THE EFFECT OF CLASSROOM DISCOURSE ON HIGH SCHOOL STUDENTS’ ARGUMENTATIVE WRITING SKILLS

A thesis presented
by
Karl D. Sineath

to
The School of Education

In partial fulfillment of the requirements for the degree of
Doctor of Education

in the field of
Education

College of Professional Studies
Northeastern University
Boston, Massachusetts
February 2014
Abstract

On the writing component of the most recent National Assessment of Educational Progress, 12th grade high school students’ scores on argumentative tasks were lower than on tasks that required them to explain or to convey an experience (National Assessment of Educational Progress, 2011). Similarly, 11th and 12th graders at the research site for this study have struggled most with writing tasks that require them to integrate sources into an argument about an issue of public discourse. In response to these challenges, this quasi-experimental study investigated the effect of classroom discourse on argumentative writing among English language arts students in 11th and 12th grades in a Boston public high school. Unlike previous work in this area, this study (1) focused on high school upperclassmen in English classes, (2) used a standardized assessment tool designed by the College Board to measure growth in argumentative writing, and (3) provided descriptions of the discourse interventions to be followed by classroom teachers, which could later be implemented by other teachers outside of the study. At the conclusion of the study, the estimated marginal mean on a College Board assessment of argumentative writing for students who experienced discourse where teachers focused on linking ideas and pressing for reasoning \( (M = 3.90, SD = 1.40) \) was significantly higher than for students who experienced procedural facilitation where teachers encouraged discursive interactions but did not explicitly prompt students to articulate their reasoning or to link their ideas to those of others \( (M = 3.32, SD = 1.43) \), \( F(1, 112) = 8.056, p = .005, \eta^2_p = .067 \). These results provide confirming evidence for sociocognitive learning theory and help fill gaps in the literature as suggested above. In addition, the present study has helped to confirm linking ideas and pressing for reasoning as discourse moves related to improvements in argumentative writing.

Keywords: discourse, argumentative writing, English education, secondary education
Acknowledgments

I would like to thank Lexie Oosting, Malcolm Ray Oosting-Sineath, Dr. Yufeng Qian, Dr. Gail Matthews-Denatale, Dr. Yamila Hussein, Josh Kronenberg, Brett Dickens, Audrey Schindler-McDonald, and all of the students who participated in this study.
Table of Contents

ABSTRACT ........................................................................................................................................... 2
ACKNOWLEDGMENTS ......................................................................................................................... 3
LIST OF TABLES ..................................................................................................................................... 9
LIST OF FIGURES ................................................................................................................................. 10
CHAPTER 1: INTRODUCTION ............................................................................................................. 11
   Statement of the Problem .................................................................................................................. 12
   Significance of the Research Problem ............................................................................................... 13
   Research Question, Definitions of Key Terms, and Hypothesis ....................................................... 15
      Research question .......................................................................................................................... 15
      Definitions of key terms ................................................................................................................ 16
         Classroom discourse ................................................................................................................. 16
         Procedural facilitation .............................................................................................................. 16
         Linking ideas and pressing for reasoning ............................................................................... 17
         Argumentative writing ............................................................................................................. 17
   Directional Hypothesis .................................................................................................................... 17
   Positionality Statement .................................................................................................................... 17
   Sociocognitive Theory: A Theoretical Framework for Classroom Discourse ................................. 19
      Overview ...................................................................................................................................... 19
      Propositions .................................................................................................................................. 22
      Variables ........................................................................................................................................ 25
      Conclusion ................................................................................................................................... 26
CHAPTER 2: LITERATURE REVIEW ..................................................................................................... 28
Section 1: Discourse as Pedagogical and Political ..........................................................29
Making meaning through discourse .............................................................................30
When discourse fails .....................................................................................................32
Teaching discourse styles ............................................................................................34
Section 2: Promising Discourse Methods for Augmenting Student Learning .............37
Discourse and student engagement .............................................................................37
Discourse’s effect on test scores .................................................................................38
Discourse and argumentative writing at the secondary and post-secondary levels ..............................................................................................................................39
The influence of teachers’ discursive choices on students’ discourse ......................41
The effects of different discursive styles on student achievement ............................42
Promising discursive techniques ..................................................................................43
Section 3: Gaps in the Literature ..................................................................................48
Review of theoretical framework ................................................................................48
Review of methods .......................................................................................................49
Review of instruments .................................................................................................50
Review of data sources and data problems ..................................................................51
Review of research on grade-levels .............................................................................53
Areas for Further Research .........................................................................................54
Conclusion ....................................................................................................................56

CHAPTER 3: RESEARCH METHODS ..............................................................................60
Research Question .......................................................................................................60
Directional Hypothesis .................................................................................................61
Research Design ........................................................................................................... 61
Population and Sampling ........................................................................................... 63
  Research site ........................................................................................................... 63
  Sampling .................................................................................................................. 65
  Power analysis ......................................................................................................... 66
  Participants ............................................................................................................... 67
Data Collection .......................................................................................................... 68
  Instruments .............................................................................................................. 68
    Instrument selection .............................................................................................. 69
    Instrument validity ................................................................................................. 70
    Measurement scale and reliability ......................................................................... 71
  Procedures .............................................................................................................. 72
    Summary of how procedures were used .............................................................. 76
    Fidelity of implementation ................................................................................... 77
    Possible extraneous variables .............................................................................. 77
Data Analysis .............................................................................................................. 80
  Preparation of the data file ..................................................................................... 80
  Transformation of the data ....................................................................................... 82
  Choice of statistical techniques .............................................................................. 85
  Answering research question ................................................................................... 86
  Steps ......................................................................................................................... 86
Validity, Reliability, and Generalizability .................................................................. 88
  Internal validity ........................................................................................................ 88
Reliability ......................................................................................................................... 88
Generalizability (external validity) ................................................................. 91
Population validity ................................................................................................. 91
Ecological validity ................................................................................................. 92
Protection of Human Subjects .............................................................................. 94
Concerns about bias .............................................................................................. 95
Conclusion ................................................................................................................ 96

CHAPTER 4: REPORT OF RESEARCH FINDING ......................................................... 97

Data Transformation ............................................................................................... 98
Demographics of School, Population, and Sample ............................................... 99
Fidelity to the Discursive Method ........................................................................... 101
Interrater Reliability ............................................................................................... 102
A Priori Power Analysis ......................................................................................... 102

Testing Assumptions .............................................................................................. 103
Approximately equal sample sizes—assumption met ........................................ 103
Absence of outliers—assumption met ................................................................. 104
Normal distribution of covariate and dependent variable—assumptions met .... 104
Linear relationship between covariate and dependent variable—assumption met .... 104
Correlation between covariate and dependent variable—assumption met ........ 104
Independence of covariate—assumption met ...................................................... 105
Homogeneity of regression slopes—assumption met ......................................... 106
Equality of error variances—assumption met ...................................................... 106
Running ANCOVA ..............................................................................................................107
  Main effect before including covariate .................................................................107
  Applying a covariate to reduce error variance ...................................................108
  Effect size and observed power analysis ............................................................109
Hypothesis Testing ........................................................................................................109
Summary ......................................................................................................................111
CHAPTER 5: DISCUSSION OF RESEARCH FINDINGS .................................................112
Results and Discussion of Research Question ....................................................112
Implications ..................................................................................................................113
  Implications for theory ..........................................................................................115
  Implications for research ......................................................................................116
  Implications for practice .......................................................................................117
  Implications for generalizability ..........................................................................118
Generalizability to a larger population .....................................................................118
Generalizability to other instructional contexts .....................................................120
Limitations ...................................................................................................................123
Areas for Further Research .....................................................................................126
Conclusion ...................................................................................................................129
References ..................................................................................................................131
Appendix A: Checklist for Procedural Facilitation ................................................145
Appendix B: Tables and Figures Associated with Assumption Checking for ANCOVA ....148
List of Tables

1  Research on Discourse by Grade-Level and Topic ........................................................... 54
2  Demographic data of Community Leadership School ....................................................... 63
3  School, Population, and Sample Demographics ................................................................. 100
4  School, Sample, District, and State Argumentative Writing Scores ................................. 101
5  Argumentative Writing Scores without Covariate ............................................................. 107
6  Estimated Marginal Means for Argumentative Writing Scores ....................................... 108
B1  Descriptive Statistics for Covariate and Dependent Variable ......................................... 151
List of Figures

1  Schematic for sociocognitive theory ...............................................................21

2  Continuum on the effects of advertising .........................................................73

B1 Boxplots for covariate and dependent variable ..............................................148

B2 Histogram for covariate (pretest argumentative writing scores) ....................149

B3 Histogram for dependent variable (posttest argumentative writing scores) ......150
The Effect of Classroom Discourse on High School Students’ Argumentative Writing Skills

Chapter 1: Introduction

In each of the last two years, more than 400,000 high school students have taken the Advanced Placement English Language and Composition exam, making it the largest AP exam (AP data: Program summary report, 2011; AP data: Program summary report, 2012). On this exam, students must write three essays, and in each of the last two years, test-takers have scored highest on the synthesis essay, which requires students to integrate sources into an argument about an issue of public discourse (English: English Language and Composition course description, 2010). On the 2011 and 2012 exams, the global means on the synthesis essay (4.9 and 5.0, respectively) were 0.5 higher than the next closest essay on the 9-point scale. Despite students’ success on this exam at the national level, students at Community Leadership School (CLS)—the research site for this study—have struggled the most with this synthesis essay (in order to protect anonymity, all proper names in this document, including the name of the school, are pseudonyms). In 2011 and 2012, the majority of seniors at CLS took the AP English Language and Composition Exam (N = 56 for both years), and in each of these years, students’ scores on the synthesis essay showed the largest gap between the school mean and the global mean (AP instructional planning report, 2011; AP instructional planning report, 2012).

Despite relative national success on the synthesis question for students who enroll in AP English Language, on the writing component of the most recent National Assessment of Educational Progress, students’ scores on argumentative tasks were lower than on tasks that required them to explain or to convey an experience (National Assessment of Educational Progress, 2011 Writing assessment, 2011). More specifically, only 14% of grade 8 students and
23% of grade 12 students scored at the competent or effective level on argumentative writing tasks. By comparison 32% of 12th-grade students scored competent or effective on tasks that required them to convey an experience, and 26% on tasks that required them to explain. For 8th-grade students, 19% scored competent or effective on tasks that required them either to convey an experience or to explain. In addition, students with a similar demographic to those at CLS—who identify as Black or Hispanic, who attend school in urban areas, and who receive lunch aid—did not perform as well on these writing tasks as students who are White or Asian, students who attend school in suburban areas, and students who do not receive lunch aid (National Assessment of Educational Progress, 2011 Writing assessment, 2011).

Taken together, these data suggest that advanced students, such as those who generally enroll in AP English courses, are experiencing success on argumentative tasks that require students to integrate and analyze evidence to substantiate claims; however, the general population of students, including those who have been historically underserved by education, struggle on these tasks. Given that some groups do experience success on these argumentative tasks, how can research inform the instruction of argumentative writing among students who struggle?

**Statement of the Problem**

Research on classroom discourse has shown promise for augmenting persuasive argumentation (Felton & Kuhn, 2001; D. Kuhn & Crowell, 2011; D. Kuhn, Shaw, & Felton, 1997; Reznitskaya et al., 2001). However, this research has historically used researcher-designed assessments of student-learning, failed to provide adequate detail of the discursive method that would allow replication, and excluded analysis of the effect of the discourse on high school upperclassmen (Murphy, Wilkinson, Soter, Hennessey, & Alexander, 2009). Such
research does not provide adequate analysis of student outcomes to inform educators’ practice in this era of state and federal accountability mandates (Dover, 2009). If qualitative research on particular styles of classroom discourse has shown promise for improving student achievement, and if the current educational climate of accountability requires improvement of standardized test-scores across all high school grade levels, then more quantitative research is needed that uses standardized measures of student-achievement at the upper grade levels and provides enough detail about the nature of the discourse to allow teachers to replicate the method of classroom discourse.

**Significance of the Research Problem**

As mentioned above, no studies of classroom discourse have met all of the following criteria: used externally-designed quantitative measures of achievement, clearly defined the nature of the classroom discourse, and evaluated the effect of the discourse on high school upperclassmen (Murphy et al., 2009). This study will help fill this gap in the research. In addition, the study will provide confirming or disconfirming empirical evidence for sociocognitive learning theory, particularly regarding the effect that classroom discourse, as a social process, has on cognitive processes that influence student achievement (Langer, 1985). Finally, the study has the potential to contribute to teachers’ practice. Like many public, urban high schools across the United States, Community Leadership School—where the author has taught for eight years—faces the dual challenge of empowering historically underserved students while simultaneously augmenting standardized test scores (Delpit, 1993; Nieto, Bode, Kang, & Raible, 2008). With a heightened focus on standardized test scores in the contemporary educational environment, many teachers at CLS have responded to this complex challenge by focusing on explicit strategy instruction for test preparation as their primary pedagogical practice.
Such a practice often narrows the curriculum, inhibits students’ feelings of self-efficacy, and actually exacerbates test scores (Slomp, 2008).

Very few instructional approaches have been shown to improve students’ critical thinking, reasoning, and argumentation about text (Murphy et al., 2009; Soter et al., 2008), and the present study has the potential to confirm or disconfirm linking ideas and pressing for reasoning as discourse moves related to these advanced skills (Wolf et al., 2005). In brief, linking ideas is the process by which discourse participants connect their ideas to the ideas of other participants (also called uptake), and pressing for reasoning is the process by which discourse participants insist on hearing and understanding justifications for ideas during discussion. These discourse moves show promise for augmenting students’ thinking, despite being used much less frequently than discussion methods that focus more on teacher evaluation of students’ responses than on the links among the responses and the reasoning behind them (Cazden, 2001).

The findings of this study will be relevant to English language arts teachers who aim to improve their students’ argumentative skills and to those who focus on academic discourse in their classrooms. After all, substantiating claims with relevant evidence and warranted reasoning has always been a central project in the English language arts classroom, and with the adoption of the Common Core State standards, logical arguments are likely to remain one of the cornerstones of secondary English assessment ("Common core standards initiative," 2010). According to these standards, which have been adopted by 45 U.S. states, students will be required to engage in academic discourse with other students and with their teachers. Finally, researchers exploring the sociocognitive effects of academic discourse and investigating methods for augmenting argumentative writing scores will also find this study relevant.
Research Question, Definitions of Key Terms, and Hypothesis

As mentioned above, the challenge of improving students’ argumentative writing test scores while also focusing on reasoning and ideas instead of test preparation permeates contemporary educational culture; therefore, the purpose of this quantitative study is to investigate the effect of classroom discourse on argumentative writing skills among English language arts students in 11\textsuperscript{th} and 12\textsuperscript{th} grades in a Boston public high school. Specifically, the following research question will be examined:

**Research question.**

What is the difference in argumentative writing scores between English language arts students in 11\textsuperscript{th} and 12\textsuperscript{th} grade who engage in classroom discourse where the teacher focuses on procedural facilitation and English language arts students in 11\textsuperscript{th} and 12\textsuperscript{th} grade who engage in classroom discourse where the teacher focuses on linking ideas and pressing for reasoning?

The purpose of the research question is to compare two methods of facilitating classroom discourse: (a) procedural facilitation, which has been shown to be the predominant discursive method of teachers (Cazden, 2001), and (b) facilitation that focuses on linking ideas and pressing for reasoning, which has shown promise for improving students’ reasoning and reading comprehension (Murphy et al., 2009; Soter et al., 2008; Wolf, Crosson, & Resnick, 2005). As predicted by sociocognitive theory, a social process of discourse that focuses on linking ideas and pressing for reasoning will increase the incidences of these ways of thinking in students when they work independently (Langer, 1987; Vygotsky, 1978). Increasing students’ reasoning through the social process of discourse shows promise for improving argumentative writing skills; however, research has not been done that uses externally designed and normed measures of argumentative writing, particularly on high school upperclassmen. Therefore, the research
question will address the potential for these particular discussion styles—\textit{a} and \textit{b} above—to affect argumentative writing in high school upperclassmen as measured by standardized assessments designed by the College Board. Because no quantitative research has looked at this effect, secondary English teachers of upperclassmen have very little guidance on how well each of the methods would translate to improvements on standardized assessments of academic writing.

\textbf{Definitions of key terms.}

\textit{Classroom discourse.}

In this study, \textit{Classroom discourse} is constitutively defined as the dynamic process of interaction where students and teachers co-construct meaning from utterances that emerge from the group and that build sequentially on one another (Nystrand, 2006). As mentioned above, for the purposes of the present study, facilitation of classroom discourse can take two forms: (a) procedural facilitation of classroom discourse, where the teacher focuses almost exclusively on the logistical and behavioral components of the discussion (Cazden, 2001), and (b) facilitation where the teacher focuses on the thinking of the students by pressing them for reasoning and by helping them link ideas (Wolf et al., 2005).

\textit{Procedural facilitation.}

\textit{Procedural facilitation} of discourse is operationally defined as teachers’ controlling the logistical and behavioral components of discussion to make sure students are on task and alternating turns but without an explicit focus on students’ reasoning. For a more detailed operational definition, see Chapter 3.
**Linking ideas and pressing for reasoning.**

The second form of discourse is operationally defined as teachers’ and students’ engaging in linking ideas and pressing for reasoning without focusing particularly on the logistical or behavioral components of the discussion (see Chapter 3 for a detailed description of each method). More specifically, Wolf et al. (2005) define *linking ideas* as when the teacher or student “explicitly connects speakers’ contributions to each other and shows how ideas/positions shared during the discussion relate to each other,” for example, by revoicing and recapping students’ ideas” (p. 52). In addition, Wolf et al. (2005) define *pressing for reasoning* as when teachers or students “ask students to explain their reasoning.” (p. 53).

**Argumentative writing.**

In this study, *argumentative writing* is operationally defined as students’ test scores on argumentative writing using externally designed and normed assessments created by the College Board.

**Directional Hypothesis.**

English language arts students in 11th and 12th grade who participate in discussions where the teacher focuses on linking ideas and pressing for reasoning will have significantly higher mean argumentative writing scores than English language arts students in 11th and 12th grade who participate in discussions where the teacher focuses on procedural facilitation.

**Positionality Statement**

As a White, middle-class man working primarily with students of color who qualify for free-or-reduced lunch, I must question my position and motivation in working with historically underserved groups. Am I exacerbating historical inequities through my blindness to implicitly biased assumptions of the *other*? To at least a small degree—in order to maintain vigilance—I
must assume that the answer to this question is yes. Although I am the first member of my immediate family to graduate from college, I grew up in a family and a community where I was introduced to mainstream academic discourse without being aware that I was learning not only a way of speaking but also a way of viewing the world (Gray, 1999; Kaplan, 1966; Whorf, 1956). Often those who speak the dominant discourse, such as myself, rely heavily on this discourse without recognizing how language can influence thought (Stewart & Bennet, 1991). As long as I remain unaware of my reliance on dominant discourses to access power, I am likely to assume incorrectly that the same reality exists for others as it does for me (Delpit, 1988; Stewart & Bennet, 1991). This is particularly important given that many students with whom I work may not have implicitly learned mainstream academic discourses, and instead, I will try to teach them through facilitation of discourse in this study. I believe that, on balance, through teaching of these mainstream academic discourses, I am lending a hand to those who have been historically treated unjustly, but I must continually ask myself how I might be wrong. I must ask what data or observations I could observe about myself that would illuminate my biases and shift the trajectory of my research.

I hope to act as a bridge between the world of theory and that of practice and influence both communities to value the legitimate but distinct evidence-based practice of the other (D. C. Short & Shindell, 2009). However, given that I am operating within a scholar-practitioner framework where I am often at the center of both research and praxis, I am particularly concerned with the issue of researcher bias (Machi & McEvoy, 2009). I need to rely on my colleagues and advisors to help ensure all of my actions are performed, first and foremost, in a way that benefits students and their learning. Only secondarily should I allow my personal research interests to drive decision-making. Throughout my research, I will state explicitly any
places where researcher bias arises or where there seems to be tension between my interests and those of students.

**Sociocognitive Theory: A Theoretical Framework for Classroom Discourse**

As mentioned above, the purpose of this quantitative study is to investigate the effect of classroom discourse on argumentative writing test scores among 11th and 12th grade English language arts students in a Boston public high school. Such research requires a guiding theory that hypothesizes about the relationship between classroom discourse and student achievement.

**Overview.**

As a theoretical framework, sociocognitive theory aligns seamlessly with classroom discourse because it postulates that the social contexts of literacy classrooms encourage particular ways of thinking and behaving; in fact, these contexts and their affiliated behaviors are inextricably linked (Applebee, Langer, Nystrand, & Gamoran, 2003; Langer, 1985). Such a relationship between social interaction and cognitive development suggests that the connection between the two needs to be carefully examined in order to determine how altering the social context and the discourse within it can ultimately augment student achievement. Such research must occur within the context of a broader theoretical framework in order to guide the questions researchers ask and the methodologies they use to explore those questions (Anfara & Mertz, 2006). In this sense, sociocognitive theory is both the lens through which classroom discourse is to be viewed and the framework that will engender predictions about student achievement (Anfara & Mertz, 2006). Though this theoretical framework is informed primarily by existing literature, the framework must be refined using the experiential knowledge of the researcher in order that it is seamlessly aligned with the particular research question and methodology of this study (Maxwell, 2005).
The particular version of sociocognitive theory to be employed in this study attempts to explain why historically underserved students, specifically students of color and students in poverty, continue to struggle on argumentative writing tasks in the context of public educational systems (Kane et al., 2009; Van de Ven, 2007). By (1) stating propositions that describe relationships among constructs, (2) defining those constructs with concepts, (3) articulating variables affiliated with the constructs, (4) forming hypotheses that express relationships among these variables, (5) providing a framework around which the experimental variable can be manipulated, and (6) providing guidance for how the dependent variable can be measured, the theoretical framework (see Figure 1) challenges the commonly held notions that students struggle because of a lack of ability, that knowledge and skills are fixed entities, and that teaching is the transmission of these fixed sets of knowledge and skills (Kerlinger & Lee, 1986; Van de Ven, 2007); instead, the theory posits that a reorganization of students’ discourse in a particular social environment can create an intermediate space between students’ home discourse and the discourse of school that allows for the learning to take place that is privileged on standardized tests (Langer & Applebee, 1986). According to the theory, shifting the nature of the discourse can create a social context that allows for cognitive processes to shift and learning to occur.
Figure 1. Schematic for sociocognitive theory.
Propositions.

Sociocognitive theory, as it applies to discourse in general, articulates relationships among three interrelated propositions: (1) Social context affects the nature of the discourse that occurs within it, (2) The discourse that occurs within a social context, in turn, affects the cognitive processes of the participants in the discourse, and (3) The cognitive processes of the participants in the discourse affect their achievement. These propositions are interrelated. The object of the first proposition becomes the subject of the second, and the object of the second proposition becomes the subject of the third. Regarding the specific application of sociocognitive theory to the present study, a logical deduction from the theory is that inter-individual classroom discourse that presses students’ for their reasoning and encourages them to make connections among ideas will develop intra-individual cognitive processes where students press themselves for their own reasoning and make connections among ideas. These cognitive moves are, in turn, more likely to emerge in students’ academic work. Therefore, after participating in discourse that focuses on linking ideas and pressing for reasoning, students will improve their performance on academic tasks (such as argumentative writing) that privilege these skills.

Each of the above propositions contains two primary constructs, overlapping in such a way—as mentioned above—that the object of the first is the subject of the second, and the object of the second is the subject of the third. Therefore, there are four major constructs in this version of the theory: social context, classroom discourse, cognitive processes, and student achievement. Their relationships are shown with the propositions that construct the theory, but at the most basic level, social context affects classroom discourse of all kinds, which affects cognitive processes, which affect student achievement. More specifically for this study, facilitating
discourse that includes more instances of specific moves (such as linking ideas and pressing for reasoning) will increase the instances of such moves in students’ thinking processes, which will increase the instances of these moves in students’ academic work. Therefore, as long as these moves are privileged by the academic task, discourse that includes more instances of linking ideas and pressing for reasoning will augment student achievement over and above discourse that does not include these historically privileged moves.

*Social context* is defined in this version of the theory as including the people (particularly students and teachers), their goals and aspirations for education, their history with education, the time, the place, the culture of the teacher(s), the cultures of the students, and the dominant culture in which the interactions take place (Langer & Applebee, 1986). This definition is constitutive in that it defines the construct of social context using concepts that could actually be constructs themselves in another study (Kerlinger & Lee, 1986). In essence, this definition of social context as it relates to education articulates that students do not simply learn to read, write, and think in a cultural, ahistorical settings; instead, students learn particular things in particular ways, and those details are often governed by the social context (Langer & Applebee, 1986). It is particularly relevant to this study to note that for historically underserved students, there is often a mismatch between the home discourse styles and the discourse styles that are privileged in school (Cazden, 2001; Heath, 1983; Michaels, 1981). According to this version of sociocognitive theory, this mismatch in social context can lead to discursive interactions between teachers and students that neither affirm the value of home discourse styles nor connect the students’ home discourse styles to the privileged discourses of school. Therefore, the social context of the classroom may influence the classroom discourse in such a way that ultimately has a deleterious effect on student achievement. Conversely, the theory also posits that if classroom
discourse can affirm students’ home discourses while also furthering their facility with privileged school discourses, then the classroom discourse can interact within the social context to augment student achievement. See the Literature Review in Chapter 2 for further explication of this idea.

A constitutive definition of classroom discourse describes the dynamic, temporal process of negotiating utterances among participants—co-constructed by the participants in such a way that it can be understood only as it emerges from the interactions. In other words, the discourse is not scripted and could not be predicted. For classroom interactions to reach the level of discourse in this theory, students’ and teachers’ utterances must be sequentially contingent—that is, each statement must be built on previous statements so that it could not have come before them (Nystrand, 2006). Taken together, these classroom interactions function as instruments for reshaping students’ experiences and their learning (Barnes, 1976). Classroom discourse is more feasibly (and more ethically) manipulated than social context, so in the study—to be described in more detail below—discourse will also be operationally defined in order to specify the details of the manipulation of the independent variable (Kerlinger & Lee, 1986). See Chapter 3 for further details.

Almost impossible to observe directly in a classroom setting (Kerlinger & Lee, 1986), cognitive processes are defined constitutively in this theoretical framework by focusing on some of the most important processes that affect student literacy achievement: comprehending, understanding, questioning, hypothesizing, and connecting ideas and experiences (Langer, 1986). In essence, the construct can be defined as “extracting and constructing meaning through interaction and involvement with the written language” (Snow, 2002, pp. 11-12). Specific to this study, the cognitive processes being integrated into the discourse require students to articulate their reasoning and to link their ideas to the ideas of others.
Given the difficulty of directly observing changes in cognitive processes, the theory assumes that useful changes can be inferred through shifts in student achievement (Kerlinger & Lee, 1986). *Student achievement* is, therefore, operationally defined as corresponding—in general—to students’ standardized test scores, for this study, in argumentative writing. Such an operational definition, in the case of this version of the theory, does not address how well standardized tests assess student achievement or their underlying cognitive processes; instead, it assumes the importance of standardized test scores in an era in which they influence the degree to which students have access to systems of power (Delpit, 1988, 1992, 1993).

**Variables.**

Given the difficulty of directly measuring and manipulating social context and cognitive processes, the two measured variables in this study—as shown in Figure 1—will be classroom discourse and student achievement (Kerlinger & Lee, 1986). One of the assumptions that bounds this theory is that changes in the measured variables signal changes in the latent variables. In other words, substantive changes in the practice of classroom discourse influence the social context, and statistically significant changes in measures of student achievement signal shifts in underlying cognitive processes. An additional assumption is that even if increases in student achievement do not signal shifts in cognitive processes, the improved test scores still have value for historically underserved students because they improve those students’ life chances by facilitating greater access to scholarships, education, and careers (Brint, 2006).

As mentioned above, the dependent variable in this study is student achievement in argumentative writing, operationally defined as students’ test scores on a standardized argumentative writing task designed and normed by the College Board. The Education Testing Service creates SAT and Advanced Placement exams that influence students life chances;
therefore, score-improvements on these exams—using the questions, rubrics, and scoring scales for them—have the possibility of improving students’ lives, above and beyond shifts in cognitive processes by providing students greater access to institutions of power (Brint, 2006). Such an operational definition does not exhaust the possible meanings of the construct, but it does make measurement possible (Kerlinger & Lee, 1986).

Taken together, these conclusions provide a justification both for using a classroom discourse intervention to augment the test scores of historically underserved students and for focusing on discursive moves that show the greatest promise for improving students’ skills. Based on the two variables in this study, sociocognitive theory posits the following: If teachers and students shift the nature of their discourse in the social context of the classroom, this will affect students’ cognitive processes and, in turn, their achievement on standardized tests. More specifically, if students and teachers significantly increase the instances of linking ideas and pressing for reasoning, students’ argumentative writing will include more instances of linking ideas and articulating reasoning; therefore, test scores will increase significantly as well given that these qualities of thinking are privileged in the scoring systems for argumentative tasks created by the College Board (see, for example, "The AP English Language and Composition Exam," 2013). This hypothesis directly responds to the problem of practice and to the research question by addressing the way that students’ test scores can be augmented using particular discursive techniques.

**Conclusion.**

The specific version of sociocognitive theory described for this study clearly states propositions that show relationships among relevant constructs. Previous research using this theory helps define these constructs using constitutive definitions for latent variables and
operational definitions for measured variables. Once these variables are clearly defined, the propositions that form the theory imply a hypothesis that posits a concrete relationship between independent and dependent variables, namely, that experiencing linking and pressing facilitation will improve students’ test scores in argumentative writing. It is this hypothesis that will be explored in the study through quantitative measures of the variables in a quasi-experimental design.
Chapter 2: Literature Review

In response to the continued struggles of Black and Hispanic students in poverty on argumentative writing tasks, academic classroom discourse shows potential for shifting students’ cognitive processes in such a way that not only affirms the thinking processes that students bring to school but also augments their argumentative writing skills.

There are a variety of pedagogical practices that simultaneously address social injustices and student learning (see Dover, 2009, for discussion), but they often fail to address the realities of the current psychometric era—where in order to be found persuasive by the educational establishment, a researcher’s work must show that it can augment standardized test scores that are used to assess student learning, teacher effectiveness, and school quality. As a general pedagogical practice, academic discourse has the potential to meet both of these criteria because it can function (a) as a political and emancipatory act by encouraging active student engagement with the social, political, and economic issues that have led to the inequitable conditions of education and society (Michaels, Sohmer, & O'Connor, 2006) and (b) as an educational practice that shows promise for improving students’ test scores (O’Connor, Michaels, & Chapin, 2013; Topping & Trickey, 2007a).

The first section of this review will explore the potential for discourse to be used as an emancipatory pedagogical practice that addresses both societal inequities outside of school and achievement inequities within school. The second section will consider the discourse methods that are most promising for improving high school students’ argumentative writing. Finally, the third section will explicitly address current gaps in the research regarding classroom discourse and argumentative writing.
Section 1: Discourse as Pedagogical and Political

Using intellectual tools in social interactions with others provides a means for socialization, learning, agency, and change (Michaels et al., 2006). In fact, recent work by Mercier and Sperber (2011) synthesizes decades of psychological research to argue that reasoning evolved, not as a means of improving knowledge and seeking truth, but as a social practice to persuade. This fits with Bruer’s (1994) findings that social interaction makes thinking public and that thought, learning, and knowledge are not simply “influenced by social factors but are irreducible social phenomenon” (p. 289).

Because most children spend such a large portion of their time in school, Michaels et al. (2006) argue that the classroom provides a space unlike any other for “linguistic, cognitive, cultural, institutional, and societal forces” to interact throughout children’s lives (p. 2351). In this sense, the basic purposes of school are achieved through communication in the classroom space (Cazden, 2001), and the “primary currency of exchange” is the linguistic interaction among students and teachers—namely, discourse (Michaels et al., 2006). This pedagogical practice has the potential to be democratic because virtually all students can participate:

Regardless of children’s race, culture, or socioeconomic status, all biologically intact children have well-developed ‘ways with words’—ways of telling stories, giving accounts, providing reasons, arguments, and evidence. (Michaels, O’Connor, Hall, & Resnick, 2010, p. 37)

Almost all children use language fluently and correctly based on the rule-oriented language communities from which they come (Pinker, 1994), which allows nearly all students to engage in a public reasoning process. In addition, because language is culture-bound, “meaning, status, and power” are inextricably linked to the language that is used to express them (Michaels et al.,
In other words, discourse is necessarily political in that it expresses the values—even if implicitly—of a culture. In this way, the drama of teaching and learning becomes a microcosm of the larger drama of society as students and teachers make their thinking public and interact with each other in the classroom space (Cazden, 2001). In addition to this political component, such inter-individual interaction develops intellectual tools well beyond those that can be developed when the interactions are intra-individual (Vygotsky, 1978).

**Making meaning through discourse.**

All students and teachers are members of multiple language communities (Gee, 1989), and these language communities have distinct, describable features and norms (Hymes, 1981). Making sense of discourse—in school or out—requires participants to determine what kind of conversation is happening and how it will likely proceed (Gumperz, 1981). This determination is made using context-bound cues connected to speakers’ diction, rhythm, tone, and procedure (Gumperz, 1981). In this sense, there is a kind of speaker-listener contract (Kolln, 2007) that encourages all parties to cooperate in order to fill in the propositional content of the conversation—including subtext—based on the particular rhetorical situation (Grice, 1989). Because this internal structure of the conversation is rule-governed, it restricts the way that participants can make meaning out of it (Michaels et al., 2006). Even if participants cannot explicitly identify the rules, the rules still exist, and they limit the kind of conclusions that can be drawn from the interaction. Michaels et al. (2006) write:

> These differently patterned discourse spaces are sites for socialization and action. They engender and constrain participants with regard to what to say and how to say it (ways with words) and what it is possible to mean and do with discourse (agency) and what kind of person it is possible to be (identity). (p. 2353)
In this regard, the patterns of language-use that are allowed and encouraged in the classroom directly affect what “counts as knowledge and what occurs as learning” (Cazden, 2001, p. 3). People can learn different discourse conventions—often unconsciously—for different communities, even when these communities occur within the same culture (Gumperz & Cook-Gumperz, 1982). This implies that students can and do learn different discourse rules not only for home and school but also for different contexts within school (Erickson, 1982).

However, it also implies that discourse styles (including thought patterns and rhetorical structures) differ across, and often even within, cultures (Kaplan, 1966). This is not simply a matter of varying syntax and diction across different language communities; instead, ways of thinking and constructing arguments are inextricably tied to language and culture (Stewart & Bennet, 1991). In this sense, non-native speakers of a particular discourse may violate the implicit contract between listener and speaker by using unfamiliar or unaccepted rhetorical modes, which are in some ways independent from syntactical rules (Kaplan, 1966).

This does not go so far as to say that language and discourse determine the way people understand reality (Whorf, 1956). Although language and discourse do influence cognitive processes in important ways, such as the way people classify and remember, other processes are better understood as nearly universal, such as the way people perceive (Gray, 1999; Honderich, 1995). Nevertheless, language, thought, and even perception are intertwined in ways that are often invisible to those who interact in the mainstream discourse—in particular those who interact in the mainstream academic discourse in the United States, which has been historically dominated by White, middle-class men (Stewart & Bennet, 1991). The distinction between mainstream academic discourse and various alternative discourses is somewhat artificial because of the ever changing landscape of academia, which includes various discourses interacting and
reciprocally influencing each other; nevertheless, it is important to note that many “American
speakers of English,” including many teachers, assume—even if implicitly—that their discourse
is simply a mechanism for expressing universal reasoning and thought processes (Stewart &
Bennet, 1991, p. 46). In this way, reality and the way that humans think about reality can appear
universal and thus independent from language, when in actuality language and its use in
discourse create a kind of “master framework” for viewing the world in the context of specific
communities (Stewart & Bennet, 1991, p. 48).

Students form socially-situated identities based on the discourse patterns of these
communities as they align their language (or choose not to align it) with the dominant patterns of
interaction, thinking, valuing, and feeling within that community (Gee, 1989). In this way,
discourse styles that are learned at home can either be reinforced or discouraged at school, and
those discourse styles learned at school can either be reinforced or discouraged at home (Heath,
1983). Because the discourse patterns of school often echo those that are deemed normative in
society at large, the primary discourse patterns of “mainstream” students usually prepare them
well for the literacy-related tasks and performances valued in school (Michaels et al., 2006).
Conversely, the primary discourse patterns of students who have been historically underserved
by education are often devalued in school. In this sense, even though students may enter school
at the same age, only the students from non-mainstream language communities are truly
beginners (Gee, 1996).

**When discourse fails.**

As suggested above, students entering school already socialized into mainstream
discourse communities are more likely to have their thinking affirmed by their teachers, and as
this interaction with their teachers deepens, so too does the students’ knowledge and skill-set. In
this light, school failure can be seen as a breakdown in the dialogical collaboration between teacher and student (Grice, 1989). A mismatched set of discourse expectations often leads to breakdowns in understanding and trust (Grice, 1989; Stewart & Bennet, 1991). In her qualitative study of sharing-time at the primary level, Cazden (2001) found that incongruous assumptions about dialectical collaboration led to tension between a “child’s intended meaning and the teacher’s valued form [of communication]” (p. 27).

Of course, school failure for historically underserved groups is much more complex than just a breakdown in discourse expectations (see, for example, Ogbu, 1994), but as Cazden (1986) and Michaels (1981) show, non-mainstream students do bring alternate discursive styles to school that are regularly perceived as deficits, often leading schools to assign these students to remedial classes. When teachers and students share expectations about how to interact, their discourse serves as a foundation for deepening their relationship, as well as students’ reasoning and literacy skills. The converse is also true when teachers and students do not share these expectations. In this way, language plays a role in identity development not only in society at large but more specifically for teachers and students in school, thus influencing the school behaviors of both parties. As students’ school-behaviors shift based on tensions arising from incompatible discourse styles, the deficits that teachers initially perceive can become real (Michaels et al., 2006), particularly because assessments of student achievement are highly context-specific and depend on the shared background knowledge of the speakers (Resnick, Salmon, Zeitz, Wathen, & Holowchak, 1993). In their mixed-methods study, Resnick et al. (1993) found that what can seem like inconsistent or even incoherent reasoning to an outsider can actually be a sophisticated multi-step argument when the discursive style and background information are shared. This suggests that academic discourse may be a particularly promising
instructional intervention for students whose home discourse does not match the discourse of school, even if the apparent quality of the students’ talk does not immediately indicate that they are ready for challenging academic discourse.

**Teaching discourse styles.**

Given that (a) particular discursive styles are privileged both informally in classrooms and formally in the way that classroom-learning is assessed and (b) that students can learn these discursive styles, schools have an obligation to teach these styles, particularly to historically underserved students (Delpit, 1992). Delpit (1993) writes that there is a strong, well-meaning but misguided, trend among teachers who believe that “their role must be to empower and politicize their most disenfranchised students by refusing to teach…dominant Discourses” (p. 291). There are several reasons that Delpit (1993) argues against this: (1) “members of society need access to dominant Discourses to (legally) have access to economic power;” (2) “such Discourses can be and have been acquired in classrooms,” and (3) “individuals have the ability to transform dominant Discourses for liberatory purposes” (p. 291-292).

However, as Michaels et al. (2010) show and as would be expected from the above discussion, all classroom talk does not lead to this type of learning—it matters what students and teachers are talking about and how they are talking about it. Focusing, even if explicitly, on the superficial features of a dominant discourse pattern will not empower students or help them grow if these features are not presented in rich, meaningful contexts that allow students to interact with both each other and the teacher using some mix of their discourse styles and the teacher’s (Michaels et al., 2010). Students often face challenges in adapting their home discourse to the discourse of school in cases where the two do not match. After all, school discourse may represent both a social category and a cultural symbol for students (Basso, 1979), and students
are sometimes concerned that in taking on the discourse of school, they may be devaluing, or even abandoning, their home discourse and identity (Ogbu, 1994).

In this sense, although imitation and eventual integration of the discourses of power may be necessary for students’ academic success, students’ identity development can be hampered if teachers frame the discourse of school as superior to students’ other discourses (Delpit, 1988). In fact, the problem is not that there is a mismatch between students’ home discourse and the discourse of school, but that institutions (including schools and teachers) often frame school discourse as normative and other discourses as inferior. Instead, teachers must help students incorporate the discourse of school—at least initially—as a kind of play where they can express concrete linguistic formulations of an abstract discourse without being asked to take on an identity that demeans their home discourse (Basso, 1979). This kind of play contains something of the desired behavior that teachers want from students (namely, “academic” discourse moves) while also including a kind of meta-statement that signals to the students’ classmates to interpret their behavior somewhat ironically (Basso, 1979).

Instead of creating this intermediary discourse space between home and school, most classroom discourses facilitated by the teacher are more simplistic. For example, Mehan (1979) showed that the dominant discourse pattern in classrooms begins with an initiation by the teacher (I), which is followed by a response by a student (R) and an evaluation by the teacher (E). In addition, Cazden (2001) has shown that this pattern has persisted for decades. Work by Heath (1978) with African-American students interacting with White teachers indicated that these teachers were more likely to ask students close-ended questions, such as “What is that?”, instead of open-ended questions similar to those asked in students’ home communities, such as “What is that like?” Close-ended questions, privileged by the school-community, were often being
ignored by African-American students because in their home language-communities, such questions were not answered when the response was obvious. From the perspective of the teachers, ignoring the questions was a sign of ignorance; when in reality, it was a sign of knowledge.

Work by Cazden (1986) demonstrated that this IRE discourse pattern almost always dominates unless an intentional intervention occurs. IRE is often necessary and appropriate, but when it becomes the dominant interactional pattern, students often get the implicit message that the curriculum is a set of facts, known by the teacher, and to be transmitted to the students (Edwards & Westgate, 1994; Freire, 1970). Because the teacher is speaking for two of these three discourse moves, the teacher’s talk-time ends up accounting for the vast majority of classroom time, and as the teacher exercises this “right” to dominate the discussion, non-mainstream discourse patterns are delegitimized (Cazden, 1986). In essence, the teacher becomes both the stage director and the principal actor in the classroom, relegating students to minor roles, even though the teacher is dependent on the students to shift the academic culture toward authentic learning and growth and away from rote banking (Cazden, 2001; Freire, 1970).

Despite these challenges, intentional interventions have occurred showing that discourse styles can be mixed to improve learning for historically underserved students. In fact, a strong body of research has revealed that students’ home discourse patterns can be harnessed at school in a variety of settings (Heath, 1978, 1983; C. D. Lee, 1993; Phillips, 1992; Piestrup, 1973). This work provides evidence that teachers can adjust pedagogical discourse practices to augment student learning. The next section will explore the most promising of these discursive methods for improving achievement.
Section 2: Promising Discourse Methods for Augmenting Student Learning

As has been indicated in the previous section, learning is a social phenomenon, and socialization often occurs through the discursive patterns that transpire in school. These discourses are necessarily political insofar as they privilege particular forms of interaction and, therefore, define what counts as learning and worthwhile knowledge. Different discourses prevail in different communities, and these discourses can be reinforced or devalued within the predominant discourse of school—which tends to privilege the mainstream discourse that is valued on standardized tests. When students’ home discourse does not match the discourse of school, teachers often struggle to turn classroom interactions into literacy experiences for these students. In this sense, school failure can be seen, at least partially, as a mismatch between discourse expectations across communities. As students have fewer literacy experiences in school, the perceived gaps in knowledge and skills can become real. However, discourse styles can be learned, thus providing promise that teachers can create a discursive space for students where their home discourse is affirmed, and they learn to interact in the discourse of the dominant culture, if they so choose. This section will review the literature on the most promising discursive methods for improving student achievement.

Discourse and student engagement.

There is both qualitative and quantitative evidence in socio-economically and linguistically diverse settings that teachers’ pedagogical choices influence students’ productive disciplinary engagement. In particular, engagement with subject matter seems to improve as teachers become more willing to problematize content through presentation of authentic controversies that can be discussed by students (Engle & Conant, 2002). Engaging in discursive practices of this type has led middle-school social studies students to report a stronger sense of
classroom community and deeper understanding of the material (Howell, Thomas, & Ardasheva, 2011). Additionally, D. Kuhn and Crowell (2011) have found—in their experimental, longitudinal study of ethnically diverse middle school students—that argumentative reasoning skills could be identified, assessed, and improved, but this improvement occurred only as part of specific discourse interventions that targeted these skills. Specific discourse interventions that encourage students to consider multiple perspectives and refer to evidence and reasoning to substantiate claims have led to these features becoming more prevalent in the argumentative writing of urban elementary school students, middle school students, and community college students (D. Kuhn et al., 1997; D. Kuhn & Udell, 2003; Reznitskaya et al., 2001). Nevertheless, these studies have used researcher-designed evaluation systems to assess the quality of the argument. This technique of assessing the effectiveness of discourse proves less useful to educators and their students who are often evaluated on the quality of their test scores, not internally-designed formative assessments.

**Discourse’s effect on test scores.**

Addressing the gap between researcher-designed and externally-designed assessments of student learning, Mercer, Wegerif, and Dawes (1999) found that after engaging in discourse in an attempt to solve a problem collaboratively, 9 and 10 year-old British children made greater gains in their reasoning test scores than did the control group, which received no intervention. In a similar assessment of the effect of collaborative problem-solving through discourse, 10 year-old British students attending an ethnically homogeneous but economically diverse school improved their scores on a cognitive abilities multiple choice test, as compared to the control group (Topping & Trickey, 2007a). These gains were not related to gender or class differences. In addition, students in the middle quartiles of pre-test ability gained more than the highest and
lowest quartiles, with the highest quartile gaining the least. In a follow up assessment performed two years after receiving the intervention, students in the treatment group continued to show higher scores on a multiple choice test of cognitive ability than the control group (Topping & Trickey, 2007b). However, the students who had initially scored in the highest quartile and shown the least improvement from pre-test to post-test were most likely to maintain the gains at the two-year follow up that they had made during the initial intervention.

Building on the work by Topping and Trickey (2007a, 2007b), O'Connor et al. (2013) found that after implementing an Accountable Talk intervention that focused on keeping students accountable to their learning community, to accurate knowledge, and to rigorous standards for thinking, students involved in the intervention outperformed students who did not receive the intervention on the Massachusetts Comprehensive Assessment System, which is used by the state to measure students’ progress toward state and federal accountability mandates. Although the intervention occurred for one-hour per day in students’ math classes in an ethnically and economically diverse elementary and middle schools, students’ summative assessment scores improved not only in math but also in English. In addition, students who received an Accountable Talk intervention that lasted only two to three class periods showed greater gains from pre- to post-tests than when they received direct instruction without a focus on classroom discourse (O'Connor et al., 2013). Therefore,

**Discourse and argumentative writing at the secondary and post-secondary levels.**

Though the work by O'Connor et al. (2013) shows promise for informing teachers’ practice, it was done in math classes and at the elementary and middle school levels, thus limiting its applicability for high school English teachers. Addressing older students and their argumentative skills in their study of both community-college and middle-school students in an
ethnically and economically diverse city, Felton and Kuhn (2001) found that after engaging in five dialogues over the course of six weeks, the adults in the study responded more strategically to argumentative tasks than the adolescents, particularly in their use of more complex argumentative moves such as counterargument. In addition, the adults were more likely than the adolescents to adapt the discursive strategies they employed depending on the context, namely, the person and the argument to which they were responding. This suggests that participants in discourse at different ages respond to that discourse in slightly different ways and, therefore, that the effect of discourse on students of different ages should be explored, particularly at the high school level given that this study included middle school and community college students.

In their study of 20 diverse middle and high schools, Applebee et al. (2003) found that after controlling for literacy levels, gender, SES, and race, students who participated in classes that used discussion-based approaches scored higher on a spring assessment of literacy. Although Applebee et al. (2003) attempt to assess the effect of discourse on secondary English classrooms, they used an assessment and a scoring system that is not broadly used to assess students and teachers outside of the schools in the study, thus limiting its broader usefulness to English teachers in this era of state and federal accountability mandates. In addition, they only had one 12th-grade classroom in their study and no 11th-grade classrooms. Finally, most of teachers in the classrooms studied by Applebee et al. (2003) taught with a traditional interaction method where the teacher did much of the talking. This means that there was relatively little variability in the amount of discourse being used from class to class. The study’s results may seem to indicate that more discourse is better, but there were not enough classes devoting large amounts of time to discourse in order to determine whether this inference is warranted. Also, no distinction was made among different kinds of discourse that occurred in the classroom—
example, the difference between teacher-led discourse with a focus on evidence and reasoning and student-led discourse with an emphasis on student exploration.

**The influence of teachers’ discursive choices on students’ discourse.**

The above research has provided evidence that in classroom discourse, teachers’ choices regarding authenticity and student involvement influence student engagement, reasoning scores, summative assessment scores, and the overall sophistication of students’ arguments. In addition, there seem to be some differences in the ways that students of different ages process discourse, suggesting that the effect of discourse should be explored for all age groups. In high school, little quantitative work has been done on 11th and 12th graders, particularly examining the effect of classroom discourse on achievement as measured by commercially available assessments of literacy. However, in their qualitative study of a high school English literature course in an ethnically and economically diverse district, Rex and McEachen (1999) found that all students learned to apply high standards for academic discourse, but the discourse choices made by the teacher played the most important role in shaping the discussion.

Providing further evidence for this conclusion from their mixed-methods case study of elementary school students in an urban district, Webb et al. (2008) found that the ways in which teachers pushed their students’ thinking during whole-class discussion corresponded to the ways in which those students pushed each others’ thinking during small-group discussion. The teachers who most consistently pushed students thinking, whether or not students expressed correct answers, also had students who most consistently expressed their reasoning in small-groups. The teachers who did this the least had students who expressed their reasoning the least. Finally, teachers’ asking students to elaborate consistently was associated with higher levels of reasoning.
Similarly, Wolf et al. (2005) found in their mixed-methods study of 21 separate classrooms in 10 schools across 3 districts that the use of Accountable Talk discourse moves was strongly associated with the level of academic rigor in the class. More frequent use of these moves (holding students accountable to accurate knowledge and rigorous thinking) by the teacher was also correlated with more frequent use of the same moves by the students, implying that the way the discourse is facilitated contributes to its quality, rigor, and level of engagement. These studies suggest that the teacher takes the lead in developing the nature of the discourse that occurs in a classroom. If that discourse is to focus on the reasoning skills that are privileged on standardized tests, then students are more likely to exhibit these skills if they are modeled by the teacher.

Although Rex and McEachen (1999) performed their study in an English high school classroom, this qualitative study did not assess whether or not changes in student-discourse led to changes in student test scores. The same issue arises in Webb et al. (2008) and Wolf et al. (2005); additionally, these studies focused on students’ math skills and reading skills in elementary and middle schools, respectively, not their argumentative writing skills in high school. Nevertheless, these studies do suggest that the discourse moves modeled by the teacher are frequently the discourse moves displayed by the students.

**The effects of different discursive styles on student achievement.**

Similar to the above studies, Nystrand and Gamoran (1991) found in their mixed-methods study of 58 middle school classes in 16 Midwestern schools that different types of engagement through discourse are related to achievement in different ways. Procedural engagement (with school processes such as following directions) was more highly correlated with student achievement in literature than disengagement but not as highly correlated with
achievement as substantive engagement (with reasoning about content through discourse).
Achievement was not simply related to engagement; it depended on what students were engaged with. Higher achievement was associated with classes where there were more connections among ideas during discussion (uptake), more open-ended questions, more cohesion from lesson to lesson, and more substantive discussion time.

Consistent with Nystrand and Gamoran (1991), Murphy et al. (2009) found in their meta-analysis of the effects of classroom discussion on students’ comprehension that many discussion approaches were effective at augmenting literal and inferential comprehension, but very few approaches improved students’ critical thinking, reasoning, and argumentation about text. Most discursive approaches did increase student talk and decrease teacher talk, but increasing student talk-time did not proportionally improve students’ comprehension or reasoning. Their findings also suggest that discourse-based instruction seems to have larger effects for students who are testing below average in terms of their literacy ability, although—as mentioned previously—above average students may be more likely to maintain the gains they do achieve through discourse-based instruction (Topping & Trickey, 2007b).

**Promising discursive techniques.**

Heretofore, this review has cited research that provides evidence for the following: For discursive moves to be used by students, they generally need to be modeled by the teacher before students are given opportunities to integrate these moves into their talk. However, different types of moves lead to different types of engagement and achievement. Substantive engagement with the content is most promising for augmenting student achievement, particularly in critical thinking, reasoning, and argumentation. This substantive engagement is most likely achieved through discourse where teachers and students incorporate others’ ideas into their propositions,
ask open-ended questions, provide cohesion from lesson to lesson, and allow for substantive discussion time. These elements of a discourse-based pedagogy are most hopeful for augmenting the achievement of students who score below average in terms of their literacy achievement.

Taking the cited research into consideration, a discourse method needs to be tested at the high school level (particularly in 11th and 12th grades) that employs assessments that are often used to measure the learning of students and the effectiveness of teachers (for example, standardized tests designed and normed by the Educational Testing Service). This discursive method would seem to require teachers to model discursive moves for their students that focus on linking ideas and pressing students for reasoning through open-ended questioning. In addition, more classroom time needs to be provided for students to express their reasoning while simultaneously allowing the teacher space to model thinking processes for students. In order to be most useful to practitioners, specific discourse moves that meet the above criteria need to be explored.

Along these lines, O'Connor and Michaels (1993) explored revoicing as just such a discourse facilitation strategy. They found in their qualitative case study that teachers’ and students’ revoicing of other students’ reasoning provides more space for students to reason by providing them with consistent opportunities to respond to open-ended questions about their reasoning. Instead of initiation by the teacher, response by the student, and evaluation by the teacher (the IRE pattern discussed above), revoicing encourages students to respond to the teacher’s (or another student’s) understanding of the initial student’s thoughts. From a socializing perspective, revoicing can also allow students to take credit for others’ substantiated inferences based on the original student’s thinking. The teacher (or another student) can revoice
a student comment (“If I hear you right, you seem to be saying X. Is that correct?”), focusing on and extending a particular part of it before asking the student whether or not the audience has understood correctly. This can help affirm the thinking of students who might be peripheral participants in the discussion. Revoicing can also serve to reframe and thus scaffold reasoning for students who are struggling to understand while simultaneously allowing more time for students to consider the same idea while it is being rephrased in other words (O'Connor et al., 2013).

In addition to revoicing as a promising, specific talk move that creates space for substantive engagement with the content, Soter et al. (2008) provide additional support for the use of authentic questions, linking ideas across the class, and including a high-density of reasoning words as likely candidates to help improve the quality of classroom discourse. The researchers coded 36 transcripts across 9 different discussion styles in an attempt to determine which types of discourse produced the most characteristics of high-level thinking and comprehension. All of the nine discussion styles had shown empirical evidence to improve students’ comprehension. The researchers grouped them into three categories: (1) expressive approaches, where the primary focus of discourse was on students’ responses to the text, (2) efferent approaches, where the primary focus was on the text itself, and (3) critical-analytical approaches where the primary focus was on interpreting and evaluating the text. Their results indicated as follows:

The most productive discussions (whether peer or teacher-led) are structured, focused, occur when students hold the floor for extended periods of time, when students are prompted to discuss texts through open-ended or authentic questions, and when
discussion incorporate a high degree of uptake [namely, linking others’ ideas to their own]. (p. 373)

The authors also found—consistent with O’Connor and Michaels (1993)—that authentic questions tend to elicit more elaborate student responses, thus giving students more space to produce deep reasoning and high-level thinking. Although affective connections with the material were necessary to engage students, the deepest reasoning consistently appeared in the critical-analytical approaches, suggesting that authentic questions, linking ideas across the class, the density of reasoning words (such as if and because), and the space for elaborated explanations are the strongest indicators of productive discourse (Soter et al., 2008). These results imply that the most productive discussions balance the control of discourse between teacher and student so that high-levels of engagement about the material can occur among students while also giving the teacher space to model and prompt the use of evidence and reasoning to substantiate claims. Despite its strengths, this study did not measure student achievement directly, but instead used empirically and theoretically supported indicators of high-level thinking in order to assess the quality of discourse. In addition, the transcripts only included students in grades 3 through 9.

Consistent with Soter et al. (2008), Nystrand, Gamoran, Kachur, and Prendergast (1997) found that the cognitive level of the teacher’s instruction did not predict student achievement in their study of Midwestern English classrooms; instead, they found that authentic questions and linking ideas across the class correlated more highly with student achievement. Similar to Cazden (2001), Nystrand et al. (1997) found that instead of focusing on dialogue between teacher and students or among students, the vast majority of instructional time was focused on teacher monologue, with only 12% of questions being coded as authentic (where the teacher did
not already know the answer) and 11% of comments including *uptake* (where the comment linked previously expressed ideas to new ones in such a way that the discourse was sequentially contingent). In addition, in the lower tracked classrooms (similar to those at the research site for this study), teachers’ instructional method tended to focus even less on students’ thinking and the content of their writing, and even more on direct instruction and procedural elements of their work (such as grammar and spelling).

These studies suggest that improvements in critical thinking, reasoning, and argumentation are most likely when teachers are modeling and allowing students space to practice revoicing of others’ ideas, linking the ideas of others’ to their own, asking authentic questions, and using a high density of reasoning words. The question remains whether or not teachers can learn these discursive techniques in such a way that benefits their students.

In their mixed-methods case study of 64 eight and nine-year-old children in the United Kingdom, Wegerif, Mercer, and Dawes (1999) explored the extent to which a discourse method established as effective for improving student reasoning in one school could be transferred to new schools. The authors found that, in the two additional schools, students improved their reasoning abilities concomitant with the extent to which their teachers engaged in the practices associated with the discourse method used in the first school. The study suggests that discourse methodologies can be taught to both teachers and students; however, if there is no observable shift in teachers’ instructional practice, there likely will not be a shift in student achievement.

This section of the review has argued that discourse that focuses on revoicing, asking authentic questions, integrating others’ ideas, and including a high-density of reasoning words is most likely to augment substantive engagement with the content, sophistication of reasoned arguments, and student achievement as measured by standardized tests. In addition, there is
some evidence that this kind of discursive method can be taught to teachers and applied in new contexts. The following section will explicitly address gaps in the literature.

**Section 3: Gaps in the Literature**

As mentioned above, research on discourse methodologies has shown promise for augmenting students’ persuasive argumentation; however, this research has often used researcher-designed assessments of student-learning instead of more independent assessments, failed to assess teachers’ fidelity to the discursive method, and excluded analysis of the effect of discourse on different high school grade-levels, particularly 11th and 12th grades. Such research does not provide adequate analysis of student outcomes to inform educators’ practice in this era of state and federal accountability mandates (Dover, 2009). If research on particular discursive techniques has shown promise for improving student achievement, and if the current educational climate of accountability requires improvement of standardized test-scores, then more quantitative research is required that uses standardized measures of individual student-achievement and practical assessments of teacher-fidelity to the discursive method in order to assess the effect of these discursive techniques on high school upperclassmen.

**Review of theoretical framework.**

Virtually all of the studies in this review have employed a sociocognitive theoretical framework that posits the social nature of learning and the interactive nature of effective teaching (Langer, 1985). In addition, as mentioned in Chapter 1, sociocognitive theory postulates that specific learning-contexts influence behavior and that these behaviors affect the meaning that students make out of their education (Langer, 1985). This theory has prompted researchers to explore classroom contexts that encourage students to interact in communities of practice that are guided, but not dominated, by the expertise of the teacher (Lave & Wenger, 1991). In this sense,
sociocognitive theory builds on Vygotsky’s (1978) conception of *inter-individual* interaction among students serving as a kind of dress rehearsal for *intra-individual* thinking processes characteristic of full participants in a community of practice (Lave & Wenger, 1991).

This theoretical framework is more constructivist than positivist in nature, making the use of quantitative measures within these studies somewhat incongruent (Ponterotto, 2005); however, given the pressure on many educators to meet state and federal accountability standards that are quantitative (Dover, 2009), mixed-methods designs—as employed in the majority of studies in this review—become more coherent: Educators and educational researchers may believe that students and teachers are co-constructing knowledge and that the best kind of evidence for this is qualitative, but if they cannot show that pedagogies built on sociocognitive theory augment test scores, their ideas will not gain purchase on the current psychometric landscape (Dover, 2009). In this sense, the most promising theory for investigating classroom discourse (as articulated in Chapter 1) remains sociocognitive, but the most promising research designs—for practical and political reasons—are shifting from a more qualitative to a more quantitative focus.

**Review of methods.**

As mentioned above, early explorations of the effectiveness of classroom discourse tended to be qualitative: Heath (1978, 1983) and Michaels (1981) are classic examples of such studies. In the 1990’s, mixed-methods work started to take root (D. Kuhn et al., 1997; Mercer et al., 1999; Nystrand & Gamoran, 1991; Resnick et al., 1993; Sandora, Beck, & McKeown, 1999; Wegerif et al., 1999), though some purely qualitative work remained (O’Connor & Michaels, 1993; Rex & McEachen, 1999). Since 2000, all of the studies directly relevant to this review have included a quantitative component (Engle & Conant, 2002; Felton & Kuhn, 2001; Howell
et al., 2011; D. Kuhn & Udell, 2003; Murphy et al., 2009; O'Connor et al., 2013; Reznitskaya et al., 2001; Soter et al., 2008; Webb et al., 2008; Wolf et al., 2005). In addition, all of the purely quantitative studies have been published since 2003, with six of the seven being published since 2007 (Applebee et al., 2003; D. Kuhn & Crowell, 2011; Murphy et al., 2009; O'Connor et al., 2013; Soter et al., 2008; Topping & Trickey, 2007a, 2007b).

This suggests a number of possible trends: (1) As sociocognitive work on discourse has progressed, the focus has shifted from exploring possibilities to explaining outcomes. This may simply be a common trend as support for a theory becomes more robust (Fraenkel, Wallen, & Hyun, 2012). (2) A broader cultural shift within education seems to have moved toward the legitimacy of positivist explanations over constructivist-interpretivist ones, shifting what kinds of research gain traction (Apple, 2008; Franklin & Johnson, 2008; Levin, 2008; Welner & Oakes, 2008). Assuming that both of these explanations might play a role in methodological shifts around research on discourse, neither of these explanations indicates a reversal of this trend in the near future (Murphy et al., 2009). Therefore, in the coming decade, research that investigates relationships between discourse and learning is likely to maintain its quantitative component.

**Review of instruments.**

Despite Resnick et al.’s (1993) ambitious attempt to develop a broadly applicable instrument for describing how students reason cooperatively in social settings, their instrument has not gained acceptance, as it was not used again by the prominent names in the field in the succeeding decades. Instead, the vast majority of studies have used measurements of student growth developed by the researchers or participants involved in the study, posing a significant threat to validity (Fraenkel et al., 2012). Mercer et al. (1999), O'Connor et al. (2013), and Topping and Trickey (2007a, 2007b) are notable exceptions, as they used externally designed
and normed measurement tools. In addition, though most researchers at least attempted to clarify their measurement of student growth, very few studies strove to assess the educators’ fidelity to the particular discursive method under consideration. Applebee et al. (2003) and Wolf et al. (2005) are notable exceptions, as they assessed fidelity using analytical rubrics.

Though the efforts of researchers in this field are admirable, looking at their work as a whole points to the need for externally-designed, well-accepted measures of student achievement, in addition to clear descriptions of the discursive method that allow for replication by practicing teachers and researchers. In their meta-analysis of classroom discourse methods, Murphy et al. (2009) indicate the importance of externally designed measures of student achievement, for even though researchers were consistently holding themselves to high standards for interrater reliability, they often designed tools that measured the very thing that was taught in the discursive intervention, limiting their conclusions about transfer of skills across contexts. For example, in Reznitskaya et al. (2001), educators taught students specific argumentative strategies such as counterargument and refutation, and then researchers assessed the quality of students’ essays on whether or not they included such argumentative strategies. Though this kind of work does help to show that particular discursive methods do actually help students learn skills that can be transferred to writing, it needs to go a step further and use commonly used measures of student achievement to show that these strategies actually have a broader effect on student learning.

**Review of data sources and data problems.**

With the wide proliferation of audio and video recording technology, researchers can gather reliable raw data on discourse with relative ease. In addition, word processing programs with search functions improve the accuracy and reliability of coding. If the current trend in
research on discourse were qualitative, with such data accepted as valid by educational leaders and policy makers, many of the challenges surrounding data collection would be relatively minor; however, attempts to quantify discursive interactions and what is learned from them prove problematic.

Assuming that the ultimate goal of this area of research is to show relationships between discourse-moves and student-learning, the first problem is in how to collect data that accurately assess student learning. Given the dominance of positivist standards for evidence in contemporary policy-making (Kelly, 2009; Levin, 2008; E. C. Short, 2008), challenges to assessment validity can be adequately resolved in the short term by using measurement tools that are broadly accepted as proxies for student achievement. For example, standardized tests results on assessments developed by the Educational Testing Service (ETS) act as gatekeepers for colleges and universities. The discussion of whether or not this should be the case is secondary for the current generation of students who face this reality; therefore, if discourse methodologies can show improvements on externally-designed and normed assessments that help students access systems of power, those discursive techniques have value—particularly given that qualitative data already support the more authentic educational value of discourse.

The second problem associated with data collection is that researchers need a way to empirically measure the discourse methodology of the educator if they are to show that it influences student test scores. A small number of researchers have attempted to develop (Wolf et al., 2005) or use (Applebee et al., 2003) rigorous tools to measure teacher behavior, but most researchers have relied on teachers’ attending professional development on the relevant discursive method as a marker of fidelity to that method; however, in any professional development, teachers will vary widely in their interpretation and implementation of methods
(Elmore & Burney, 1999). This poses significant problems for collecting data about how an individual teacher implements discourse.

Theoretically, taping, transcribing, and recording discourse and representing this discourse quantitatively shows great promise for articulating the relationship between discourse and student achievement. O'Connor et al. (2013) used this method to excellent effect, as they code, count, and report discourse moves in their quantitative study. However, this method is not practical for everyday classroom teachers to assess their own discursive moves. Instead, a compromise needs to be developed that allows teachers to assess their own fidelity to a discursive method in such a way that has been shown to be accurate and reliable enough to improve student achievement. Though a full development of this compromise is beyond the scope of this study, Chapter 3 will provide some recommendations for an assessment of the discursive method that can potentially be connected to student achievement.

**Review of research on grade-levels.**

Although research on discourse with a quantitative component has been done at the elementary, middle, high, and post-secondary levels, no research has been done on the effect of discussion on argumentative writing skills among 11th- and 12th-graders (see Table 1), and this study aims to fill this gap in the research.
Table 1

Research on Discourse by Grade-Level and Topic

<table>
<thead>
<tr>
<th>Level</th>
<th>Topic</th>
<th>Argumentative Skills</th>
<th>Reasoning Skills</th>
<th>Comprehension Skills</th>
<th>General Literacy Achievement</th>
<th>Engagement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Middle</td>
<td>Felton and Kuhn (2001); D. Kuhn and Crowell (2011); D. Kuhn et al. (1997); D. Kuhn et al. (1997); Topping and Trickey (2007a, 2007b)</td>
<td>Wolf et al. (2005); Sandora et al. (1999); Applebee et al. (2003); Soter et al. (2008)</td>
<td>Nystrand and Gamoran (1991); O'Connor et al. (2013)</td>
<td>Nystrand and Gamoran (1991)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post-Secondary</td>
<td>Felton and Kuhn (2001); D. Kuhn et al. (1997)</td>
<td>Resnick et al. (1993)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Areas for Further Research

Future research on discourse and learning will be built on (1) a foundation of sociocognitive theory (Langer, 1985; Vygotsky, 1978), (2) qualitative research that has explored the possibilities for discourse to influence student learning (Michaels, 1981), (3) mixed-methods research that has narrowed the focus to the relationship between particular discourse methods.
and student learning (Soter et al., 2008; Wolf et al., 2005), and (4) quantitative research that has applied more systematic measures of both teacher-behavior and student-achievement (Applebee et al., 2003; O'Connor et al., 2013; Topping & Trickey, 2007a, 2007b). Within this context, further research needs to acknowledge the political reality of this psychometric era—that educators are accountable to state and federal mandates for student learning, as measured by quantitative assessments. Qualitative and mixed-methods research has already shown that engaging students in discourse around controversial topics generally improves their inferential, argumentative, and comprehension skills (Murphy et al., 2009), but very little work has been done that addresses whether or not similar methodologies augment students’ standardized test scores, particularly in argumentative writing at the upper high school level. The use of broadly adopted standardized measures of student achievement (such as the SAT and AP exams) within studies on argumentation would test the degree to which authentic argumentative skills transfer to improvements in test scores. In addition, the large amounts of data that are already available for such assessments make it possible to compare student pretest and posttest performance to the global mean and standard deviation throughout the country, minimizing threats to validity that arise due to differences in the assessments themselves.

For such studies to make clear and useful connections between discourse methods and student learning, researchers must develop more straightforward assessments of the fidelity to which an educator holds to a discourse method. Though Wolf et al. (2005) attempted to use rubrics to measure this, the method was too cumbersome and not subject to ready tests of reliability. The questions remain, what would robust fidelity data look like, and is there a way to quantify this without returning to intractable philosophical problems about the validity of using quantitative methods to assess qualitative phenomena? Of course, this quantification simplifies
the fluid facilitation of student talk and potentially occludes the real variable affecting student achievement; however, to take a pragmatic stance, such quantification could be the beginning of broader implementation of loosely codified discourse methodologies, and thus even further research on their effectiveness.

**Conclusion**

In a review of literature on teaching for social justice Dover (2009) writes:

While these resources provide essential curricular, philosophic, and pedagogical information about the scope of teachers’ attempts to teach for social justice, they fail to provide sufficient analysis of student outcomes to meet the needs of teachers and researchers constrained by state and federal accountability mandates. (p. 518)

The research in academic discourse in the urban English classroom is subject to a similar critique. Further, in the current era of psychometrics, researchers must provide not only qualitative data that discourse pedagogies raise students’ critical consciousness but also quantitative data that these pedagogies provide students with the reasoning skills necessary to access systems of power.

Darder (1991) argues that results of standardized testing seem valid to society at large because they appear to be value-free scientific tools. In fact, all forms of reasoning, scientific methods, and pedagogies are bound to their particular time and place (Bruner, 1996; hooks, 1994; T. Kuhn, 1996). As long as reason and language enjoy relative value-freedom, they remain insulated from criticism because they disqualify competitors without having to engage them. When this happens, reasoning and language become tools of oppression (Darder, Baltodano, & Torres, 2003; Love, 2004). Twenty-first century urban English teachers must, therefore, teach and assess language and reasoning (in its privileged forms) as codes of power
(Delpit, 1988) to help students access systems of power through academic “achievement” while simultaneously providing them with tools to critique the dominant culture in its own terms through discourse. As Sledd (1996) suggests reasoning and critical analysis need to be taught as potential weapons that if not controlled can oppress. In this way, historically underserved groups move toward emancipation. This freedom—deeply encoded in the American ethos—requires choice, and in general, students’ choices broaden in their academic and professional careers as their test scores improve. As people and groups of people have more real choices, they also have more freedom. This necessitates rigorously researching how to teach codes of power through discourse in such a way that augments test scores while also affirming the languages and cultures of underserved groups. As shown through the above review, affirming students’ discourses while explicitly teaching mainstream discourse styles shows much promise in this area.

In summary, research cited in this review has shown that learning is a social phenomenon, and socialization often occurs through the discursive patterns that transpire in school. These discourses are necessarily political insofar as they privilege particular forms of interaction and, therefore, define what counts as learning and worthwhile knowledge. Different discourses prevail in different communities, and these discourses can be reinforced or devalued within the predominant discourse of school—which tends to privilege the mainstream discourse that is valued on standardized tests. When students’ home discourse does not match the discourse of school, teachers often struggle to turn classroom interactions into literacy experiences for these students. In this sense, school failure can be seen, at least partially, as a mismatch between discourse expectations across communities, and as students have less literacy experiences in school, the perceived gaps in knowledge and skills can become real. However, discourse styles can be learned, providing promise that teachers can create a discursive space for
students where their home discourse is affirmed and they learn to interact in the discourse of the dominant culture, if they so choose.

For discursive moves to be used by students, they generally need to be modeled by the teacher before students are given opportunities to integrate these moves into their talk. However, different types of moves lead to different types of engagement and achievement. Substantive engagement with the content is most promising for augmenting student achievement, particularly in critical thinking, reasoning, and argumentation. This substantive engagement is most likely achieved through discourse where teachers and students focus on revoicing, asking authentic questions, integrating others’ ideas, and including a high-density of reasoning words. Such a discursive methodology is most likely to augment substantive engagement with the content, sophistication of reasoned arguments, and student achievement as measured by standardized tests. In addition, there is some evidence that this kind of discursive method can be taught to teachers and applied in new contexts.

As mentioned above, future research on discourse and learning will be built on a foundation of sociocognitive theory (Langer, 1985; Vygotsky, 1978), qualitative research that has explored the possibilities for discourse to influence student learning (Michaels, 1981), mixed-methods research that has narrowed the focus to the relationship between particular discourse methods and student learning (Wolf et al., 2005), and quantitative research that has applied more systematic measures of both teacher-behavior and student-achievement (Applebee et al., 2003; O'Connor et al., 2013; Topping & Trickey, 2007a, 2007b). Within this context, further research needs to acknowledge the political reality of this psychometric era—that educators are accountable to state and federal mandates for student learning, as measured by quantitative assessments. Qualitative and mixed-methods research has already shown that
engaging students in discourse around controversial topics generally improves their inferential, argumentative, and comprehension skills (Murphy et al., 2009), but very little work has been done that addresses whether or not similar methodologies augment students’ standardized test scores, particularly in argumentative writing at the upper high school level. The use of broadly adopted standardized measures of student achievement (such as the SAT and AP exams) within studies on reasoning would test the degree to which authentic argumentative skills transfer to improvements in test scores.

This study intends to fill this gap in the research by assessing the effectiveness of the most promising discourse methodologies—as described above—on 11th and 12th grade high school students’ argumentative writing, as measured by scores on standardized tests designed and normed by the College Board.
Chapter 3: Research Methods

This study employed a quasi-experimental, nonequivalent control-group design to investigate the effectiveness of particular discourse moves in augmenting students’ argumentative writing. Six classes of 11th and 12th grade English language arts students at a Boston public high school were given a pretest designed by the College Board to establish baseline scores for argumentative writing. Three of the six classes were assigned to the treatment group and three to the comparison. The treatment group received a discourse intervention that focused on linking ideas and pressing for reasoning, and the comparison group participated in discourse where the teacher focused only on procedural facilitation. At the end of the two-week intervention, both groups took a posttest on argumentative writing. With a continuous dependent variable (argumentative writing scores), a nominal independent variable (discourse method), and pretest scores as a covariate, ANCOVA was used (after testing necessary assumptions) to compare the adjusted means of the dependent variable. The following chapter articulates this methodology in more detail, discussing the population, sampling, statistical power, data collection methods, instruments, procedures for the intervention, possible extraneous variables, steps in the data analysis, validity, reliability, generalizability, protection of human subjects, and possible concerns about researcher bias.

Research Question

What is the difference in argumentative writing scores between English language arts students in 11th and 12th grade who engage in classroom discourse where the teacher focuses on procedural facilitation and English language arts students in 11th and 12th grade who engage in classroom discourse where the teacher focuses on linking ideas and pressing for reasoning?
Directional Hypothesis

English language arts students in 11th and 12th grade who participate in discussions where the teacher focuses on linking ideas and pressing for reasoning will have significantly higher mean argumentative writing scores than English language arts students in 11th and 12th grade who participate in discussions where the teacher focuses on procedural facilitation.

Research Design

This study employed a quasi-experimental, nonequivalent control-group design (Gall, Borg, & Gall, 1996). Despite (1) the use of treatment and comparison groups, (2) researcher manipulation of the independent variable, (3) controls for possible extraneous variables, and (4) random assignment of full intact classes of students to the treatment or comparison group, random assignment of students to these classes was not possible because the students’ classes predated the study (Creswell, 2012). In short, the grouping patterns prevented student assignment into English classes from being random and the research design from being experimental. Manipulation of student scheduling in order to achieve randomization for the sake of the study would have been unethical because it would have privileged the needs of the researcher over the study. As such, a quasi-experimental design was used.

All groups were given a pretest to establish baseline scores for students’ argumentative writing. Half of the classes were randomly assigned to the treatment group and half to the comparison group. The treatment group received a discourse intervention that focused on linking ideas and pressing for reasoning, and the comparison group participated in discourse that focused only on procedural facilitation. At the end of the two-week instructional intervention, both groups took a posttest on argumentative writing.
The study was two-weeks in length because this duration allowed both (1) for more rigorous controls of extraneous variables than a longer study would permit, and (2) for an intervention to be tested that could reasonably be implemented by classroom teachers without requiring an onerous reworking of their curriculum. Prior research in this area where the intervention encompassed much, if not all, of an entire school year tended to be either descriptive in nature (Applebee et al., 2003) or required a substantive overhaul of much of the curriculum (D. Kuhn & Crowell, 2011; Topping & Trickey, 2007a, 2007b). Research similar to the present study that not only focused more closely on particular discursive interventions in the context of an already functioning class but also that attempted to control for potential confounding variables, such as curricular and teacher effects, tended to be of a similar duration to the present study (Mercer et al., 1999; O'Connor et al., 2013; Sandora et al., 1999; Webb et al., 2008).

The original motivation for beginning this research—namely, that students both at the research site and in the country at large tend to struggle on argumentative tasks more than on other writing tasks (see Chapter 1 for discussion)—drove this decision for the duration of the study to be two full weeks of classroom instruction. An intervention was essential that helped teachers support students in moving forward in argumentative writing, particularly an intervention where effects could be seen in a relatively short period of time. After all, state and national standards require English teachers to improve students’ learning in a large variety of areas, including writing and reading in many genres, speaking and listening, and media literacy ("Common core standards initiative," 2010). This means that teachers and students do not have the luxury of spending many months going into all topics and skills in great depth. Often, it is important to see growth on both informal assessments and on standardized assessments in a relatively short period of time.
The nonequivalent control-group design used in this study differs from an experimental control-group design only insofar as the students were not randomly assigned to classes. The primary threat to internal validity (as discussed in detail below) was that group differences on the posttest may have been due to preexisting differences rather than to the effect of the treatment. To mitigate this threat, a statistical technique was used (Analysis of Covariance) that minimized the effect of the initial group differences by adjusting the posttest means based on the pretest means (Gall et al., 1996).

**Population and Sampling**

**Research site.**

Research for this study was conducted at Community Leadership School (CLS). CLS is a pilot high school in the Boston Public Schools (BPS) with enrollment of approximately 500 students. As compared to the district at large, African-American students, females, and low-income students are over-represented; whereas, students in special education, students with limited English proficiency, and males are under-represented (see Table 2 for further demographic information).

Table 2

*Demographic data of Community Leadership School* ("Enrollment data: Boston," 2012)

<table>
<thead>
<tr>
<th>Category</th>
<th>Percentage of Population at Community Leadership School</th>
<th>Percentage of Population in the Boston Public Schools</th>
</tr>
</thead>
<tbody>
<tr>
<td>African-American</td>
<td>43.5</td>
<td>33.7</td>
</tr>
<tr>
<td>Hispanic</td>
<td>43.2</td>
<td>43</td>
</tr>
<tr>
<td>White</td>
<td>8.6</td>
<td>12.6</td>
</tr>
<tr>
<td>Asian</td>
<td>3.8</td>
<td>8.3</td>
</tr>
<tr>
<td>Female</td>
<td>54</td>
<td>48</td>
</tr>
<tr>
<td>Male</td>
<td>46</td>
<td>52</td>
</tr>
<tr>
<td>Special Education</td>
<td>16</td>
<td>19</td>
</tr>
<tr>
<td>Limited English Proficient</td>
<td>27</td>
<td>30</td>
</tr>
<tr>
<td>Low Income</td>
<td>75</td>
<td>70</td>
</tr>
</tbody>
</table>
Historically, as a pilot school in BPS, CLS has had five autonomies: staffing, budget, governance-and-policies, school calendar, and curriculum-and-assessment ("Boston Pilot/Horace Mann Schools Network," 2011). With budget cuts in recent years, however, the staffing and budgetary autonomies have been effectively trimmed from the list, leaving pilot schools in Boston with less autonomy than they had previously.

In this context of being able to make independent decisions about curriculum and assessment but with a dwindling budget and diminishing autonomy, CLS faces the challenge of continually improving student learning among a population of historically underserved students. Of secondary but not insignificant importance is the necessity of CLS’s meeting both district and state mandates. Since the founding of the school in 2002, CLS students have underperformed as compared to the state on the Massachusetts Comprehensive Assessment System (MCAS), a high-stakes graduation exam. For example, on the 2012 MCAS, 77% of students in the school scored proficient or advanced in English language arts, as compared to 88% in the state—with only 8% of CLS’s students scoring advanced, compared to 37% at the state level. In addition, SAT and AP scores fall below national averages. For example, on the most recent SAT, the mean critical reading score of CLS students was 402, compared to 515 in the nation as a whole, and the mean writing score was 396, compared to 506 ("SAT report: Only 43 percent of 2012 college-bound seniors are college ready," 2012). On the two most recent AP English exams (language and literature), only 11% of students earned a qualifying score of 3, 4, or 5—compared to approximately 60% in the country at large ("The 8th annual AP report to the nation," 2012). In this sense, CLS faces challenges associated with other urban high schools throughout Boston and across the country (Kane et al., 2009).
Sampling.

Convenience sampling was used to choose this site for the study because the researcher has been serving as an English teacher in the school for the past nine years. English language arts was the subject area to be studied. Although both math and English are assessed on the state graduation exam, on the SAT, and on AP exams, students have historically performed better in math than English at CLS, with—for example—42% of CLS students scoring advanced on the math portion of the 2012 graduation exam and only 8% in English.

Although the total population of the school is approximately 500, the target population for this study was 11th and 12th grade students in general education classes at CLS, namely, students who were not placed into substantially separate learning environments due to special needs or limited English proficiency (LEP). Removing students with special needs and students who are designated LEP from the school’s total population of 11th and 12th graders yielded a target population of 169 students. The primary general education teacher at each grade-level participated in the study, yielding an initial sample size of approximately 130 students. Three classes in each grade participated (6 total classes), allowing half of these classes to be randomly assigned to the treatment group and half assigned to the comparison group.

Across both grades and each of the teachers’ classes, students’ race and income level were nearly evenly distributed so that the treatment and comparison groups contained near equal proportions of students across all demographic categories (see Chapter 4 for further discussion).

As mentioned above, the target population was the total general education population of 11th and 12th graders in the school, and a relatively large sample was drawn from this population in an attempt to determine the effect of discourse on argumentative writing in English language arts for CLS upperclassmen. Given that the demographic data of CLS are similar to the district
at large, some limited generalizations may be drawn, requiring replication of this study using randomized samples for further substantiation.

This sampling strategy allowed the researcher to address the difference in student achievement between CLS English language arts upperclassmen who engage in classroom discourse where the teacher focuses on procedural facilitation and CLS English language arts students who engage in classroom discourse where the teacher focuses on linking ideas and pressing for reasoning. Though not ideal, the sampling strategy described above was necessary because a simple random sample of the population of the school was not practical or ethical. As mentioned above, the formation of the classes predated the study, and it would have been impracticable to alter scheduling for the sake of the study.

**Power analysis.**

The statistical software G*Power allows users to calculate a priori statistical power for ANCOVA by inputting effect size, error probability ($\alpha$), statistical power ($1 - \beta$), degrees of freedom, number of groups, and number of covariates. Essentially, by setting the probability of Type I error at 0.05 (that is, the probability of rejecting the null hypothesis when it is actually true) and the probability of Type II error at 0.2 (that is, the probability of accepting the null hypothesis when it is actually false), a sample size can be arrived at that makes the study worth undertaking (Mayers, 2013). In essence, $\alpha$ and $\beta$ have an inverse relationship such that decreasing the likelihood of Type I error increases the likelihood of Type II error and vice versa. As such, a balance must be struck between $\alpha$ and $\beta$. Keeping $\alpha$ at $p = 0.05$, $\beta$ can be decreased (and statistical power increased) by applying a stronger treatment or increasing the sample size (Tabachnick & Fidell, 2007b). In this study, applying a stronger treatment was not consistent with a review of the literature and increasing the sample size was not practically possible given
that all teachers in the school who taught more than one 11th or 12th grade English class were already participating.

Using G*Power before the study to calculate this necessary sample size for one-way ANCOVA, a sample size of 90 was necessary to achieve statistical power of 0.8 with \( p = 0.05 \), 1 degree of freedom, 2 groups, a medium effect size, and 1 covariate (Mayers, 2013; Tabachnick & Fidell, 2007a). As such, the planned sample size for this study (130) was sufficient to achieve a high degree of statistical power. Even with a relatively high attrition rate, the sample size was likely to stay above 90. Essentially, this means that given the conditions listed above, the statistical power calculated before the study was high enough to make the study worth pursuing (Tabachnick & Fidell, 2007b). Keeping the significance level fixed, increasing the sample size would also increase the power (Weiss, 1999); however, assuming a small observed effect-size, either the significance level would have to be lowered or the sample size raised to bring the measure of power above 0.8 (Weiss, 1999). In short, the study as proposed, with a sample size of 130, would make the detection of a false null hypothesis high as long as the observed effect is not very small (Cohen, 1988; Weiss, 1999).

**Participants.**

Participants in the study were 11th and 12th grade English language arts teachers and students at CLS. Students at CLS have the demographic profile described in Table 2 (see Table 3 for sample demographics). The participating teachers were both White males with approximately 10 years teaching experience. The author of this study was the primary teacher of 12th grade students in their English language arts classes, and the other participating teacher was the primary teacher for 11th grade students in their English language arts classes. In total, these two teachers taught six English language arts classes, all of which were involved in the study.
All of the additional 11th or 12th grade instructors not involved in the study taught English language arts either to only one section of students or to classes of students who were composed primarily of students with special needs or students who have been designated as Limited English Proficient. Studying the relatively small numbers of these students in special populations was beyond the scope and methodology of this study (see Chapter 5 for further discussion).

Teachers and students were to benefit from this study insofar as they would gain a better understanding of the effect of particular discussion methodologies on student learning. The author of this study played a similar role in the school to other teachers, did not hold any position of power in the school, and did not perform any evaluative duties of any teachers. Because of their participation in this study, teachers, students, and administrators received information about the effectiveness of discourse interventions currently in use in the school. Given the importance of student achievement scores for accessing systems of power (Delpit, 1988), for preserving teachers’ jobs, and for maintaining a viable school, such information about the effectiveness of interventions was highly sought after in the school community.

Data Collection

Instruments.

This study aimed to assess the difference in argumentative writing achievement test scores between high school English language arts students who engaged in classroom discourse where the teacher focused on procedural facilitation and high school English language arts students who engaged in classroom discourse where the teacher focused on linking ideas and pressing for reasoning. As such, a valid and reliable instrument was required to measure argumentative writing. Standardized achievement test scores served as a proxy for this
construct. Such an approach was justified because of the importance of achievement test scores in providing historically underserved students—similar to those in the target population—access to systems of power (Delpit, 1988).

**Instrument selection.**

Instruments designed by the College Board were used in this study because their validity and reliability have been extensively tested (see, for example, E. Lee, Lee, & Brennan, 2012; Patterson, Packman, & Kobrin, 2011; Shaw, Kobrin, Patterson, & Mattern, 2012) and because performance on these exams provides the target population of students with access to scholarships and/or college credit (see, for example, "Boston University Boston high school scholarship," 2012).

In particular, teachers and administrators at the research site are moving toward using the College Board’s Advanced Placement English Language exam as the foundation of a general literacy assessment. This exam is given almost exclusively to 11th and 12th grade students to assess their argumentative and analytical skills in English. As such, it was a perfect fit as an assessment of argumentative writing among 11th and 12th graders at the research site. More specifically, one of the exam’s essay questions requires students to synthesize information from at least three (of approximately six) sources and incorporate it into a well-developed argument that responds to a prompt. The College Board has released global mean and standard deviation data for the 2007 – 2012 essay questions. These exams include prompts about the effectiveness of advertising, the elimination of the penny coin, space exploration, technology in schools, the locavore movement, and the viability of the United States Postal Service (see, for example, "The AP English Language and Composition Exam," 2013). The College Board granted the researcher permission to use these instruments for this study. The purpose of this argumentative
essay instrument developed and tested by College Board is to measure 11th and 12th grade high school students’ competency in argumentative writing, particularly in their ability to synthesize evidence from a number of different sources and incorporate it into a well-developed essay that argues a position on an issue of public discourse. This instrument was a perfect match for assessing the transfer of these same skills from classroom discourse to individual writing.

**Instrument validity.**

Given that this study employed a pretest and a posttest, two of these exams were selected by the participating teachers for the study. All of these questions were written by a test development committee that was composed of three to four Advanced Placement high school English teachers and three to four college English professors—providing content-related evidence of validity ("Course & exam development," 2012).

Criterion-related evidence of validity is provided by studies on the relationship between AP exams and college GPA. In general, Patterson et al. (2011) found that students who received a score of 2, 3, 4, or 5 on AP English exams outperformed reference groups by a statistically significant margin in college GPA in English. Patterson et al. (2011) provide evidence of both predictive validity (by examining the relationship between AP English exam scores for high school students and eventual college GPA) and concurrent validity (by examining the relationship between AP English exam scores for college students and their concurrent GPA).

More specifically, Mattern, Shaw, and Xiong (2009) found that students who scored a 1 or a 2 on the AP English Language and Composition exam outperformed a comparison group who did not take the exam by 0.139 points in the students’ four-year college grade point average, though with a small effect size of 0.187 (Cohen, 1988). In addition, students who scored a 3, 4, or 5 on the AP English Language and Composition exam outperformed those who scored a 1 or a
2 by 0.423 GPA points, with a medium effect size of 0.571 (Mattern et al., 2009). Finally, students who scored a 3, 4, or 5 on the AP English Language and Composition exam outperformed a comparison group who did not take the course by GPA 0.562 points, with a medium effect size of 0.758 (Mattern et al., 2009).

Construct-related evidence of validity is provided as follows: (1) In the judgment of the content experts on the test development committee, the tasks on this exam require argumentative writing skills, (2) the test development committee indicates that the tasks are relevant and representative of argumentative writing, and (3) a high correlation exists between scores on the test and GPA in English (English: English Language and Composition course description, 2010; Patterson et al., 2011).

**Measurement scale and reliability.**

In addition to data on the global mean and standard deviation for each question, the College Board provides public access to rubrics for scoring. Also, the researcher—who did the scoring—has been trained by the Educational Testing Service to score synthesis essays on the AP English Language and Composition exam at the annual AP Reading. An additional ETS employee scored a random sample of 30% of the essays in the study to assess the reliability of the scores to the College Board rubric. The essay rubrics use a 0 – 9 interval scale that was designed by college professors at the AP Reading. Taken together these documents and this information provided the researcher with a means for assessing student achievement on an interval scale by scoring the pre- and post-tests. Using such an interval scale requires an assumption that there are relatively equal levels of achievement between each of the 10 scores. This assumption allows for more sensitive data analysis and is justified when using standardized measures of achievement in educational research (Fraenkel et al., 2012).
When trained College Board readers, such as the ones who were utilized in the study, use these tools, reliability remains high—allowing for accurate comparisons between students’ scores and the global mean (E. Lee et al., 2012). For example, E. Lee et al. (2012) found a score correlation of 0.765 across two separate scorings of the 2007 AP English Language and Composition exam.

**Procedures.**

This study was two weeks in duration, employing a nonequivalent control-group design that exposed randomly assigned classes of students to procedural facilitation of discussion or to facilitation that focused on linking ideas and pressing for reasoning. All three general education classes taught by each of the two participating teachers were involved in the study, yielding six total participating classes. All of the students in these classes took a pretest that assessed argumentative writing skills. One of the 11th grade classes was assigned randomly to receive procedural facilitation, and the other two 11th grade classes received facilitation that focused on linking and pressing. In contrast, one of the 12th grade classes was assigned randomly to receive linking and pressing facilitation, and the other 12th grade classes received procedural facilitation. If linking and pressing facilitation were to be found to have a significant effect on students’ argumentative writing, then all classes that did not receive a two-week bout of the treatment would receive it after the conclusion of the study, giving each of the classes an equal amount of time experiencing linking and pressing facilitation.

Building on the description in Chapters 1 and 2, procedural facilitation consisted of the following 13 elements:

1. The teacher provides the class with an argumentative prompt provided by the College Board for a synthesis question on the AP English Language and Composition exam.
2. Students respond to the prompt in writing without reading the sources.

3. Students place their view on a continuum that corresponds to their viewpoint (see Figure 2).

<table>
<thead>
<tr>
<th>What are the effects of advertising?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entirely Negative ------ Mostly Negative ------ Neutral ------ Mostly Positive ------ Entirely Positive</td>
</tr>
</tbody>
</table>

Figure 2. Continuum on the effects of advertising.

4. Students discuss their viewpoint with a partner.

5. The teacher opens class discussion to the whole class, asking students to share their perspectives. The teacher calls on students in the order they raise their hands, allowing students to take the lead in expressing their perspectives and in responding to each other.

6. Students are broken into groups so that each group of three to five students is responsible for reading and explaining one of the six to eight sources that correspond to the prompt.

7. Students present the central message of each source to their classmates and place the source on the continuum in a place that they think corresponds with the authors’ view.

8. The teacher asks the class if they wish to discuss the group’s explanation.

9. If so, the teacher calls on students in the approximate order that they raise their hands.

10. After each of the groups has presented their source and placed the source on the continuum, the teacher groups students by similar view, with three to five students per group.

11. Each group is given 15 minutes to discuss their view and prepare a presentation that makes a claim and substantiates that claim with evidence and reasoning.

12. Each group shares their viewpoint.
13. The teacher opens class discussion to the whole class and calls on students in the order that they raise their hands.

Building on the brief introduction in Chapters 1 and 2, facilitation that focuses on linking and pressing consisted of the following 19 elements:

1. The teacher provides the class with an argumentative prompt provided by College Board for a synthesis question on the AP English Language and Composition exam.

2. Students respond to the prompt in writing without reading the sources.

3. Students place their view on a continuum that corresponds to their viewpoint (See Figure 2 above).

4. Students discuss their viewpoint with a partner. The teacher directs students to explain their view, support their view with evidence, and explain their reasoning.

5. The teacher opens class discussion to the whole class, asking students to share their perspectives; however, in contrast to procedural facilitation, the teacher rephrases what students say, asks them to explain their reasoning, and asks them to link their ideas to the evidence and/or to what has previously been said by their classmates.

6. Students are broken into groups so that each group of three to five students is responsible for reading and explaining one of the six to eight sources that correspond to the prompt.

7. The first group of students presents the central message of each source to their classmates and place the source on the continuum in a place that they think corresponds with the authors’ view.

8. The teacher rephrases the students’ articulation of the central argument and asks if s/he has understood their thoughts accurately.
9. If the students respond negatively, the teacher asks them to explain again, and the teacher rephrases again.

10. Once the students affirm that the teacher has understood their articulation of the source’s central message, the teacher presses for reasoning by asking the students to explain how they determined the central message and why they placed it where they did on the continuum.

11. Once the students have explained, the teacher asks if the class has any questions. If they do, the teacher calls on students in the approximate order that they raise their hands.

12. For the second group and all subsequent groups, the teacher follows steps 7 to 11.

13. In addition, the teacher asks the students to explain the relationship between their source and the previous groups source by asking whether their author would agree with, disagree with, or qualify the view of at least one of the previous authors.

14. After students respond to this question, the teacher rephrases their thinking and asks if that rephrasing is accurate.

15. Once the students affirm that the rephrasing is accurate, the teacher moves to the next group and follows steps 7 through 14.

16. After each of the groups has presented their source and placed the source on the continuum, the teacher groups students by similar view, with three to five students per group.

17. Each group is given 15 minutes to discuss their view and prepare a presentation that makes a claim and substantiates that claim with evidence and reasoning.

18. Each group shares their viewpoint.
19. The teacher opens class discussion to the whole class and calls on students in the order that they raise their hands; however, in contrast to procedural facilitation, the teacher rephrases what students say, asks them to explain their reasoning, and asks them to link their ideas to the sources and/or to what has previously been said by their classmates.

Three of the six participating classes engaged in discussions where the teacher focused on procedural facilitation and three of the six classes engaged in discussions where the teacher focused on linking and pressing facilitation. The linking and pressing facilitation aimed to improve students’ facility with sophisticated argumentation. Specifically, all classes assigned to procedural facilitation followed the 13 steps enumerated above, and all classes assigned to linking and pressing facilitation followed the 19 steps above.

**Summary of how procedures were used.**

The study lasted two weeks, during which time all teachers used materials provided by the College Board for the synthesis essays (which included an argumentative prompt and six to eight related sources) to teach lessons on argumentative writing (see, for example, "The AP English Language and Composition Exam," 2013). Lessons were standardized across all classes so that the only substantive difference were the discussion portion of each lesson (namely, the independent variable).

This synthesis essay has been given to high school students since 2007, and two versions of the exam were publicly released from 2007 to 2011, with a single exam being released in 2012 and 2013, and the College Board also releasing statistical data, such as global mean and standard deviation, for 7 of these 12 exams. Two of these six were used for the pre- and post-tests. Participating teachers then chose among the 10 remaining exams to use for daily lessons with two synthesis essays being taught and discussed each of the two weeks in the study. In
summary, of the 12 total synthesis essays, 2 were used for assessment on the pretest and postest (on advertising and the elimination of the penny coin) and 4 were used for lessons (on canonical literature in high school English class, individuality and conformity in schools, creating memorials, and space exploration). The author received permission from the College Board to use these assessments as long as the assessments were not published in this manuscript.

At the end of the two-week study, all students took a posttest, assessing whether or not their achievement scores in argumentative writing had changed.

**Fidelity of implementation.**

Following the standards of Johnson, Mellard, Fuchs, and McKnight (2006), in order to ensure fidelity of implementation to the discussion intervention, the researcher minimized the complexity of each of the interventions by creating a checklist similar to the procedure described above (see Appendix A). The researcher also provided all resources and materials necessary for the lessons. Finally, the researcher recorded the first two lessons of each teacher in each intervention and compared the teachers’ procedures with those on the checklist.

If teachers were not following the checklist, the researcher would view the recording with the teacher, and they would discuss how the researcher could support the teacher in more closely following the checklist. The researcher would, then, videotape the next two lessons. If fidelity did not improve, that teacher’s data would be omitted from the study. In addition, the researcher would also videotape the first lesson of each teacher in the second week of the study and follow the same process listed above to ensure fidelity.

**Possible extraneous variables.**

Taken together, Fraenkel et al. (2012) and Gall et al. (1996) provide a comprehensive list of 12 potential extraneous variables. By employing a quasi-experimental, nonequivalent control-
group design where (1) the same teachers implemented discussion for both the comparison and treatment groups, (2) the sample was demographically similar to the population, (3) a pre-test and post-test was used, (4) the pretests and posttests were intermixed and scored at the end of the study, and (5) the statistical technique for analysis minimized the affects of preexisting differences in argumentative writing skills, the design and statistical methodology provided some controls for maturation, testing threats, instrument decay, and regression (Fraenkel et al., 2012; Gall et al., 1996); however, location, history, data-collector characteristics, and implementation variability could have posed serious threats to the study. As such, they are addressed below, in addition to experimental treatment diffusion, resentful demoralization of the control group, compensatory rivalry by the control group, and compensatory equalization of treatments.

Maturation was controlled by using the same pretest and posttest for each group. This provided for the greatest likelihood under the conditions of the study for maturation effects to be spread evenly across the treatment and control groups. In a similar way, testing threats posed relatively little challenge to this study due to the similar goals of both the comparison and the treatment groups. Students were told that the unit was focused on argumentation, and they had the opportunity to take the same tests in both the treatment and comparison groups; therefore, any gain that was shown simply because of students’ experience with the pretest was likely to occur across both the treatment and comparison groups.

Instrument decay was controlled by using a single, College Board trained scorer (the author) who scored the essays blind to group assignment and pretest/posttest. The scorer only had access to the text of each student’s essay and the College Board rubric. In addition, a random sample of the essays (30%) was rescored blind by a second rater to establish a high degree of interrater reliability. Administering the pretest provided some control for statistical
regression because the posttests scores were adjusted based on the pretest scores for both the treatment and control groups.

Location threat was also minimized in this particular design because the same classrooms and teachers were used for the treatment and control groups. In addition, the school operates on a rotating schedule, so each class met throughout the day across the two-week study.

Threats to internal validity due to data-collector characteristics and data-collector bias were minimized in this study by establishing clear protocols for testing and using the same teachers to implement the tests across both phases of the study. Because the teachers were implementing the intervention, they needed to know that the study aimed to compare differences in students’ argumentative writing abilities. This awareness could have led to teachers’ subconsciously favoring one method over another; however, this effect was minimized by standardizing all testing and implementation procedures (see Procedures section).

As mentioned above, implementation variability was controlled by assessing fidelity to the discourse method through the use of checklists and periodic assessments of classroom recordings by the researcher. In addition, as suggested by Fraenkel et al. (2012), both discussion methods were implemented by all teachers with their preference for one or the other stated in advance. This information is discussed in Chapter 5 so as to report on, but not necessarily control for, this particular extraneous variable.

Experimental treatment diffusion, where the comparison group wishes to be part of the treatment group, was minimized by aligning the materials and methods of each class as closely as possible. The only difference across the groups was the extent to which the teacher pressed for reasoning and linked ideas across the classes. Because of this consistency, resentful
demoralization of the control group was less likely, particularly because there were no obvious reasons that students would prefer more facilitation from their teachers.

Compensatory rivalry by the control group, where the control group begins competing with the treatment group, was an unlikely threat because both groups were engaged with the same units of study. There was no outstanding reason to believe that one group would have a greater desire to compete than is normally the case across classes. In addition, compensatory equalization of treatments, where the treatment group received goods or services not provided to the comparison group, was not a threat because all classes and all groups received the same materials and resources.

Finally, history and differential selection posed potential problems because the researcher could not be assured that the treatment and comparison groups would be similar in all respects except for the facilitation treatment they received. It was likely that the demographic information of the sample would be similar to that of the population, but because students’ placement into their classes could not be controlled, this fact could not be established until students entered their classes. At that point, effects due to history and differential selection could be reported. See Chapters 4 and 5 for further discussion.

Data Analysis

Preparation of the data file.

When students took the pre- and post-tests, they were assigned a unique identification number to record on the exams instead of their names. In IBM’s statistical analysis software SPSS, the researcher input grade, treatment, and demographic information associated with each student’s number and then deleted the names from the data file. The researcher named these variables, classified them by type, and labeled the level of measurement. After students
completed the pretest and posttest, the rater scored the tests and input scores for each of the participants in the study. At this point, all names had been deleted.

Frequency distributions were used to view how many participants belonged to different categories (such as grade-level, teacher, gender, ethnicity, treatment group, and comparison group), and histograms were used to determine how many participants had earned each score. Given that the data for the dependent variable were classified as continuous (interval data with 10 score points), if the data were normally distributed, mean would be used as a measure of central tendency. Median would be used if the data were severely skewed and outliers posed a problem. In addition, standard deviation and variance were used as measures of spread.

Regarding data cleaning, after the data were entered into SPSS, the researcher verified that the data values were correct and that they conformed to logical values for each variable—for example, assuring that inputs for the variable gender had only two values: male and female, that inputs for grade-level were either 11 or 12, and that inputs for argumentative writing were from 0 through 9 (as defined by the rubrics for the instrument). To be included in analysis, all variables should have conformed to reasonable limitations that matched the real-world concepts they intended to measure (Osborne & Overbay, 2008).

After data were checked as described above, a boxplot was used to examine outliers. An outlier is a data point that is “far outside the norm for a variable or population” (Osborne & Overbay, 2008, p. 205). There is not a precise set of criteria that define outliers; however, a practical guideline is that outliers are three or more standard deviations away from the mean (Osborne & Overbay, 2008). They are of concern because they deviate so significantly from the other collected data that they may be affected by a mechanism that has not been accounted for in the study (Osborne & Overbay, 2008). As such, outliers can have negative effects on statistical
analyses. Outliers increase error rates and decrease statistical power. In addition, if the outliers are non-randomly distributed, they decrease the normal distribution of the data—increasing the likelihood of Type I and Type II errors. Finally, the presence of outliers in data sets distorts parameter and statistical estimates whether the researcher is using parametric or nonparametric tests (Osborne & Overbay, 2008). However, outliers should not necessarily be eliminated from the data set. This choice depends on reasoned inferences about the cause of the outliers.

Osborne and Overbay (2008) enumerate six potential causes for outliers: data error, intentional or motivated misreporting, sampling error, standardization failure, faulty distributional assumptions, and legitimate cases sampled from the correct population that are worthy of further inquiry. The inferred cause of the outlier motivated the decision for including or omitting the data (see Chapter 4 for further discussion).

As mentioned above, a histogram was used to assess the normality of the data collected for the dependent variable. If the assumption of normality was met, parametric tests (t-test and ANCOVA) would be used to test the hypotheses. If the assumption of normality was not met, the researcher would consider non-parametric tests or transforming the skewed data (see Chapter 4).

**Transformation of the data.**

As mentioned above, if the data were skewed, they would be transformed in a manner that corresponded to the degree and direction of their skewness (Tabachnick & Fidell, 2007b).

In addition, in order to improve reliability, students’ scores were equated across all test forms used in the study. The College Board has been producing new test forms for the synthesis argumentative essay on the AP English Language and Composition exam since 2007. Although these test forms are expressly designed to measure the same argumentative skills, it is impossible
to construct test forms that have identical characteristics (Kolen & Brennan, 2004). As such, comparisons of raw scores across test forms would not have produced reliable gain scores; however, equating different forms of the test addressed these issues (Kolen, 1988).

In order to use test equating, alternate forms of the test need to be built to the same test specification. Additionally, test content and conditions of measurement need to be held constant (Kolen & Brennan, 2004). The test forms and proctoring instructions for the AP English Language exam meet these criteria.

Regarding the present study, Kolen and Brennan (2004, pp. 7-8) suggest the following steps for implementing equating:

1) Decide on the purpose for equating.
2) Construct alternate forms.
3) Choose a design for data collection.
4) Implement the data collection design.
5) Define what constitutes correspondence between scores on alternate forms.
6) Specify the statistical methods used to estimate the defined score correspondence.

For the present study, the purpose for equating was to measure improvement in argumentative writing while using different forms of the same instrument. Using different forms in conjunction with a pretest/posttest nonequivalent control-group design minimized practice effects (Fraenkel et al., 2012). As mentioned above, the College Board has created alternate forms of the same exam every year since 2007. The design for data collection was nonequivalent-groups, where groups of test-takers with similar demographic characteristics took each form of the exam (Livingston, 2004). This design was justified because very large
numbers of students have taken the AP English Language exam each year since 2007 \((N > 280,000)\) ("AP data: Program summary report," 2007).

Linear equating was used to define correspondence between scores on alternate forms for the following reasons: (1) the College Board reports the global mean and standard deviation for these test forms, (2) differences in difficulty for test forms are not necessarily equal across the entire score range due to the nature of ordinal data, and (3) the sample size was so large these exams when they were given at the national level (Kolen, 1988; Livingston, 2004). Linear equating allowed for the possibility that the “difference in relative difficulty between two forms [was]…variable along the score scale” (Kolen, 1988, p. 33). For example, some test forms might have been more difficult for students who score on the low range than for students who score on the high range. The range of the original scores was from 0 to 9, as described on the College Board rubrics (see, for example, "The AP English Language and Composition Exam," 2013). However, because the means for the possible test-forms vary from 4.62 to 5.0, and the standard deviations vary from 1.58 to 1.74, it was possible for the equated scores to range from approximately -0.5 to 9.5, depending on the order in which the forms were given.

Finally, the linear conversion is defined by setting the standardized scores on the two forms equal as shown below (Kolen, 1988):

\[
\frac{X_1 - \bar{X}_1}{S_1} = \frac{X_2 - \bar{X}_2}{S_2}
\]

Solving for \(X_1\) yields the following:

\[
X_1 = \frac{S_1}{S_2} X_2 + (\bar{X}_1 - \frac{S_1}{S_2} \bar{X}_2)
\]
where $X_1$ is the form 1 equivalent of a form 2 score, $X_2$ is the corresponding score on form 2 (post-test), $\bar{X}_1$ is the global mean on form 1 (pre-test), $\bar{X}_2$ is the global mean on form 2, $S_1$ is the global standard deviation for form 1, and $S_2$ is the global standard deviation for form 2. This equation yields the form 1 equivalent of a form 2 score, allowing the researcher to compare equated test scores instead of raw scores. This method of test equating augments reliability across different test forms (Kolen & Brennan, 2004), allowing for reliable computation of gain scores from pre-test to post-test even in conditions where the difficulty of the test varies across the score scale.

**Choice of statistical techniques.**

Given that (a) the research design was quasi-experimental (b) the groups of participants were preexisting (11th and 12th grade English classes), (c) the dependent variable was continuous (argumentative writing scores), and (d) the independent variable was nominal at two levels (discourse method), one-way ANOVA was used initially to assess the difference in the means on the posttest. The purpose of this test was to compare the means of the dependent variable between two groups before adding a covariate (Muijs, 2011).

One-way ANCOVA was then used as a statistical method in order to control for pretest scores as a covariate. Similar to ANOVA, the purpose of one-way ANCOVA was to compare potential differences in the mean of the dependent variable between two groups, except that one-way ANCOVA allowed for this comparison while also controlling for pretest scores (Muijs, 2011). As a measure of effect size, partial eta squared ($\eta_p^2$) was used to approximate the percentage of variability in argumentative writing scores that could be accounted for by the independent variable while controlling for the covariate (Muijs, 2011).
Answering research question.

ANCOVA was used to answer the research question about the difference in argumentative writing scores between high school English language arts students who engage in classroom discourse where the teacher focuses on procedural facilitation and high school English language arts students who engage in classroom discourse where the teacher focuses on linking ideas. The statistical method of one-way ANCOVA allowed for the covariate of pre-test scores to be controlled while doing the above comparison.

Steps.

The following steps were followed when analyzing the data:

1) Test assumptions for use of ANCOVA (Mayers, 2013; Tabachnick & Fidell, 2007b)
   a) Equal sample sizes
   b) Absence of outliers, assessed using boxplot to determine if any data points fall more than three standard deviations from the mean
   c) Normal distribution of both covariate and dependent variable, assessed using histogram and both skewness and kurtosis criteria (values between -1 and +1)
   d) Linear relationship between covariate and dependent variable, assessed using a statistical test of linearity
   e) Correlation between covariate and independent variable, assessed using Pearson’s correlation coefficient, which should fall between 0.3 and 0.9.
   f) Independence of the covariate between groups, assessed using an independent samples t-test
   g) Homogeneity of regression slopes, assessed by building a custom model in IBM’s SPSS to determine if there is a significant interaction between the covariate and
the dependent variable

h) Equality of error variances, assessed using Levene’s test of equality of error variances. This is essentially another method for assessing the homogeneity of regression slopes by checking whether or not the covariate has an equal regression coefficient associated with both groups of the independent variable—procedural facilitation and linking/pressing.

2) If all of the above assumptions are met, perform ANCOVA. If, however, some are not met, the consequence of violating that particular assumption will be discussed.

3) Assess the mean and standard deviation for each group of the independent variable to see if they are different.

4) Run ANOVA to assess the effect of the independent variable (discursive method) on the dependent variable (argument writing) before controlling for pretest scores.

5) Add the covariate—pre-test scores—to the model and use a full factorial model to assess whether or not statistically significant differences in argumentative writing scores across the two groups still manifest themselves in the data when the covariate is controlled for. In other words, do facilitation methods have statistically significant difference in the means for argumentative writing scores when controlling for pre-test scores?

6) Look at the test statistics to determine if the covariate has a significant impact on the difference in the means. Look at the means adjusted for the covariates, because ANCOVA re-estimates the means (making them hypothetical) to account for the effect of the covariate.

7) Look at the effect size using partial eta squared to approximate the percentage of
variability in argumentative writing scores that can be accounted for by the independent variable while controlling for the covariate.

Validity, Reliability, and Generalizability

Internal validity.

Internal validity is “the extent to which extraneous variables have been controlled by the researcher, so that any observed effect can be attributed solely to the treatment variable” (Gall et al., 1996, p. 467). Taken together, Fraenkel et al. (2012) and Gall et al. (1996) provide a comprehensive list of 12 potential threats to internal validity. As discussed in the Data Collection section under Possible Extraneous Variables, possible threats to internal validity of the study included location, history, data-collector characteristics, implementation variability, maturation, testing threats, instrument decay, and regression (Fraenkel et al., 2012; Gall et al., 1996). As such these threats were addressed specifically above; however, it is worth noting here that most of these threats were mitigated by using a quasi-experimental, nonequivalent control-group design where (1) the same teachers implemented discussion for both the comparison and treatment groups, (2) the sample was demographically similar to the population, (3) a pretest and posttest was used, (4) the pretests and posttests were intermixed and scored at the end of the study, and (5) the statistical technique for analysis minimized the effects of preexisting differences in argumentative writing skills.

Reliability.

Reliability “refers to the consistency of scores or answers provided by an instrument” (Fraenkel et al., 2012, p. 162). For this study, reliability needed to be established by (1) standardizing how teachers implemented the instrument, (2) indicating the method for scoring the argumentative essays, (3) reporting a reliability coefficient for previous uses of this
instrument, (4) explaining how gain scores would be calculated in relationship to the global mean through a process of score equating, and (5) describing the method for establishing interrater reliability.

Regarding standardizing implementation, the College Board provides detailed instructions for teachers to proctor exams. Teachers involved in this study followed these procedures. In addition, testing was done on the same students at the same time of day in the same room from pretest to posttest.

Regarding the scoring method, the College Board provides public access to prompts and rubrics for scoring (see, for example, "The AP English Language and Composition Exam," 2013). Although the College Board granted permission to use these instruments, the organization requested that they not be published in this paper; nevertheless, they are readily available through the website cited above. The essay rubrics use a 0 – 9 interval scale that was designed by college professors at the AP reading. Taken together, these documents and this information provided the researcher with a method for assessing changes in student achievement on an interval scale by scoring the pre- and post-tests. The researcher has also been trained by ETS to score the synthesis essay for the College Board at the annual AP Reading.

Regarding reliability coefficients, when trained College Board readers, such as the ones who were involved in the study, use these tools, reliability remains high—allowing for accurate comparisons between students’ scores over time (E. Lee et al., 2012). For example, E. Lee et al. (2012) used an alternate forms, test-retest method for assessing reliability and found a score correlation of 0.765 across two separate scorings of the 2007 AP English Language and Composition exam.
As mentioned in the section “Transformation of Data,” in order to improve reliability even further, students’ scores were equated across all test forms used in the study. The College Board has been producing new test forms for the synthesis essay on the AP English Language and Composition exams since 2007. Although these test forms are expressly designed to measure the same argumentative skills, it is impossible to construct test forms that have identical characteristics (Kolen & Brennan, 2004). As such, comparison of raw scores across test forms would not have produced reliable gain scores; however, equating the different forms of the test addressed these issues (Kolen, 1988). This method of test equating augmented reliability across the different test forms used in this study (Kolen & Brennan, 2004), allowing for reliable computation of gain scores from pretest to posttest (see Transformation of Data section above for more detailed information).

Regarding interrater reliability, a single rater—the author—trained by the College Board scored all the exams using the rubrics and scoring worksheets provided by the College Board. This rater scored the exams blind to treatment, time, and identity of participants. In other words, the rater did not know which essays correspond to which groups. In addition, the order that the rater scored the essays was randomized. Finally, since the dependent variable was interval (with 10 score points), an interrater reliability analysis was performed using Pearson’s correlation coefficient (if the data were normally distributed) and Cronbach’s alpha to assess the consistency and reliability of the primary rater. A second College Board trained rater scored 30% of the essays, and these scores were compared with the initial scores on the essays to determine interrater reliability.
**Generalizability (external validity).**

Generalizability (or external validity) is “the extent to which the findings of an experiment can be applied to individuals and settings beyond those that were studied” (Gall et al., 1996, p. 473). Bracht and Glass (1968) enumerate 12 potential threats to external validity, two under *population validity* and ten under *ecological validity*. Each of these will be addressed in the subsequent sections.

**Population validity.**

Population validity is the degree to which a researcher can generalize from the sample used in a study to a larger population (Gall et al., 1996). In the present study, the initial sample size constituted more than 75% of the target population of 11th and 12th graders and was composed of approximately equal numbers of students in both grades; therefore, it was reasonable to assume that findings from the study would be generalizable to the experimentally accessible population—namely, other students in the school who are upperclassmen and who are not in special education or English as a second language programs (Gall et al., 1996). However, generalizing to the target population of all high school students in the Boston Public Schools would be riskier and would require a careful comparison of the characteristics between the sample and the district at large (Gall et al., 1996).

As compared to the district at large, African-American students, females, and low-income students were over-represented in the sample, as compared to the target population; whereas, students in special education, students with limited English proficiency, and males were under-represented (see Table 2 for further demographic information). These comparisons do not take into account the extent to which the instructional intervention in the present study interacted with student characteristics that were not demographic in nature. For example, different results may
have been obtained with students at different schools with different personological variables such as extraversion-introversion and anxiety levels (Gall et al., 1996). If the population were to differ appreciably from the sample, the generalizability of the findings would be limited (Gall et al., 1996).

**Ecological validity.**

Ecological validity refers to the degree to which “the results of an experiment can be generalized from the set of environmental conditions created by the researcher to different environmental conditions” (Gall et al., 1996, p. 475). A study has high ecological validity when the results can be obtained in a variety of environments by a variety of researchers. This requires an explicit description of the experimental treatment so that the study can be replicated (for a detailed description, see the subsection Procedures in the Data Collection section of this chapter).

In the design, the Hawthorne effect was minimized because all the teachers and students were participating equally in the study, each receiving the same curriculum and resources. Novelty and disruption effects also needed to be minimized. These effects provided a threat to the ecological validity of this study because multiple teachers were involved, and it was impossible to assess whether the treatments would be novel or disruptive in each classroom. In addition, because the treatment length was relatively short, as compared to a full school year, the results of the study may not be generalizable to consistent use throughout the year. Nevertheless, the researcher asked the teachers how much their everyday instruction differed from the instruction provided in the treatments of the study and reported the findings (see Chapter 4).

Concerns about experimenter effects were relatively low. Each teacher implemented both procedural facilitation and linking/pressing facilitation, prima facie controlling for the effect of the teacher; however, the teachers knew that they were participating in a study, and their
preference for one treatment over another may have meant that they—consciously or
subconsciously—altered their interactions with the students in such a way that yielded
unintended consequences. Although a clear procedure was provided for each treatment, the
potential risk did not disappear. As such, in addition to recording the first two lessons of each
week, the researcher also asked the participating teachers if they had a preexisting preference for
one of the treatments over the other, and this information is discussed in Chapter 5.

The pretest used in the study may have posed a threat to ecological validity by sensitizing
students to the elements of the study (Gall et al., 1996). Though this is a serious threat on its
face, the threat was minimized due to the normal practice of testing students on their
argumentative writing in schools. In addition, all students across both treatments should have
been affected equally by the pretest.

Because the post-tests are analogous to essay exams that are given throughout students’
schooling, it was unlikely that the post-test served as a unique learning experience in and of itself
that would bias the study. In addition, all students across all treatments received the same
posttest; therefore, the effect should have been uniform. History and treatment effects may have
interacted, creating unique conditions surrounding the study that made the effectiveness of the
intervention more likely; however, given that progressive teaching practices have been in place
in the school for a number of years, it is unlikely that the discourse methods being used in the
study were seen as particularly innovative. Nonetheless, the researcher asked the participating
teachers the extent to which they thought the treatments were innovative and the extent to which
they were disenchanted with current discourse methods. See Chapter 5 for further discussion.

The discourse method may have been limited in its effect on the dependent variable
(argumentative skills) if the dependent variable had been measured using other means. For
example, if the dependent variable had been measured using multiple choice instead of essays, the results may not have been generalizable; however, given that essay writing is an integral part of virtually all English exams, even if the results do not apply to other ways of measuring the dependent variable, they still provide information about an important segment of English instruction.

Finally, the time of measurement and treatment effects may have interacted in such a way that the effects of any of the treatments may have deteriorated over time. In other words, results on posttests that occurred immediately after the intervention may not have persisted as time passed. Long-term assessment of this potential deterioration depended heavily on the results of the initial two-week intervention. As such, this issue will be discussed further in Chapter 5.

**Protection of Human Subjects**

Although this study included adolescent participants, there was no conceivable way that they could be harmed physically or psychologically because of the study; in fact, the greater likelihood was that they would benefit. The discourse methodologies being compared were both currently in use throughout the school. This study only served to provide a systematic, quantitative comparison of their effects. All aspects of this study, including the discourse methods and observation techniques, were accepted parts of the school culture, and the school’s headmaster fully approved all elements of this study. In addition, all teachers in the study signed informed consent forms so that they understood they could terminate their service as a subject at any time. For scheduling reasons, entire classes received the intervention, and individual teachers and school administrators provided consent for the use of the intervention.

All teacher and student data taken from the study was coded by number. No student or teacher names appeared in any of the final data sets used by the researcher. The researcher kept
all data and assessment materials in a locked cabinet in a locked room at the research site. All of these materials will be destroyed at the full conclusion of the study. Finally, no participants in this study were deceived in any way.

**Concerns about bias.**

Given that the researcher was at the center of both research and praxis in this study, research bias was of particular concern (Machi & McEvoy, 2009). In prior years, the researcher has implemented various discourse methodologies in his classroom, hypothesizing that they would be effective in augmenting student achievement. Improved test scores occurred after the intervention, with qualifying scores on the Advanced Placement English Language and Composition exam increasing by 150%. No causal relationship can be established from these informal observations, but the researcher must recognize his bias toward believing that one may exist. For this reason, the researcher is particularly susceptible to confirmation bias where he may only notice the evidence that supports the hypothesis. This is one of the primary reasons that this study was quantitative, using externally designed, normed, and scored assessments produced by ETS. It was also one of the reasons that the essays were made anonymous and scored blind.

With qualitative, anecdotal evidence that discourse interventions are effective with the target population—secondary English students at CLS—it would have been unethical to withhold the intervention for any substantive period of time. This created a dilemma because to use this quasi-experimental design, a comparison group was needed, from which the treatment would be—at least temporarily— withheld. In particular, because of the researcher’s position in the White, male hetero-normative hegemony in the United States, withholding quality education in order to complete dissertation research seemed unprofessional, even unscrupulous. Given this
positionality, in order to create a realistic comparison group that reasonably approximated status quo discussion in ELA classes, teachers in the comparison group procedurally facilitated discussion. This allowed for all students to participate in discussion throughout the study so that the social process of learning through group talk was active across all groups. This created a more stringent requirement for the test of the effectiveness of the linking and pressing treatment while also meeting essential ethical standards. It is also likely diminished the effect size of the intervention (see Chapter 5 for further discussion). In addition, participating teachers agreed to implement the treatment intervention with the control group after the conclusion of the study if the intervention showed promise for augmenting student achievement. Consistent with this agreement, at the time of this writing, participating teachers were using the results of the study to inform their instructional practice (see Chapters 4 and 5 for results and further discussion).

Conclusion

This section indicates how the research project responded to the problem of practice by addressing a series of overlapping gaps in the literature. Assessing the effect of classroom discourse on argumentative writing is an immensely complex endeavor that could not be addressed comprehensively in this study; however, progress was made in assessing the effectiveness of particular discursive techniques (namely, procedural facilitation and linking and pressing facilitation) in augmenting argumentative writing test scores of 11th and 12th grade high school students in English language arts classes in a Boston Public School. In fact, the process and results of the research informed teachers’ learning about how to use authentic discourse to improve test scores in argumentative writing such that teachers continued employing the most effective discourse methodologies after the conclusion of the study.
Chapter 4: Report of Research Findings

The purpose of this quantitative study was to investigate the effect of classroom discourse on argumentative writing among English language arts students in 11th and 12th grades in a Boston public high school. The study employed a quasi-experimental, nonequivalent control-group design, which is similar to a pretest posttest randomized experiment except that—for both practical and ethical reasons—students could not be randomly assigned to groups (Gall et al., 1996). In the study, six classes of 11th and 12th grade English language arts students ($N = 115$) at a Boston public high school were given a pretest designed by the College Board to establish baseline scores for argumentative writing. Three of the six classes were randomly assigned to the treatment group and three to the comparison group. The treatment group received a discourse intervention that focused on linking ideas and pressing for reasoning, and the comparison group participated in discourse where the teacher focused only on procedural facilitation. The initial class sessions of both participating teachers were recorded to assess fidelity to the method of discourse (for further detail see Chapter 3 and Appendix A). At the end of the two-week instructional intervention, all students took a posttest on argumentative writing that was also created by the College Board. The pretest and posttest were scored by two employees of the Educational Testing Service (including the author) who have been trained by the College Board to score such essays.

The data were transformed using linear equating to ensure that the two test forms were of equivalent difficulty. With a continuous dependent variable (argumentative writing scores), a nominal independent variable (discourse method), and pre-test scores as a covariate, ANCOVA was used (after testing necessary assumptions) to compare the adjusted means of the dependent variable. The following chapter provides a summary of the findings as they relate to the research
question for this study: What is the difference in argumentative writing scores between English language arts students in 11th and 12th grade who engage in classroom discourse where the teacher focuses on procedural facilitation and English language arts students in 11th and 12th grade who engage in classroom discourse where the teacher focuses on linking ideas and pressing for reasoning?

**Data Transformation**

As mentioned in Chapter 3, linear equating was used to define correspondence between scores on alternate forms (Kolen, 1988; Livingston, 2004). Linear equating allows for the possibility that the “difference in relative difficulty between two forms is...variable along the score scale” (Kolen, 1988, p. 33). For example, some test forms might be more difficult for students who score on the low range than for students who score on the high range. The range of the original scores will be from 0 to 9, as described on the College Board rubrics (see, for example, "The AP English Language and Composition Exam," 2013).

The linear conversion is defined by setting the standardized scores on the two forms equal as shown below (Kolen, 1988):

\[
\frac{X_1 - \bar{X}_1}{S_1} = \frac{X_2 - \bar{X}_2}{S_2}
\]

Solving for \(X_1\) yields the following:

\[
X_1 = \frac{S_1}{S_2}X_2 + \left(\bar{X}_1 - \frac{S_1}{S_2}\bar{X}_2\right)
\]

where \(X_1\) is the pretest equivalent of a posttest score, \(X_2\) is the corresponding score on the posttest, \(\bar{X}_1\) is the global mean on the pretest (4.85), \(\bar{X}_2\) is the global mean on the posttest (4.62), \(S_1\) is the global standard deviation for the pretest (1.67), and \(S_2\) is the global standard deviation.
for the posttest (1.64). This equation yields the pretest equivalent of a posttest score, allowing the researcher to compare equated test scores instead of raw scores. This method of test equating augments reliability across different test forms (Kolen & Brennan, 2004), allowing for reliable computation of gain scores from pretest to posttest even in conditions where the difficulty of the test varies across the score scale. The data for this study were transformed using this method.

**Demographics of School, Population, and Sample**

The enrollment numbers and demographics for Community Leadership School change slightly from year to year, and even from month to month, but at the time of the study, 520 students were enrolled with females and Hispanic students overrepresented as compared to the district at large (see table 3). Students in 11th and 12th grade shared a similar demographic profile to those in the school; however, the general education population, which is also the population for this study, contained a higher percentage of female students, with the other demographic indicators being nearly equivalent. Despite this difference between the 11th and 12th grade student population and the population for this study, there was very little difference between the demographics of the sample and the demographics of the population for this study—11th and 12th grade students in general education classes who have not been place into substantially separate environments because of limited English proficiency or special education requirements.
Table 3

**School, Population, and Sample Demographics**

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Female</th>
<th>Male</th>
<th>Black</th>
<th>Hispanic</th>
<th>Asian</th>
<th>White</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Whole School</td>
<td>520</td>
<td>54</td>
<td>46</td>
<td>39</td>
<td>47.5</td>
<td>4</td>
<td>7.9</td>
<td>1.5</td>
</tr>
<tr>
<td>11th &amp; 12th Grade Students in Total</td>
<td>217</td>
<td>52.3</td>
<td>47.7</td>
<td>41.2</td>
<td>46.8</td>
<td>4.6</td>
<td>6.5</td>
<td>0.5</td>
</tr>
<tr>
<td>11th &amp; 12th Grade General Education Students (Population)</td>
<td>188</td>
<td>57.2</td>
<td>42.8</td>
<td>42.3</td>
<td>46.5</td>
<td>4.8</td>
<td>5.9</td>
<td>0.5</td>
</tr>
<tr>
<td>11th and 12th Grade Students in the Study (Sample)</td>
<td>115</td>
<td>60.8</td>
<td>39.1</td>
<td>41.8</td>
<td>48.7</td>
<td>4.4</td>
<td>5.2</td>
<td>0</td>
</tr>
</tbody>
</table>

*Note.* All numbers listed above are percentages except for the numbers listed under N.

Prior to 2013, the school collected lunch forms from students, and information from these forms was used to determine not only whether students would receive lunch aid but also the income status of families; however, at the beginning of this academic year, the school joined a national program called the “Community Eligibility Option,” which provides federal aid to Boston Public Schools for offering free breakfast and lunch to all students ("BPS offers universal free meals for every child," 2013). This is a boon for Boston students, including those at the research site, but the shift does mean that the school no longer has detailed income information on their families. Despite this fact, it is worth noting that low-income students have historically been overrepresented at CLS, with 75% or more being classified as such (see Table 2).

Table 4 indicates composite argumentative writing scores (out of 20) on a state standardized test, namely, the long composition on the Massachusetts Comprehensive Assessment System (MCAS), where students are asked to write arguments to support claims.
using reasoning and evidence. Although this test is administered in the spring of the 10th grade year, it does help paint a picture of the students’ baseline writing scores on standardized tests as they become upperclassmen. Because the students in the study are juniors and seniors, they do not all take another argumentative writing prompt that would provide similar data until their senior year (after the beginning of the study). As shown in Table 4, the school and district perform similarly on this argumentative task. The sample slightly outperforms the district. This is not particularly surprising because, unlike the school and district populations, the sample is composed solely of students in general education English classes at Community Leadership School. Nevertheless, the state still outperforms the school on this task. Data were not available for the general education populations of the school, district, or state.

Table 4

<table>
<thead>
<tr>
<th></th>
<th>Sample</th>
<th>School</th>
<th>District</th>
<th>State</th>
</tr>
</thead>
<tbody>
<tr>
<td>11th and 12th Grade</td>
<td>13.6</td>
<td>13.3</td>
<td>13.2</td>
<td>14.3</td>
</tr>
<tr>
<td>Students</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fidelity to the Discursive Method

Both participating teachers engaged in both procedural facilitation and facilitation that focused on linking ideas and pressing for reasoning. Each teacher followed a checklist to verify his fidelity to the method (Johnson et al., 2006), and the initial lessons of each teacher were
recorded to assess fidelity. In all cases, both teachers followed the checklists precisely (see Appendix A for checklists).

**Interrater Reliability**

Regarding students’ argumentative writing scores, the primary scorer was the author of this study, who has been trained and employed by the Educational Testing Service to score essays such as these for the College Board. A second ETS trained reader scored 30% of the essays, and 78% exact agreement was obtained between the two raters. Cronbach’s alpha coefficient was .94, indicating high internal consistency of the scores (Salkind, 2010). In addition, with normally distributed scores (see assumption checking below), Pearson’s correlation coefficient was .88, suggesting a high level of the rating tendency between the two raters’ scores (Salkind, 2010; Stemler, 2004).

**A Priori Power Analysis**

The statistical software G*Power allows users to calculate a priori statistical power for ANCOVA by inputting effect size, error probability (α), statistical power (1 – β), degrees of freedom, number of groups, and number of covariates. Essentially, by setting the probability of Type I error at 0.05 (that is, the probability of rejecting the null hypothesis when it is actually true) and the probability of Type II error at 0.2 (that is, the probability of accepting the null hypothesis when it is actually false), a sample size can be arrived at that makes the study worth undertaking. In essence, α and β have an inverse relationship such that decreasing the likelihood of Type I error increases the likelihood of Type II error and vice versa. As such, a balance must be struck between α and β. Keeping α at p = 0.05, β can be decreased (and statistical power increased) by applying a stronger treatment or increasing the sample size (Tabachnick & Fidell,
In this study, applying a stronger treatment was not consistent with a review of the literature and increasing the sample size was not practically possible given that all teachers in the school who taught more than one 11th or 12th grade English class were already participating.

Using G*Power before the study to calculate this necessary sample size, a sample size of 90 was necessary to achieve statistical power of 0.8 with \( p = 0.05 \), 1 degree of freedom, 2 groups, 1 covariate, and medium effect size (Mayers, 2013; Tabachnick & Fidell, 2007a). As such, the planned sample size for this study (130) was sufficient to achieve a high degree of statistical power. Even with a relatively high attrition rate, the sample size was likely to stay above 90. In reality, the sample size was 115, meeting the standard for an a priori power of 0.8 or higher. Essentially, this means that given the conditions listed above, the statistical power calculated before the study was high enough to make the study worth pursuing (Tabachnick & Fidell, 2007b).

**Testing Assumptions**

A series of assumptions must be met before running ANCOVA. A comprehensive assessment of these assumptions was performed as follows:

**Approximately equal sample sizes—assumption met.**

Although the initial plan was to have a sample size of approximately 130, school-level factors—such as students switching schools, students switching classes, and chronic absenteeism—left a real sample size of 115 for this study. Using one-way ANCOVA requires the size of the treatment and comparison groups to be approximately equal (Tabachnick & Fidell, 2007b). With 54 students in the treatment group and 61 students in the comparison group, the sample size requirements for one-way ANCOVA are met (Tabachnick & Fidell, 2007b). Given
that all teachers who teach upperclassmen in more than one class section were already involved in the study, there was no way to increase sample size beyond 115.

**Absence of outliers—assumption met.**

With a pretest mean of 2.32 ($SD = 1.34$), a minimum score of 0, and a maximum score of 6, none of the data points for the covariate are more than three standard deviations from the mean. With a posttest mean of 3.59 ($SD = 1.45$), a minimum score of 1.16, a maximum score of 7.27, none of the data points for the dependent variable are more than three standard deviations from the mean (see Figure B1 in Appendix B for boxplot). Therefore, the absence of outliers assumption is met.

**Normal distribution of covariate and dependent variable—assumptions met.**

To appropriately use ANCOVA, the covariate and the dependent variable should be normally distributed across both groups of the independent variable (Mayers, 2013). The histograms for these data (see Figures B2 & B3 in Appendix B) show that the covariate and the dependent variable are both approximately normally distributed. In addition, the covariate and the independent variable meet the skewness and kurtosis criteria because all four of these values lie between -1 and +1 (see Table B1 in Appendix B). Visual (histogram) and quantitative (skewness and kurtosis) methods of assessing normality are both met, and therefore, the assumption of normality is also met.

**Linear relationship between covariate and dependent variable—assumption met.**

For the covariate to be appropriately included in the analysis, it must have a linear relationship with the dependent variable (Mayers, 2013). Using the test of linearity, the
relationship between pretest argumentative writing scores and posttest argumentative writing scores was linear at a statistically significant level, $F(1, 114) = 78.29, p < .001$. Since the probability of the test statistic is lower than our designated $\alpha$ value (0.05), the null hypothesis that there was not a linear relationship is rejected; therefore, the assumption of linearity is met.

**Correlation between covariate and dependent variable—assumption met.**

If there is not a reasonable correlation (with Pearson’s correlation coefficient between $r = 0.30$ and $r = 0.90$) between the covariate and the dependent variable, the covariate cannot be justifiably included in the statistical analysis, and thus ANCOVA cannot be used (Mayers, 2013). In the present study, the correlation between the covariate (pretest scores) and the dependent variable (posttest scores) is $0.638$, falling between the accepted values for $r$. Therefore, the assumption of a correlation between the covariate and the dependent variable is met.

**Independence of covariate—assumption met.**

In order to use a covariate in the statistical analysis, the covariate should not differ significantly across groups of the independent variable (Mayers, 2013). Because there are two groups, an independent samples t-test can be used to evaluate this assumption. There is no significant difference, $t(113) = 0.643, p = .522$, in argumentative writing pretest scores between students who eventually participated in procedural facilitation ($n = 61, M = 2.25, SD = 1.41$) and students who eventually participated in linking/pressing facilitation, ($n = 54, M = 2.41, SD = 1.27$). Therefore, the assumption of independence of the covariate has been met, and the pretest scores can be included as a covariate in the statistical analysis to reduce error variance (Mayers, 2013).
Homogeneity of regression slopes—assumption met.

Homogeneity of regression slopes suggests that there is not a different dependent variable/covariate slope in some cells of the design and that there is no interaction between the independent variable and the covariate (Tabachnick & Fidell, 2007b). If there are interaction effects between the independent variable and the covariate, that means that the relationship between the covariate and the dependent variable is different across different levels of the independent variable, and thus, ANCOVA cannot be used (Tabachnick & Fidell, 2007b). In order to assess whether or not the regression slopes differ significantly between these groups, a custom model will be built using IBM’s SPSS that assesses the interaction between the covariate and the independent variable (Mayers, 2013).

There is no significant interaction between the covariate (pretest scores) and the dependent variable (posttest scores), $F(1, 111) = .182, p = .670$, failing to reject the null hypothesis that there is no interaction effect. A significant interaction would have meant that the assumption had been violated, and the use of ANCOVA would be inappropriate because the correlation between the covariate and the dependent variable would differ between groups of the independent variable. Since in instead there is homogeneity of regression slopes, the assumption is met.

Equality of error variances—assumption met.

The variance between the groups should be approximately homogeneous. Levene’s test of equality of error variances will be used to test this assumption. With $p > 0.05$, there is a failure to reject the null hypothesis, meaning that the error variance of the dependent variable is approximately equal across both groupings of the independent variable, $F(1,113) = .309, p = .580$, and the assumption is met.
Running ANCOVA

As shown above, the data for the present study meet all of the assumptions for ANCOVA, including fulfilling all the necessary assumptions to include pretest scores as a covariate. As such, the results from ANCOVA are presented below.

Main effect before including covariate.

Prior to running ANCOVA, the main effect of the independent variable (discourse group) on the dependent variable (argumentative writing scores) should be explored before controlling for the covariate of pretest scores (Mayers, 2013). Table 5 suggests that argumentative writing scores are higher for the group that received linking and pressing facilitation, but the descriptive statistics do not indicate whether or not the difference is statistically significant.

Table 5

<table>
<thead>
<tr>
<th>Group</th>
<th>n</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linking and Pressing Facilitation</td>
<td>54</td>
<td>3.96</td>
<td>1.40</td>
</tr>
<tr>
<td>Procedural Facilitation</td>
<td>61</td>
<td>3.27</td>
<td>1.43</td>
</tr>
</tbody>
</table>

Running ANCOVA without the covariate of pretest scores (effectively ANOVA) indicates that the main effect of discourse on argumentative writing scores is statistically significant at the $p = .05$ level, $F(1, 113) = 6.765, p = .011$. Despite this significance level, the effect of pretest scores as a covariate on the error variance needs to be explored to confirm that the statistical significance holds when the covariate is included in the analysis.
Applying a covariate to reduce error variance.

Table 6 shows the estimated marginal means after applying pretest scores as a covariate. The estimated marginal mean of posttest argumentative writing scores for the group receiving linking and pressing facilitation has decreased very slightly, and the estimated marginal mean for the group receiving procedural facilitation has increased very slightly. Nevertheless, the standard error has decreased, increasing the likelihood that the difference between the two groups remains statistically significant (Mayers, 2013).

Table 6

<table>
<thead>
<tr>
<th>Group</th>
<th>n</th>
<th>Estimated Marginal Mean</th>
<th>SE</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linking and Pressing Facilitation</td>
<td>54</td>
<td>3.90</td>
<td>0.148</td>
<td>[3.60, 4.20]</td>
</tr>
<tr>
<td>Procedural Facilitation</td>
<td>61</td>
<td>3.32</td>
<td>0.139</td>
<td>[3.04, 3.60]</td>
</tr>
</tbody>
</table>

The purpose of including pretest argumentative writing scores as a covariate when running ANCOVA is to reduce error variance by controlling for prior ability in argumentative writing. When this is done, the main effect of discourse group (the independent variable) on posttest argumentative writing scores (the dependent variable) remains significant at the 0.05 level, $F(1, 112) = 8.056, p = .005$.

For the unadjusted main effect, the residual mean square was 2.001. After applying the covariate, the residual mean square was 1.185. Statistical significance is determined by the F-ratio, which can be calculated by dividing the explained variance by the error variance.
Therefore, by reducing the error variance, the F-ratio has increased, subsequently increasing the chances of finding a statistically significant result (Mayers, 2013).

**Effect size and observed power analysis.**

In ANCOVA, partial eta squared ($\eta^2_p$) is often used as a measure of effect size with designs that have non-independent cells. For example, in this study the students who did the pretest and the posttest ($N = 115$) were different from each other (between subjects effect), but every student did take both tests (within subjects effect), where the pretest score was included as a covariate. The $\eta^2_p$ statistic approximates the percentage of variability in argumentative writing scores that can be accounted for by the independent variable while controlling for the covariate (Tabachnick & Fidell, 2007a). The effect size for the study was small, $\eta^2_p = .067$ (Muijs, 2011). This means that 6.7% of the between subjects variance is accounted for by group assignment.

Regarding statistical power, an a priori power analysis was done to determine an appropriate sample size for this study (see above). Some researchers argue that *observed power* should also be calculated by considering the significance criterion, the actual sample size, and the observed effect size (see, for example, Mayers, 2013). Although there is some disagreement about the value of observed power (O'Keefe, 2007), it will be included here for the sake of propriety. Using G*Power to calculate observed power (given $\alpha = .05$, $\eta^2_p = .067$, $N = 115$, two groups, one numerator degree of freedom, and a single covariate) yields observed power of .813.

**Hypothesis Testing**

The hypothesis for this study is as follow: English language arts students in 11th and 12th grade who participate in discussions where the teacher focuses on linking ideas and pressing for reasoning will have significantly higher mean argumentative writing scores than English
language arts students in 11th and 12th grade who participate in discussions where the teacher focuses on procedural facilitation.

As shown above, after including pretest scores as a covariate in the statistical test ANCOVA, a statistically significant difference at the $p < 0.05$ level was found between argumentative writing scores for students who participated in discussions where the teacher focused on linking ideas and pressing for reasoning and argumentative writing scores for students who participated in discussions where the teacher focused on procedural facilitation, $F(1, 112) = 8.056, p = .005$. The effect size was calculated using partial eta squared, $\eta_p^2 = .067$. Therefore, despite the small effect size, this study provides evidence for (1) rejecting the null hypothesis—that there is no difference in argumentative writing test scores between the group who participated in procedural facilitation and the group who participated in linking/pressing facilitation—and (2) accepting the directional hypothesis—that there is a difference in test scores between these two groups.

Table 5

*Dependent Variable (DV) and Covariate (CV) Scores, by Discussion Group (IV)*

<table>
<thead>
<tr>
<th></th>
<th>Linking and Pressing Facilitation ($n = 54$)</th>
<th>Procedural Facilitation ($n = 61$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>$SD$</td>
</tr>
<tr>
<td>Posttest Argumentative Writing</td>
<td>3.95</td>
<td>1.40</td>
</tr>
<tr>
<td>Writing Scores (DV) (Unadjusted Means)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretest Argumentative Writing</td>
<td>2.41</td>
<td>1.27</td>
</tr>
<tr>
<td>Writing Scores (CV)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Summary

After finding high fidelity to the discursive method, finding high interrater reliability, discovering little difference in demographics between the population and sample, transforming the data using linear equating, and meeting all the necessary assumptions for ANCOVA, the null hypothesis was rejected, and the directional hypothesis was accepted. In essence, the study provides evidence that students who participate in discussions where the teacher focuses on linking ideas and pressing for reasoning have significantly higher mean argumentative writing scores ($M = 3.90$, $SD = 1.40$) than students in participate in discussions where the teacher focuses on procedural facilitation ($M = 3.32$, $SD = 1.43$), $F(1, 112) = 8.056$, $p = .005$, $\eta^2_p = .067$. 
Chapter 5: Discussion of Research Findings

This study investigated the effect of classroom discourse on argumentative writing among English language arts students in 11th and 12th grades in a Boston public high school. After assessing students on initial argumentative writing skills with a pretest, implementing a discourse intervention, and reassessing students on argumentative writing with a posttest, one-way ANCOVA was used (after testing necessary assumptions) to compare the adjusted means of the dependent variable. The following chapter discusses the findings described in Chapter 4, considers possible explanations for the findings, evaluates the implications of the findings, and suggests further research.

Results and Discussion of Research Question

The research question for this study examines the difference in argumentative writing scores depending on the style of discourse implemented by the teacher. More precisely, what is the difference in argumentative writing scores between English language arts students in 11th and 12th grade who engage in classroom discourse where the teacher focuses on procedural facilitation and English language arts students in 11th and 12th grade who engage in classroom discourse where the teacher focuses on linking ideas and pressing for reasoning?

This study helps fill gaps in the literature insofar as it (1) provides clear and straightforward descriptions of the discourse interventions that could be followed by classroom teachers, (2) uses a standardized assessment tool designed by the College Board to measure growth in argumentative writing, and (3) focuses on high school upperclassmen in ELA classes. The descriptive statistics show that the mean on this College Board assessment for students who experienced discourse where the teacher focused on linking ideas and pressing for reasoning ($M = 3.95, SD = 1.40$) was significantly higher than for students who experienced procedural
facilitation where teachers encouraged discursive interactions but did not explicitly prompt students to articulate their reasoning or to link their ideas to those of others ($M = 3.27$, $SD = 1.43$), $F(1, 113) = 6.765$, $p = .011$

As has been discussed, random assignment was not possible in this study, and although attempting to control for prior ability in argumentative writing by including pretest scores as a covariate does not substitute for a true experimental design, it is worth noting that the mean scores for students who experienced linking and pressing facilitation remained higher even after the means were adjusted to control for the pretest on argumentative writing (linking and pressing facilitation—$M = 3.90$, $SD = 1.40$; procedural facilitation—$M = 3.32$, $SD = 1.43$, $F(1, 112) = 8.056$, $p = .005$, $\eta^2_p = .067$).

These results suggest that students in the population addressed in this study are more likely to improve their argumentative writing, at least in the short term, by experiencing discourse where the teacher plays an active role in rephrasing students’ thinking, asking about their reasoning, and helping students make connections among the ideas that arise during the discussion. It is, however, worth mentioning that across both intervention groups, the mean scores improved from the pretest; they just improved more for the linking and pressing group. This will be discussed further below.

**Implications**

These findings are particularly relevant in light of the original motivation for beginning this research—namely, that students both at the research site and in the country at large tend to struggle on argumentative tasks more than on other writing tasks (see Chapter 1 for discussion). For this reason, an intervention is essential that helps teachers support students in moving forward on this important skill, particularly an intervention where effects can be seen in a
relatively short period of time. After all, state and national standards require English teachers to improve students’ learning in a large variety of areas, including not only reading and writing in many genres but also speaking, listening, and media literacy ("Common core standards initiative," 2010). This means that teachers and students do not have the luxury of spending many months going into all topics and skills in great depth. Often, it is important to see growth on both informal assessments and on standardized assessments in a relatively short period of time. The particular discourse moves used in this study provide some promise (even for teachers who are already devoting large amounts of time to discourse) that by integrating moves that link students’ ideas and press for reasoning, teachers can augment student learning over and above what they are already seeing through their focus on classroom discourse.

Although the effect size is relatively small ($\eta_p^2 = .067$), this is unsurprising with an intervention that lasted two weeks, particularly one in which the comparison group also experienced large amounts of discourse. A larger effect would have been more likely if there had been a greater difference between the treatment and comparison groups (Tabachnick & Fidell, 2007b)—for example, if the comparison group did not engage in discourse but instead lecture. This was not done in this study because of ethical considerations, given that the literature review strongly suggested that discourse in general would have a favorable effect on students’ argumentative writing skills. Nevertheless, the effect size in this study is still of practical significance to educators. Integration of a particular set of discourse moves—over the course of two weeks—had the effect of augmenting students’ argumentative writing scores by more than half a point on a 9 point scale. In the most recent year that scores were available, the achievement gaps between essay scores of students at the research site (almost exclusively students in poverty and students of color) and the global mean on similar essays was .6 or less.
Although the magnitude of these gains would have to be maintained, the results of this study indicate that this particular discourse intervention shows the promise of narrowing historically resilient achievement gaps in argumentative writing.

**Implications for theory.**

Regarding the significance of the study to the theoretical framework, sociocognitive learning theory suggests, in general, that the discourse that occurs within a particular social context affects the cognitive processes of the participants in the discourse. In turn, the cognitive processes of the participants in the discourse affect their achievement (see Chapter 1). For the purposes of this study, as particular discourse moves are made in the social context of discussion, these moves will become more prevalent in students’ intra-individual thinking processes. As the moves become more prevalent in students’ cognitive processes, they will also become more prevalent in academic work similar to the discourse. Finally, as these moves become more prevalent in students’ academic work, their achievement will improve on standardized assessments that privilege such moves (see Chapter 1). Therefore, in accordance with the theory, a discourse that encourages students to use the argumentative moves privileged on standardized assessments will help students improve their scores on such assessments.

More specifically, the study has provided evidence—at least for the population in question—that inter-individual discourse where the teacher focuses on linking students’ ideas and encouraging students to articulate their reasoning is likely to increase the incidences of linking ideas and articulating reasoning in students’ intra-individual cognitive processes. As these incidences increase in students’ thought processes, these moves are also more likely to appear in academic work that calls on skills similar to those employed in the discourse, such as argumentative writing and argumentative class discussions. As these moves appear more
regularly in students’ argumentative writing, students’ argumentative writing scores will increase. In short, this study has provided confirming empirical evidence for this theory, regarding the effect that classroom discourse, as a social process, has on cognitive processes that influence student achievement (Langer, 1985).

**Implications for research.**

Regarding the scholarly significance of the study, foregoing research in the field has provided confirming evidence of sociocognitive learning theory (see Chapter 2), but prior to this study, no work on classroom discourse had met all of the following criteria: used externally-designed quantitative measures of achievement, clearly defined the nature of the classroom discourse, and evaluated the effect of the discourse on high school upperclassmen’s argumentative writing (Murphy et al., 2009). This study helped fill the gap in the research.

More specifically, research cited in Chapter 2 has shown that learning is a social phenomenon, and socialization often occurs through the discursive patterns that transpire in school (see, for example, Cazden, 2001). The present study has helped confirm that teachers can encourage particular discourse moves that, in turn, may help augment argumentative writing skills, and thus standardized test scores in argumentative writing.

It cannot be forgotten, however, that academic discourse is necessarily political insofar as it generally privileges particular forms of interaction and, therefore, defines what counts as learning and worthwhile knowledge. As discussed in Chapter 2, different discourses prevail in different communities, and these discourses can be reinforced or devalued within the predominant discourse of school—which tends to privilege the mainstream discourse that is valued on standardized tests. When students’ home discourse does not match the discourse of school, teachers often struggle to turn classroom interactions into literacy experiences for these
students (see, for example, Heath, 1978). In this sense, school failure can be seen, at least partially, as a mismatch between discourse expectations across communities, and as students have less literacy experiences in school, the perceived gaps in knowledge and skills can become real (Michaels et al., 2006). However, this study presents evidence that teachers can implement effective discursive moves using the checklists in this study, that students can learn these moves through teachers’ modeling, and that when given space to practice students can integrate these moves into their academic work, thus improving their test scores. Taken together, these conclusions provide promise that teachers can create a discursive space for students where their home discourse is affirmed and where they can learn to interact in the discourse of the dominant culture, if they so choose.

Implications for practice.

In this regard, the study has the potential to contribute to teachers’ practice. Like many public, urban high schools across the United States, Community Leadership School faces the dual challenge of empowering historically underserved students while simultaneously augmenting standardized test scores (Delpit, 1993; Nieto et al., 2008). With a heightened focus on standardized test scores in the contemporary educational environment, many teachers at CLS have responded to this complex challenge by focusing on explicit strategy instruction for test preparation as their primary pedagogical practice (Sineath, 2012). Such a practice often narrows the curriculum, inhibits students’ feelings of self-efficacy, and actually exacerbates test scores (Slomp, 2008).

Very few instructional approaches have been shown to improve students’ critical thinking, reasoning, and argumentation about text (Murphy et al., 2009; Soter et al., 2008), and the present study has helped to confirm linking ideas and pressing for reasoning as discourse
moves that augment these advanced skills. These discourse moves show promise for extending students’ thinking, despite being used much less frequently than discussion methods that focus more on teacher evaluation of students’ responses than on the links among the responses and the reasoning behind them (Cazden, 2001; Wolf et al., 2005).

The findings of this study are relevant to English language arts teachers who aim to improve their students’ argumentative skills and to those who focus on academic discourse in their classrooms. Substantiating claims with relevant evidence and warranted reasoning has always been a central project in the English language arts classroom, and with the adoption of the Common Core State standards, logical arguments are likely to remain one of the cornerstones of secondary English assessment ("Common core standards initiative," 2010). According to these standards, which have been adopted by 45 U.S. states, students will be required to engage in academic discourse with other students and with their teachers, and this study could help inform teachers about how to optimize this discourse to improve argumentative writing.

Implications for generalizability.

Generalizability (or external validity) is “the extent to which the findings of an experiment can be applied to individuals and settings beyond those that were studied” (Gall et al., 1996, p. 473). This section will discuss two kinds of generalizability: generalizability to a population (also called population validity) and generalizability to different environmental conditions (also called ecological validity).

Generalizability to a larger population.

In the present study, the sample size constituted more than 65% of the target population and was composed of approximately equal numbers of students in grades 11 and 12; therefore, it is reasonable to assume that findings from the study are generalizable to the experimentally
accessible population—namely, other students in the school who are upperclassmen and who are not in special education or English as a second language programs (Gall et al., 1996). However, generalizing to the population of the school as a whole or to all high school students in the Boston Public Schools is riskier because included in these larger populations are students with special needs and students who have been coded as having limited English proficiency. In addition, the present study looked only at upperclassmen; generalizing the methodology to lowerclassmen may be ineffective. As compared to the district at large, African-American students, females, and low-income students are over-represented in the sample; whereas, students in special education, students with limited English proficiency, and males are under-represented (see Table 2 for further demographic information).

To be clear, this study did not take into account the extent to which the instructional interventions interacted with student characteristics that are not demographic in nature. For example, different results may be obtained with students at different schools with different personological variables such as extraversion-introversion and anxiety levels (Gall et al., 1996). If the population differs appreciably from the sample, the generalizability of the findings are limited (Gall et al., 1996).

The most reasonable generalization of the study’s results is likely to students in general education within the school. It is also possible that general education students with similar demographic and achievement profiles may benefit from the discourse interventions described herein; however, the educational community would benefit from more research in this area. See below for further discussion.
Generalizability to other instructional contexts.

Ecological validity refers to the degree to which “the results of an experiment can be generalized from the set of environmental conditions created by the researcher to different environmental conditions” (Gall et al., 1996, p. 475). A study has high generalizability in this regard when the results can be obtained in a variety of environments by a variety of researchers. This requires an explicit description of the experimental treatment so that the study can be replicated (for a detailed checklist, see Appendix A).

In the design, the Hawthorne effect was minimized because all the teachers and students participated equally in the study, each receiving the same curriculum and resources. Novelty and disruption effects do provide a threat to the generalizability of this study because multiple teachers were involved, and it is impossible to assess whether the treatments will be novel or disruptive in each classroom. In addition, because the treatment length was relatively short, as compared to a full school year, the results of the study may not be generalizable to consistent use throughout the year. Nevertheless, the researcher did ask participating teachers how much their everyday instruction differed from the instruction provided in the treatments of the study and found that the discourse interventions were not particularly novel in that both teachers regularly employ both kinds of discourse used in the study (at least weekly and often daily). Despite this fact, the intervention did occur at the beginning of the school year, allowing for the possibility that the intervention was novel compared to the experience of the students in prior years. This may mean that the effect of the discourse could be attributable to the novelty of the intervention instead of to the instructional effects of the discourse itself. Although this novelty effect may help to explain that both groups improved from pretest to posttest, it would not seem to account
for the fact that the linking and pressing group improved more than the procedural facilitation group, thus providing a relatively minor threat to the stated implications of the study.

Concerns about experimenter effects were relatively low. Each teacher implemented both procedural facilitation and linking/pressing facilitation, prima facie controlling for the effect of the teacher; however, the teachers did know that they were participating in a study, and their preference for one treatment over another may have meant that they—consciously or subconsciously—altered their interactions with the students in such a way that yielded unintended consequences. Although a clear procedure was provided for each treatment and fidelity to this treatment was assessed using checklists and recordings, the potential risk did not disappear. After all, both teachers expressed a slight preexisting preference for linking and pressing facilitation over procedural facilitation. Neither teacher expressed confidence that one group would perform better than the other, but both teachers thought it more likely that the linking and pressing group would outperform the procedural facilitation group, as opposed to the opposite. Changing these conditions in different instructional contexts could lead to different results, even with demographically similar students.

The pretest used in the study may have posed a threat to generalizability across differing contexts by sensitizing students to the elements of the study (Gall et al., 1996). Though this is a serious threat on its face, the threat was minimized due to the normal practice of testing students on their argumentative writing in schools. In addition, all students across both treatments were affected equally by the pretest.

Because the posttests are analogous to essay exams that are given throughout students’ schooling, it is unlikely that the posttest served as a unique learning experience in and of itself that biased the study. In addition, all students across all treatments received the same posttest;
therefore, the effect should have been uniform. History and treatment effects may have interacted, creating unique conditions surrounding the study that make the effectiveness of the intervention more likely; however, given that progressive teaching practices have been in place in the school for a number of years for both of the teachers (including the use of discourse as a pedagogical technique), it is unlikely that the discourse methods being used in the study were seen as particularly innovative by the students. Both participating teachers reported that the interventions did not seem particularly innovative. In addition, neither participating teacher expressed particular disenchantment with current discourse methods in the school or in their classes. Of course, this style of self-report cannot eliminate these threats to generalizability; however, the teachers’ responses do seem to mitigate the possibility that the observed effects in this study could be attributable to an interaction of history and treatment effects instead of to the discourse interventions.

The discourse method might be limited in its effect on the dependent variable (argumentative skills) if the dependent variable was measured using other means. For example, if the dependent variable was measured using multiple-choice instead of essays, the results may not be generalizable; however, given that essay writing is an integral part of virtually all English exams, even if the results do not apply to other ways of measuring the dependent variable, they still provide information about an important segment of English instruction.

Finally, the time of measurement and treatment effects may have interacted in such a way that the effects of any of the treatments may deteriorate over time. In other words, results on posttests that occur immediately after the intervention may not persist as time passes. As a part of the normal battery of school-wide assessments, all of the participants in the study will take another version of the posttest as a midterm and as a final exam. For ethical reasons, now that an
initial effect has been seen for linking and pressing facilitation, it would be unfair to withhold this intervention from the students who were initially assigned to the comparison group.

Because all students will now be receiving versions of linking and pressing facilitation, it would not be valid to include students’ midterm and final exams as part of this particular study. However, this issue of possible deterioration of the treatment effects does point to an area for further research (see below for further discussion).

**Limitations**

In addition to the specific discussion of implications for generalizability above, the following section discusses further limitations of the present study.

Given that the author is at the center of both research and praxis in this study, research bias is of particular concern (Machi & McEvoy, 2009). In prior years, the researcher has implemented various discourse methodologies in his classroom, hypothesizing that they would be effective in augmenting student achievement. Improved test scores occurred after some of these interventions, and their effectiveness helped motivate the current study. For this reason, the researcher is particularly susceptible to confirmation bias where he may have only noticed the evidence that supported the hypothesis. This is one of the primary reasons that this study was quantitative, using externally designed, normed, and scored assessments produced by ETS. It was also one of the reasons that the essays were anonymized and scored blind.

Despite some of these basic checks against bias, participating teachers in the study—including the researcher—may have unwittingly engaged with students in such a way that unintentionally favored the treatment over the comparison. Students may have been responding to researcher expectations that they improve more in one group than the other. Alternatively, participating teachers may have introduced confounding variables (level of enthusiasm,
optimism, or energy, for example) that were not accounted for in the study but may have contributed to the main effect.

It is also possible that teachers in the study were simply more adept at facilitating discourse using a linking and pressing methodology than they were at procedural facilitation. This is unlikely because the procedural facilitation is in many ways less complex; nevertheless, the main effect could have more to do with the individual teachers skill in facilitating a particular methodology as opposed to implicit differences in the way the methodology influences student learning. Although this would be inconsistent with prior research and theory in this field, it is still worth considering.

Finally, both teachers in the study were White males with approximately 10 years teaching experience in high poverty, urban schools. Their ability to integrate the methodology into their teaching styles may have been more adept than could be expected from less experienced teachers in similar settings. Alternatively, more skilled teachers may have been able to more effectively integrate the methodologies, thus producing an even larger effect than observed in the present study.

The intention of having the same teachers facilitate both treatment and comparison groups was to minimize these teacher effects; however, without a larger-scale experimental design, it is impractical to control for all possible extraneous variables. Again, this points to an area for further research that will be discussed below.

In order to create a realistic comparison group that reasonably approximated status quo discussion in ELA classes, teachers in the comparison group procedurally facilitated discussion. This allowed for all students to participate in discussion throughout the study so that the social process of learning through group talk was active across all groups. This created a more
stringent requirement for the test of the effectiveness of the linking and pressing treatment while also meeting essential ethical standards. Nevertheless, this could have contributed to a smaller effect size than would have been observed had there been a larger difference between the treatment and comparison groups.

Many of these problems would be ameliorated, if not solved, with a true experimental design that included random sampling and random assignment of students and teachers to treatment and control groups. It would also be invaluable to have a sample size of teachers large enough that teachers could be treated as their own sample, thus assessing teacher effects in a way that was not possible or warranted in the present study. Sadly, the scope and cost of such a study makes its implementation improbable. In addition, random assignment of students to groups and to teachers for the sake of scholarship, as opposed to intentional assignment that optimizes their education, is unethical on its face in most settings.

Two limitations that would not be improved through experimental design are the challenge of assessing the quality of the classroom discourse and the length of the study. Chapter 2 of this study suggests that the quality of discourse is affected by the social interactions about content that occur within it (see, for example, Cazden, 2001). The quality of these interactions is very difficult to assess. Although some researchers have attempted to develop systems for assessing reasoning in conversation (see, for example, Resnick et al., 1993), these methods have neither gained widespread acceptance in the field nor are they practical for educators who are attempting to improve the quality of their classroom discourse. This is why a basic checklist was used in this study, in an attempt to simplify complex discursive moves without making them facile. Whether or not this goal was achieved is up for debate, but at present, there is no clearer pathway for educators who wish to implement discourse
interventions. Nevertheless, the methodology for facilitating discourse may have been simplistic, leading to an overly formulaic use of discursive moves. Smoother and more authentic integration of the discourse moves may have led to even greater gains for students in the treatment group.

Finally, the length of study is potentially problematic insofar as it only addresses students’ growth over a two-week period. On the one hand, as mentioned above, it is advantageous for researchers and practitioners to know that a discourse intervention as short as two weeks has a large enough effect to appear in quantitative research and to provide practical benefits for students. On the other hand, it is imperative to determine if these gains are lasting and to what extent further discourse is required for students to maintain the gains they originally achieved.

Areas for Further Research

Future research on discourse and learning will be built on (1) a foundation of sociocognitive theory (Langer, 1985; Vygotsky, 1978), (2) qualitative research that has explored the possibilities for discourse to influence student learning (Michaels, 1981), (3) mixed-methods research that has narrowed the focus to the relationship between particular discourse methods and student learning (Wolf et al., 2005), and (4) quantitative research that has applied more systematic measures of both teacher-behavior and student-achievement (Applebee et al., 2003; O'Connor et al., 2013; Topping & Trickey, 2007a, 2007b). Within this context, further research needs to acknowledge the political reality of this psychometric era—that educators are accountable to state and federal mandates for student learning, as measured by quantitative assessments. Qualitative and mixed-methods research has already shown that engaging students in discourse around controversial topics generally improves their inferential, argumentative, and
comprehension skills (Murphy et al., 2009), but very little work has been done other than the 
present study that addresses whether or not similar methodologies augment students’ 
standardized test scores, particularly in argumentative writing at the upper high school level. 
The use of broadly adopted standardized measures of student achievement (such as the SAT and 
AP exams), as was done in this study, would test the degree to which authentic argumentative 
skills transfer to improvements in test scores. 

However, more experimental studies are necessary that, through their use of random 
assignment, help minimize the kinds of extraneous variables that threaten the validity of this 
study. As mentioned above, such studies are often not practical, and sometimes not ethical. This 
puts scholar-practitioners in a quandary—where they need additional empirical evidence to 
support the use of discourse interventions but where the very nature of educational settings limits 
the extent to which this empirical evidence can be obtained. This leaves researchers and 
practitioners to replicate studies, such as this one, in slightly different contexts in the hope of 
compiling a body of evidence that supports particular interventions. For example, a study similar 
to the current one could be done across multiple high schools in a single district, employing 
many more teachers. Even if random assignment were not possible, such a study would help to 
persuasive, although imperfect, empirical support for a broadly applicable discursive 
intervention.

In addition, more research on discourse needs to be performed with special populations of 
students, including students with special needs and students who have been designated as 
Limited English Proficient. For students with special needs, researchers need to explore the 
effectiveness of discourse for students with particular learning profiles. Although sample sizes 
would likely be too small to use the research design employed herein, single subject research
(and alternative versions thereof) may provide evidence of the effectiveness of particular discourse moves for enriching the learning of students with special needs. A similar argument applies to students who have been designated as Limited English Proficient. To what extent would high school students who are learning English benefit from discourse generally, and linking and pressing more specifically? This question remains to be answered.

It is also important to reiterate that the population for this study was almost exclusively students of color who attend a school in an urban setting where the vast majority of students’ families are classified as low-income and where the majority of students score on the low-end of college-level assessments like the ones used in this study. This is important because, as suggested above, students at the research site may have responded differently to the discourse interventions in the study than students with different demographic profiles who attend other schools. For example, it may be the case that if students had been consistently pressing each other for reasoning and linking ideas during discourse prior to the commencement of the study, then procedural facilitation may have been more effective because students would have been modeling these discourse moves for each other, making the teachers’ role less important. In this regard, further research is essential using methodologies similar to this study but with students with different demographic profiles.

As mentioned above, further studies are necessary that examine the varying effectiveness of teachers who implement the same facilitation method. Such a study would require the sample size of teachers to be large enough to examine their results, as well as the students’ results. Although a study of this kind would require a substantive investment of resources, it would doubtless benefit the field.
Finally, as indicated in Table 1 (see Chapter 2), very little work has been done on the relationship between discourse and engagement for high school and college students. As such, a useful supplement to this study would be additional work that explores the relationship between the discursive moves used in this study and student engagement. Furthermore, a study is necessary that examines the relationship between student engagement in argumentative discourse and student achievement in argumentative writing.

Conclusion

In response to the continued struggles of Black and Hispanic students in poverty on argumentative writing tasks, the present study provides evidence that academic classroom discourse can help shift students’ cognitive processes in such a way that augments their argumentative writing skills and improves their test scores. Although there are a variety of well-researched pedagogical practices that employ discourse, they often fail to address (1) the practical challenges that educator’s face in attempting to implement an instructional intervention and (2) the realities of the current psychometric era—where in order to be found persuasive by the educational establishment, researcher’s work must show that it can augment standardized test scores that are used to assess student learning, teacher effectiveness, and school quality. The linking and pressing facilitation of discourse used in this study meets both of these criteria because it was described in a relatively simple checklist format and was shown to improve test scores on an assessment of argumentative writing designed by the College Board.

Ideally, researchers and practitioners will take up these results and begin scholar-practitioner collaboration that addresses some of the limitations mentioned above. In the meantime, English teachers, particularly in schools similar to the research site for this study, can
use the linking and pressing checklists included in this study with a measured degree of confidence that their use will help augment students’ argumentative writing skills.
References


O'Keefe, D. J. (2007). Post hoc power, observed power, a priori power, retrospective power, prospective power, achieved power: Sorting out appropriate uses of statistical power analyses. *Communication Methods and Measures, 1*(4), 291-299.


Retrieved from http://PAREonline.net/getvn.asp?v=9&n=4


## Appendix A

### Checklist for Procedural Facilitation

**Date:**

<table>
<thead>
<tr>
<th>Step Number</th>
<th>Description</th>
<th>Step Completed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Provide the class with the argumentative prompt for this week.</td>
<td>Yes = Y, No = N</td>
</tr>
<tr>
<td>2</td>
<td>Students respond to the prompt in writing without reading the sources.</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Students place their view on a continuum that corresponds to their viewpoint.</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Students discuss their viewpoint with a partner.</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Open class discussion to the whole class, asking students to share their perspectives. Call on students in the order they raise their hands, allowing students to take the lead in expressing their perspectives and in responding to each other.</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Break students into groups so that each group of three to five students is responsible for reading and explaining one of the six to eight sources that correspond to the prompt.</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Students present the central message of each source to their classmates and place the source on the continuum in a place that they think corresponds with the authors’ view.</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Ask the class if they wish to discuss the group’s explanation.</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Call on students in the approximate order that they raise their hands.</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>After each of the groups has presented their source and placed the source on the continuum, Group students by similar view, with three to five students per group.</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Give each group 15 minutes to discuss their view and prepare a presentation that makes a claim and substantiates that claim with evidence and reasoning.</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Each group shares their viewpoint.</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Open class discussion to the whole class and calls on students in the order that they raise their hands.</td>
<td></td>
</tr>
</tbody>
</table>

**Notes:**
**Checklist for Linking and Pressing Facilitation**

**Date:**

<table>
<thead>
<tr>
<th>Step Number</th>
<th>Description</th>
<th>Step Completed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Provide the class with the argumentative prompt for this week.</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Students respond to the prompt in writing without reading the sources.</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Students place their view on a continuum that corresponds to their viewpoint.</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Students discuss their viewpoint with a partner. Direct students to explain their view, support their view with evidence, and explain their reasoning.</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Open class discussion to the whole class, asking students to share their perspectives; however, in contrast to procedural facilitation, rephrase what students say, ask them to explain their reasoning, and ask them to link their ideas to evidence and/or to what has previously been said by their classmates.</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Break students into groups so that each group of three to five students is responsible for reading and explaining one of the six to eight sources that correspond to the prompt.</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Direct the first group of students to present the central message of each source to their classmates and place the source on the continuum in a place that they think corresponds with the authors’ view.</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Rephrases the students’ articulation of the central argument and asks if you have understood their thoughts accurately.</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>If the students respond negatively, ask them to explain again, or ask another group to help explain the first group’s thinking. Rephrase again.</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Once the students affirm that you have understood their articulation of the source’s central message, press for reasoning by asking the students to explain how they determined the central message and why they placed it where they did on the continuum.</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Once the students have explained, ask if the class has any questions. If they do, call on students in the approximate order that they raise their hands.</td>
<td></td>
</tr>
<tr>
<td>Step</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>For the second group and all subsequent groups, follow steps 7 to 11.</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>In addition, ask the students to explain the relationship between their source and the previous groups source by asking whether their author would agree with, disagree with, or qualify the view of at least one of the previous authors.</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>After students respond to this question, rephrase their thinking and asks if that rephrasing is accurate.</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Once the students affirm that the rephrasing is accurate, move to the next group and follows steps 7 through 14.</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>After each of the groups has presented their source and placed the source on the continuum, group students by similar view, with three to five students per group.</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Give each group approximately 15 minutes to discuss their view and prepare a presentation.</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>Each group shares their viewpoint that makes a claim and substantiates that claim with evidence and reasoning.</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>Open class discussion to the whole class and calls on students in the order that they raise their hands; however, in contrast to procedural facilitation, the teacher rephrases what students say, asks them to explain their reasoning, and asks them to link their ideas to the sources and/or to what has previously been said by their classmates.</td>
<td></td>
</tr>
</tbody>
</table>

Notes:
Appendix B

Tables and Figures Associated with Assumption Checking for ANCOVA

Figure B1. Boxplots for covariate and dependent variable.
Figure B2. Histogram for covariate (pretest argumentative writing scores).
Figure B3. Histogram for dependent variable (posttest argumentative writing scores).
Table B1

*Descriptive Statistics for Covariate and Dependent Variable*

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Min</th>
<th>Max</th>
<th>M</th>
<th>SD</th>
<th>Skewness Statistic</th>
<th>Skewness SE</th>
<th>Kurtosis Statistic</th>
<th>Kurtosis SE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pretest Scores</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Covariate)</td>
<td>115</td>
<td>0</td>
<td>6</td>
<td>2.32</td>
<td>1.34</td>
<td>0.611</td>
<td>0.226</td>
<td>-0.179</td>
<td>0.447</td>
</tr>
<tr>
<td><strong>Posttest Scores</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Dependent Variable)</td>
<td>115</td>
<td>1.16</td>
<td>7.27</td>
<td>3.59</td>
<td>1.45</td>
<td>0.335</td>
<td>.226</td>
<td>-0.443</td>
<td>0.447</td>
</tr>
</tbody>
</table>