagrees with Maria?” Once a student responds, a teacher could further probe that student’s thinking by asking, “Why do you think that?” And when a student responds, the teacher could encourage him/her to continue speaking by using active listening indicators such as, “Oh, that’s interesting” or “Go on.”

These three features are considered more dialogic in that they promote talk that encourages students to learn through dialogue. The LIDO also captures three types of teacher content questions, one of which is considered highly dialogic, while the others are more indicative of the aforementioned IRE sequences often seen in classrooms. Dialogic questions are open, discussable, and contestable in nature. For example, when a teacher asks students to explain their thinking, without requiring a “correct” answer, they exemplify dialogic talk. On the other hand, the less dialogic teacher questions captured by the LIDO include asking semi-open questions, where the answer is not contestable but offers more latitude (e.g. “Explain the process of osmosis.”), and closed, “test-like” questions that assume there is a correct answer at which students need to arrive (“What do you get when you divide these two numbers?”)
As for the student-produced utterances during discussion, the LIDO captures more dialogic moves such as direct and indirect student interactions about content (i.e. instances where students are speaking to each other directly or are referring to each other’s ideas), as well as stating claims that are then supported through evidence and reasoning (e.g. “I agree with the author because we read about a similar story in Social Studies.”) On the other hand, less dialogic student utterances include responses that are either one clause or less (e.g. in response to a teacher’s question, “Do you agree or disagree?”, a student could answer, “I disagree” without further elaboration) or more than one clause (statements that involve longer student utterances that are not directed to other students, and which do not contain a claim with supportive evidence; e.g. “I really don’t know what to think about this.”) Taken together, the LIDO therefore helps us understand the various forms of talk produced by the teacher and students during whole class discussion, and can pave a path toward understanding how these talk moves can shape students’ learning.

To understand the relationship between these talk moves and student outcomes, I conducted two empirical studies that examined the role of talk among students and between students and teachers in classroom settings where the Word Generation program was implemented (Snow, Lawrence, & White, 2009). The program involves fourth through eighth grade students mostly in urban schools in the Northeastern United States, who discuss controversial dilemmas ranging from whether everyone should learn a second language (fourth grade), to whether the
United States has an obligation to address past inequities against African Americans and Native Americans (eighth grade).

Over 80% of students in the sample came from low-income families, as indicated by their eligibility for reduced or free school lunches. Using audio and video recordings of discussions, the two studies used the LIDO to capture students’ and teachers’ talk during whole class discussions. The studies sought to investigate the relationship between whole class discussion (produced by teachers and students) and post-discussion persuasive essay writing. The role of another classroom factor, student participation rate, was also examined. In the following sections, findings from both studies are discussed as answers to the questions investigated in both studies. Also, though causal inferences cannot be made using findings from these two studies alone, some considerations for future classroom practice can be drawn from the results reported in both papers.

**Research Findings**

**Are higher dialogic discussion practices related to higher essay scores?**

Findings suggest that teacher discourse practices deemed more dialogic are positively associated with post-discussion scores, even after taking into account students’ participation rate, the topic of discussion, students’ prior academic language scores, and individual differences among students (race/ethnicity, gender, socioeconomic status, English learner status and special education status). In other words, when teachers: a) encouraged students to respond to one another, b) followed up with a student by asking for an elaboration or the use of evidence
and reasoning to support a stated claim, c) used active listening to keep a student talking, and d) asked open, contestable questions, students’ post-discussion essay scores were higher. The relationship found was strong, in that, on average, each additional high dialogic teacher utterance per ten minutes of discussion time was associated with an increase of 1.33% in essay scores (Figure 2).

On the other hand, the studies conducted found a negative relationship between high dialogic discussion produced by students and their writing outcomes. In other words, students’ direct and indirect talk about content during whole class discussion was negatively associated with their post-discussion writing scores. This counter-intuitive finding can be further understood by including all student talk in a given classroom – that which is produced during whole class discussion (captured by the LIDO), and talk that occurs in small group discussions (not captured by the LIDO). We cannot fully understand the relationship between talk and discussion without first acquiring complete data about the range of student input during discussion.

Also, when examined closely, student discourse moves deemed dialogic did not necessarily challenge students’ previously-held beliefs, and students who engaged in direct or indirect talk with their peers did not seem to incorporate different perspectives in their writing. To illustrate, the exchange in Table 1 shows that Student A held on to his assertion despite protest from other students, and eventually quit the discussion altogether. Though dialogic in form, this exchange
failed to rise to the level of *academically productive talk* that serves to further
students’ understanding through dialogue (Chapin & O’Connor, 2007).

**Are less dialogic discussion practices related to lower essay scores?**

Less dialogic discussion practices (produced by both teachers and students,
thereby contributing to overall low dialogic classrooms), were negatively
associated with essay scores. That is, teacher discourse practices such as asking
semi-open or quiz-like questions during whole class discussions, as well as student
practices such as using minimal talk, were negatively associated with essay scores,
even after taking into account individual differences, prior test scores, and
classroom participation rate. This association, however, was not as strong as the
association between higher dialogic teacher discussion and essay scores, in that on
average, each additional low dialogic move every ten minutes of discussion time
was associated with a slight decrease of 0.28% in students’ essay scores (Figure
2). If established to be causal, this finding is encouraging in that teachers can be
coached to add more dialogic questions and practices to their repertoire of
discussion skills in order to better assist students’ learning.

***Insert Figure 2 Here***

**What is the relationship between participation and writing?**

The first study found that a higher participation rate can help students
improve their writing, above and beyond the contribution of topic, teacher- and
student-produced talk, and student demographic variables. That is, by simply
encouraging more students to participate during a discussion, teachers may be able
to improve their post-discussion persuasive writing. While some students may not benefit from increased participation in a whole class setting (i.e. some students may learn by silently reflecting on content discussed or engaging in small group discussions), being in classrooms where there is a high level of participation can help improve their writing. On average, an increase in overall student participation rate (e.g. from 25-50% participation to 50-75% participation) has an associated increase of 5.8% in essay scores.

This finding is significant in that it was observed despite accounting for the fact that the relationship between participation and writing may be influenced by topic. One mechanism through which this may be possible is that high-participation classrooms may allow for a wide range of ideas and perspectives, which in turn could help students consider multiple sources of evidence and counter-evidence when writing about the same topic previously discussed.

Implications for Classroom Practice

Given these empirical findings, the following instructional practices can be implemented during whole class discussions.

Emphasize quality and quantity

While high classroom participation is associated with better writing outcomes, it is important to consider the quality of talk during whole class student interactions. That is, if the goal is to build students’ critical thinking and perspective-taking skills (which in turn can improve their persuasive writing
skills), teachers need to facilitate discussions that produce high quality dialogic talk among students. But what is considered ‘high quality’?

Though the LIDO captures ‘dialogicality’ of talk, it does not capture quality of discourse, particularly when assessing student talk. To assess quality of discussion, teachers must make note of whether students’ talk is aligned with the goals of improving critical thinking and perspective-taking. For example, while students’ talk to one another can be dialogic, students may not yet have developed the listening skills needed to understand and build upon other students’ perspectives. Teachers, however, can revoice a particular student’s contribution during discussion (O’Connor & Michaels, 1993, 1996) in order to teach students that “learning means learning from others, taking advantage of others’ ideas” (Hiebert et al., 1996, p.17).

Prioritize contestable questions

When leading whole class discussions, teachers can help improve students’ writing outcomes by posing more contestable questions, and fewer semi-open or quiz-like questions. Although semi-open and closed, quiz-like questions certainly play a role in students’ learning (e.g. they can help educators assess whether students have learned important concepts or processes), they can come at the expense of contestable questions. For example, as both semi-open and quiz-like questions expect a correct response or set of responses, they are more likely to produce talk exchanges that resemble IRE sequences where students respond in minimal ways to answer the teacher’s question. On the other hand, an open,
contestable question does not have a “correct” answer; as such, these forms of questions can foster learning environments that are truly dialogic, where “intellectual openness, and possibilities for critique and creative thought” are encouraged (O’Connor & Michaels, 2007, p.277). Additionally, contestable questions encourage wide-ranging responses that can help students explore various ideas and, with teacher facilitation, shape and build upon their own previously-held views.

However not all topics can generate contestable questions; that is, not all topics are “discussable,” and when students are not interested in (or sufficiently prepared for) a topic, they may lose interest and not engage in dialogic ways, even when contestable questions are posed. Thus, teachers must decide which topics lend themselves to productive dialogue or discussion, and importantly, how to express questions in ways that encourage dialogic practice among students. Unlike semi-open and quiz-like questions, contestable questions aim to extend students’ thinking by challenging them to think beyond what is known to the teacher. To illustrate, consider the following three examples:

What was the outcome of the 2016 election?

What is the process of becoming president?

What led Americans to vote for Trump in 2016?

 Though all three questions in the previous example begin with “What,” they serve very different functions and aim to elicit varied student responses. The first question can be classified as a “quiz-like” question, as the answer is already
known to the teacher. Here, the teacher is testing students’ knowledge, and can expect one correct response. Unlike the first question, the second question examines students’ understanding of becoming president in the United States. As there are a set of correct answers to this procedural question, it is classified as a semi-open question. The third question, however, requires students to examine various economic, political, and social factors that led to Trump’s victory. Furthermore, the teacher cannot expect one or more “correct” responses to this question; students can disagree with one another and will likely present evidence and reasoning to support their claims. This makes this question truly contestable, discussable, and open-ended, and will serve to develop students’ critical thinking and argumentation skills.

**Additional Considerations**

The two papers highlighted the need to further explore other aspects of conducting dialogic discussions that may be pertinent to students’ learning outcomes. For example, students’ writing outcomes may be related to the time allotted to a given topic, with the assumption that more time allows students to catalyze their comprehension (and thereby, writing). Additional time for essay writing, including drafting, editing, and re-writing, may also be necessary in order for students to fine-tune their persuasive writing skills. Finally, no discussion can be productive without previously-set classroom norms that encourage students to share ideas in an equitable and productive manner while also scaffolding students to effectively engage in discussions.
Conclusions

Despite their associations with positive academic outcomes, dialogic discussions are rare; one study found that dialogic discourse occurred during 52 seconds per class hour in eighth grade (Nystrand & Gamoran, 1991). Some teachers also report either not having time to conduct dialogic, whole class discussions, or felt unprepared to manage such discussions when they occur. Despite these challenges, talk is a recurrent feature of any k-12 classroom; despite its ubiquity, however, it varies widely from one classroom to another. This paper outlined main findings from two empirical studies and related implications for practice that can help teachers lead and facilitate discussions that are beneficial for fourth through eighth grade students’ persuasive writing.
References


Table 1

Discussion among Students during Unit #3: “Where is the Justice in our Justice System?” with Student-to-Student Codes Italicized

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Utterance</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>Why did you say it takes five years to change a person?</td>
</tr>
<tr>
<td>A</td>
<td>The purpose is, like, to wake the person up and say, ‘you’re not doing stuff that is right and that you should stop doing it.’</td>
</tr>
<tr>
<td>C</td>
<td>You just said it again, a criminal is a criminal so they won’t change.</td>
</tr>
<tr>
<td>A</td>
<td>No, I do not say that.</td>
</tr>
<tr>
<td>D</td>
<td>Yes, you did!</td>
</tr>
<tr>
<td>A</td>
<td>I did not say that.</td>
</tr>
<tr>
<td>D</td>
<td>You said a criminal is a criminal.</td>
</tr>
<tr>
<td>A</td>
<td>I said a criminal is a criminal and a criminal is gonna do the crime while he’s a criminal</td>
</tr>
<tr>
<td>C</td>
<td>Yeah! You just said it again.</td>
</tr>
<tr>
<td>Teacher</td>
<td>So he said that they won’t be deterred by what?</td>
</tr>
<tr>
<td>A</td>
<td>By punishment. I never said that they can’t change.</td>
</tr>
<tr>
<td>C</td>
<td>You just said a criminal is a criminal.</td>
</tr>
<tr>
<td>A</td>
<td>Yeah, a criminal is a criminal</td>
</tr>
<tr>
<td>D</td>
<td>Before they get help</td>
</tr>
<tr>
<td>A</td>
<td>Before they get help and then if they get help, they can change.</td>
</tr>
</tbody>
</table>

*(several exchanges later)*

| A       | I’m done. |
| Teacher | No, no, you’re not done. |
| A       | I am done. |
| Teacher | No |
| A       | I am done. |
| C       | A just threw in the white flag. |
Figure 1. Teacher and student talk moves captured by the Low Inference Discourse Observation Tool (LIDO). Moves at the top of the pyramids are hypothesized to be more dialogic than those at the bottom of the pyramid.
Figure 2. Predicted effects of high teacher dialogic discussion and low overall dialogic discussion on essay scores per 10 minutes of discussion time
Chapter 5: Thesis Conclusion and Future Directions

The work presented in this dissertation began with a simple question, *Can discussion improve writing?* While the data used in this study are drawn from the larger *Catalyzing Comprehension through Discussion and Debate* study, a large randomized trial comparing intervention and control classrooms, the nature of the empirical work presented in Study 1 and Study 2 does not lend itself to drawing causal inferences about the relationship between classroom talk and writing outcomes. However, the work provides revealing insights into how classroom discussion is realized throughout various fourth through eighth grade classrooms in the Northeastern United States, as well as implications for pedagogical practice.

In this chapter, findings will be summarized and reflected upon, and recommendations for future research in this field will be suggested. Finally, I will end with a discussion on approaches to help move the field of study in directions that make use of advances in the field of Artificial Intelligence.

I. Summary of Findings

The first paper began with examining classroom-level data. Using the Low Inference Discourse Observation Tool (LIDO) to code all whole class interactions around content, the paper examined 42 classrooms and the 471 resulting essays from 369 students in grades four through seven. The multilevel analysis revealed that when students are part of classrooms that are characterized by more dialogic teacher approaches, their associated essay scores were higher. Conversely, classrooms characterized by low teacher- and student-produced talk were
associated with essays that were relatively lower, on average. In this paper, classroom-level student contributions that were coded as dialogic using the LIDO were also associated with lower essay scores, contrary to the hypothesized positive relationship. Another classroom feature, participation rate, was also positively associated with essay scores, even when accounting for the effects of topic on participation rate. Puzzling findings about student participation from the first study prompted the investigation presented in Study 2, which examined student-level data in one classroom over eight discussion sessions.

In the second study, fifteen eighth grade students discussed four topics over eight class sessions. Though the topics all focused on civic dilemmas, the types of discourse produced by students varied across subjects. On average, higher dialogic features were least common across all topics, while lower-dialogic features (Minimal Talk) were most prevalent. Also, while on average, dialogic moves increased over time, Minimal Talk decreased. Despite these discussions becoming more dialogic as time passed, student dialogic discussions were negatively associated with essay scores – a finding that was consistent with the previous study.

To understand this finding, it is important to reflect upon the way dialogic talk was coded using the LIDO. When capturing teacher talk, the LIDO distinguishes between three types of questions asked by teachers: contestable questions, semi-open questions, and quiz-like questions. This distinction can be understood as a proxy for quality, as it rests on the assumption that quiz-like
questions do not adequately prepare students to engage in critical thinking (Alexander, 2008; Cazden, 1988). Specifically, quiz-like questions help students catalyze knowledge such as facts learned, while semi-open questions can solidify students’ understanding of processes and procedures; these outcomes, however, are not as pertinent for persuasive writing as they are for comprehension of content. On the other hand, open-ended questions elicit the back-and-forth discourse that allows students to not only hear various perspectives, but also challenge and question these views – skills that are necessary for persuasive writing. Additionally, the teacher, as the expert, can be presumed to adhere to the topic and ask worthwhile questions. Conversely, when capturing student talk, the LIDO does not have any such distinguishing categories, and therefore only focuses on occurrences of a hypothesized ‘dialogic’ feature rather than what can be approximated to quality of that feature. That is, what may be dialogic in form (e.g. students speaking to one another directly about content), cannot be assumed to also be of high quality in nature given students’ novice status as interlocutors. Additionally, the LIDO does not capture all student talk during a given discussion as it excludes any talk occurring in small groups. As such, conclusions about student talk and its value for promoting (or hindering) writing cannot be made without first capturing the full range of student output during discussion.

In the second study, participation rate played no role in students’ post-discussion essay, most likely due to the high variability among students’ proportion of talk in a relatively small sample. These findings highlight the
importance of evaluating the quality of talk, rather than solely focusing on quantity of student talk, as the mere presence of dialogic talk does not guarantee that students will be engaged in academically productive talk that improves academic skills associated with persuasive thinking and writing.

The third paper explores instructional principles that arise from the findings in Study 1 and Study 2. These include emphasizing both increased talk among students (quantity) while also emphasizing quality of talk; that is, student talk should be truly dialogic in the sense that it not only involves dialogue, but also dialogue that improves learning and thinking. Teachers can help ensure student talk is of high quality by playing an active role in facilitating student discussions, and using dialogic moves described that have been shown to be associated with better writing outcomes (see Study 1).

Another important takeaway is the need to incorporate more contestable questions, and fewer semi-open or quiz-like questions during whole class discussion. While the latter two forms of questions certainly have a place in the classroom, they do not exemplify the dialogic talk associated with improved persuasive essay scores. Teachers are also reminded that not all topics can be discussable, but when they are, sufficient time should be allotted for discussion.

II. Recommendations for Future Research

The importance of discussion in today’s classrooms cannot be overstated. As instruction moves beyond the classroom (either entirely or through use of supplementary online platforms), discussion becomes the most distinguishing
aspect of traditional classrooms. Unlike online environments, discussions held in classrooms are face-to-face, in person, and are facilitated by the teacher. To maximize the impact of discussions on learning outcomes, however, it is important to use tools like the LIDO to help teachers understand how language can be used to further thinking, how to facilitate discussion (e.g. by using active listening and pressing a student to clarify or support a response), and which types of questions produce the desired learning outcomes (e.g. contestable vs. semi-open or quiz-like questions).

The work presented in this dissertation began with investigating classroom-level factors that predict writing (Study 1) followed with the relationship between student-level classroom discussion and persuasive writing (Study 2). Findings from these two papers highlight important future directions, some of which will be outlined in this next section.

Data Collection Considerations

The field of research exploring the relationship between dialogic discussion and student outcomes such as writing has helped further our understanding of classroom factors that can shape students’ learning (e.g. Mercer, Wegerif, & Dawes, 1999; Nystrand, Wu, Gamoran, Zeiser, & Long, 2003; O’Connor & Michaels, 2007; Michaels, O’Connor & Resnick, 2008; Kuhn & Crowell, 2011; Snow, Lawrence & White, 2009). However, the studies presented in this dissertation offer a unique insight into: a) how to capture all student and teacher talk during whole class discussion using a unique, validated instrument (LIDO),
and b) understanding the predictive relationship between various features of whole class talk and writing as a learning outcome.

As the goal of educators is ultimately to improve classroom practice, the next phase of this work should provide evidence-based tools and resources that educators can use in their classrooms to inspire dialogue that supports positive outcomes for their students. In order to do so, however, the field must build upon the rich insights provided by observational and correlational work with relatively small samples of students and head toward intervention studies that can help establish causality between discussion and outcomes such as reading comprehension and writing quality.

Future investigations should also make use of large databases of student data at the district or state level that include demographic variables and standardized test scores that provide researchers with various control variables and outcomes, which, when used in conjunction with classroom data on student discourse, can shed light on how classrooms characterized by higher dialogic practices and those characterized by lower levels of dialogic practices relate to various student outcomes.

Collecting classroom discourse data can also be accomplished without intervention studies; that is, as classroom talk (in one form or another) already occurs in various classrooms, it can be recorded and related to student learning outcomes. If it can be established, with strong evidence, that higher dialogic discussion does in fact positively predict outcomes such as reading comprehension
and writing quality, interventions that provide teachers with professional
development or coaching can be implemented to help teachers facilitate
discussions that are more dialogic in nature.

One issue with collecting discussion data, however, is the length of
classroom discussions and the complex nature of classroom talk, which can
require significant time and resources to manage and analyze using tools such as
the Low Inference Discourse Observation Tool. To address this issue, shorter
recordings of discussions can be captured to help shed light on their association
with student outcomes. If these shorter recordings of classroom discussions are
collected over multiple class sessions as opposed to one time point, they will
enable easier (and less time-consuming) analyses of discussion using tools such as
the LIDO. They will also provide more robust findings given that they represent a
more accurate assessment of classroom discourse, regardless of topic and other
factors that can vary from one day to another. In addition to data collection, the
field can (and should) take advantage of technological advances that help analyze
oral and written language produced by teachers and students.

**Measuring Student Outcomes**

The data presented in this dissertation explored one post-discussion student
learning outcome: persuasive writing quality. This is due to the fact that
intervention classrooms using the Word Generation curriculum were designed
such that students wrote persuasive essays on the same topic they discussed during
the week. Other classrooms that were part of the *Catalyzing Comprehension*
through Discussion and Debate study (e.g. control classrooms where whole class discussions were also held and coded using the LIDO) did not have an outcome that was closely related to the topic of discussion. As such, only intervention classrooms were examined, and only writing quality was used as an outcome in this dissertation. Future investigations can further the field by also exploring literacy outcomes such as reading comprehension, perspective taking, and reasoning skills, using large-scale data that will provide more robust associations between discussions and these important student learning outcomes.

**Group Differences**

Prior to entering schools, students’ learning outcomes differ based on factors such as socioeconomic background. This is due to the literacy environments that shape their acquisition of vocabulary (Hart & Risley, 1995), vocabulary growth (Rowe, Raudenbush, & Goldin-Meadow, 2012), and comprehension and sentence complexity (Fenson et al., 1994). Unfortunately, these differences persist throughout a child’s schooling, as seen by the wide and persistent gap in eighth grade reading among students based on family income level, native- and non-native speakers of English, and ethnic/racial groups (U.S. Department of Education, 2016). Schools, in this way, reproduce inequities observed in society, as children “can become the casualties of invisible criteria in school” (Zwiers, 2014, p.9) that expects them to participate in discourse contexts that differ from those into which they have been socialized (Schleppegrell, 2004). This results in what Macedo (1994) calls the “pedagogy of entrapment,” whereby
students are required to engage in and master academic skills that are not explicitly taught in schools (p.34).

An equitable educational system, however, will ensure that all students’ academic needs are met, regardless of their families’ income levels, the home languages spoken at home and/or students’ emergent bilingual status, ethnic or racial backgrounds, and other individual factors. If education is to be the “great equalizer,” as Horace Mann once said, then we must, as educators, adjust our instruction to provide high quality instruction to all students, especially our most vulnerable ones. Could the positive effects of teacher dialogic discussion, for example, matter more for students who come from low-income backgrounds? Could students who are emergent bilinguals and thus not as proficient in their oral and written language skills benefit more from discussion-based pedagogy? To answer these questions, we must first understand how discussion can predict different students’ literacy outcomes.

Most participants in these studies overwhelmingly came from low-income backgrounds, a group that has traditionally lagged behind their peers in attaining literacy outcomes such as reading and writing (U.S. Department of Education, 2016). However, research suggests that students deemed ‘struggling’ are likely to benefit more from dialogic pedagogical approaches than their higher-achieving counterparts (Kong & Pearson, 2003; Murphy et al., 2009). As such, it is important to encourage as well as equip teachers with the tools required to create classroom environments that prioritize dialogic talk, and that prepare students for
persuasive writing and other academic outcomes through discussion. Future research could also explore, more explicitly, the difference in post-discussion achievement between students from various socioeconomic backgrounds using samples of students with equitable distributions of income levels.

Another direction worth exploring in future studies on discussion is difference by grade level. Though participants in Study 1 came from four grade groups (fourth through seventh), differences by grade was not examined due to insufficient representation at the higher grades. Specifically, there were only three sixth grade and five seventh grade classrooms compared to 16 fourth grade and 18 fifth grade classrooms. Future investigations of this topic should examine discussion at various grade levels in order to better assess the following: a) how does dialogic discussion develop or change by grade? and b) does the relationship between discussion and student outcomes such as reading comprehension and writing differ by grade?

The second study focused on a different group, eighth graders, and though some findings were consistent (i.e. in both groups, high student dialogic discourse was negatively predictive of essay scores), others were not (participation level did not predict writing scores in the second paper as it had done in the first.) This raises the possibility that there may be differences in student participation by grade level, furthering the need to explore this group difference in future investigations.
Discussion and Writing Mechanisms

The studies presented in this dissertation highlight that not all discussion is good discussion. Additionally, though all teachers led classroom discussions, they varied widely in their implementation of discussions that enhance learning and thinking. This highlights the importance of working with teachers to broaden their understanding of discourse practices that lead to better student outcomes, as well as their capacity to lead such discussions. The same principle can be applied to writing. Though all students wrote persuasive essays, not all essays were, in fact, persuasive. Zwiers (2014) emphasizes that it is not sufficient for teachers and students to know the tools required for success in academic contexts; rather, teachers must ensure that students learn the skills that enable them to effectively use such tools, and this can only be done through explicit instruction and practice. Teachers should endeavor to help students achieve the goals outlined in the Common Core State Standards by teaching students how to organize their ideas into essays that are not only well-structured, but are also well-supported using evidence or reasoning, and concisely written using appropriate academic language features.

Discussion Type and Form

Though the Low Inference Discourse Observation tool was created for the specific purpose of studying whole class interactions, it is worth noting that student participation level may vary by student depending on whether the discussion takes place in whole class or small group settings. For example, quieter
students may not feel comfortable participating in whole class discussions, but may either feel more comfortable or may be unable to avoid participation in small group settings. Also, while teacher participation is essential for any whole class discussion, small groups, by their very nature, limit the teacher’s presence during discussions, a variable which could change the way students interact with one another. It is therefore imperative to compare discourse and participation level among students in whole class settings as well as in small group settings, and to then assess whether and to what degree participation in different groups predicts post-discussion academic outcomes.

Teachers employ various approaches when facilitating discussions. To illustrate, though there was only one teacher in Study 2, whole class discussion manifested differently, with the teacher sometimes allowing whole class interactions among all students while at other times limiting debate to small-group representatives. Using larger samples would allow for an investigation into the full array of discussion types to evaluate whether some forms of discussion produce better outcomes than others.

The recommendations for future work outlined in this section aim to extend the work beyond correlational studies and into the domain of causal analyses. Robust links established between dialogic classroom talk and various student outcomes can consequently enable the development of evidence-based pedagogy that can leverage research findings for better student outcomes such as reading comprehension, writing quality, and development of reasoning skills. In the next
and final section of this chapter, I will discuss a path forward for research in this domain, inspired by advances made in other fields such as computer science and implemented in fields such as medicine.

III. Toward Big Data in Education

Advances in artificial intelligence (AI) continue to infiltrate nearly every aspect of modern life, from the way we drive (self-driving cars), to the way we play (augmented reality), and even the way we heal (surgical robots). The technologies that are at the heart of machine learning systems include speech-to-text (e.g. Google home), video stream processing (e.g. self-driving cars), and natural language processing (Google’s search engine); all three of these fundamental components of artificial intelligence have immense potential to assist educational researchers and practitioners alike in their pursuit of better, evidence-based pedagogy. The adoption of these technologies will hinge on their ability to integrate with, rather than compete with, the insights of teachers. Just as AI has eliminated redundant tasks in other industries, so too can it within our educational community.

There are three specific ways that existing AI tools can facilitate the collection of new data in educational settings: 1) speech-to-text systems can be used for the passive monitoring of student activity in classrooms as well as the transcription of that activity into text; 2) video stream processing can be used to monitor and even understand non-verbal student-student and student-teacher interactions; and 3) natural language processing can be used to automatically
transcribe student hand-written essays, and perform semantic and syntactic analyses on digital essays. The value of these data modalities, when collected together, exceeds the value of any when collected individually. That is, with access to the video, audio, and text from classrooms, we can start to collect datasets that for the first time, provide a truly comprehensive insight into classroom practices.

Of course, the opportunities provided by such data will come with unique challenges. More specifically, the challenges associated with the collection of such data include: a) building networks of trust with schools and districts to allow for the collection of such data; b) the responsible storage and meticulous curation of data; and c) computational frameworks to enable the analysis and distribution of the data to the wider community of experts. A demonstrable digital research ecosystem that responsibly stores, de-identifies, and curates data will help solve the first two of these three challenges. AI will play a central role in the creation of this digital ecosystem.

Recent advances have made the automated and rapid de-identification of sensitive documents, videos, and images in other domains such as medicine possible for the first time (Neamatullah et al., 2008). These advances may be translated and applied in our educational domain, to ensure that all collected data are in compliance with the strictest of federal and state regulations that protect the privacy of children in schools. As has been accomplished with sensitive medical records, once de-identified, such data may be shared with the wider community of
experts (Johnson et al., 2016). The sheer volume of such data will present exciting opportunities for research, but it will also introduce new computational challenges. The video data used in the second study of this dissertation was already 21GB in size despite the fact that it contained data from a mere 15 students across eight discussions. As we build partnerships and begin to collect hundreds of hours of data from thousands of students, the management of such data itself will become a challenging task. Instead of requiring investigators around the world to buy and manage their own storage, I propose a centralized approach to the problem. Specifically, all data may be housed in highly secured data servers, where curated content may be searched, and selectively downloaded in accordance with the specific needs of the research team.

Closing Thoughts

The work presented in this dissertation is a culmination of years of research that was first inspired by my work as a classroom teacher in a culturally and linguistically diverse middle school. Ultimately, the goal of my work is to ensure that all students, no matter their linguistic, cultural, or socioeconomic background, have access to public education that is both responsive to the personal needs of students and is also driven by robust research findings. To that end, I seek to further the work in this domain by elaborating on the dissertation journey that initially began with the question, Can discussion improve writing?
References


Bibliography


