Distribution Agreement

In presenting this thesis as a partial fulfillment of the requirements for a degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis in whole or in part in all forms of media, now or hereafter now, including display on the World Wide Web. I understand that I may select some access restrictions as part of the online submission of this thesis. I retain all ownership rights to the copyright of the thesis. I also retain the right to use in future works (such as articles or books) all or part of this thesis.

Shawna Dempsey April 11, 2020

"*Ain't* Ain't a Word": Stigma Against Southern Speech in the Classroom

by

Shawna Dempsey

Dr. Susan Tamasi
Adviser

Program in Linguistics

Dr. Susan Tamasi

Adviser

Robyn Clarke

Committee Member

Dr. Elizabeth Kim

Committee Member

2021

"*Ain't* Ain't a Word": Stigma Against Southern Speech in the Classroom

By

Shawna Dempsey

Dr. Susan Tamasi

Adviser

An abstract of
a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Arts with Honors

Program in Linguistics

2021

Abstract

"*Ain't* Ain't a Word": Stigma Against Southern Speech in the Classroom
By Shawna Dempsey

This research explores whether teachers' prior experience with a student's dialect influences grading outcomes. Specifically, I am interested in how negative language attitudes about Southern American English affect how teachers grade students who speak this variety as compared to "standard"-sounding students. This question stems from the nature of the education system as a language authority, which plays a crucial role in spreading standard language ideology (Lippi-Green, 2012). As such, studying the implicit biases that these authorities exercise toward students who speak with non-standard dialects when providing feedback is crucial for equity in the education system. In this study, participants were randomly assigned to one of two conditions: Southern accent or standard accent. Participants evaluated two student assignments, a student presentation and an essay. First, participants observed a student presentation with all variables held constant except the student's accent, which reflected the condition to which participants had been assigned. Participants then provided feedback regarding the presentation. Then, all participants evaluated and graded a student essay (identical between conditions) on an unrelated topic, which was explicitly attributed to the student presenter. The feedback and grades of each assignment were then compared between conditions. Moreover, grader comments were analyzed to further understand the grades administered. While the grades were not statistically different between conditions, Southern condition grades varied more widely than standard condition grades. This result demonstrates the lack of consensus that Southern condition graders had as compared to standard condition graders. These results are discussed in light of educator characteristics and research design. Practical implications are also discussed.

"*Ain't* Ain't a Word": Stigma Against Southern Speech in the Classroom

By

Shawna Dempsey

Dr. Susan Tamasi

Adviser

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Arts with Honors

Program in Linguistics

2021

Table of Contents

Chapter One: Introduction

"*Ain't* ain't a word, and I ain't gonna say it." I learned this phrase in an elementary school classroom along with countless other Southern-speaking peers. Even as I type, I am met with a red line to remind me that *ain't* ain't a word. In the pursuit of education, many Southern American English speakers learn that their speech is not acceptable in the classroom. More precisely, they learn that their speech is incorrect—full of errors and mispronunciations. As a result, these speakers are expected to abandon their language and learn the "correct" form that is taught in school. With this narrative in mind, I question the implications of speaking Southern American English for students in the education system.

The U.S. education system upholds and spreads standard language ideology, a set of beliefs that promotes an idealized form of a language and demonizes language variation (Lippi-Green, 2012; Milroy & Milroy, 2012). Standard language ideology postures a "standard"[1] form of a language as the "correct" form of that language, thereby subordinating other varieties of the language to the "standard" form. Ironically, no such form exists in practice as language variation and change are unavoidable characteristics of language. Even so, standard language ideology is widely accepted in the U.S. and other countries (Preston, 1999; Montgomery & Cramer, 2016). The U.S. education system is no exception; in fact, it perpetuates standard language ideology by disguising it as language instruction. For example, the U.S. education system teaches that one variety of English is "correct" while others, including Southern American English, are "incorrect." Consequently, students who speak different varieties become the target of

[1] As I explain in the following sentences, a "standard" form of a spoken language does not and cannot exist (see the Standard Language Ideology section for a more extensive discussion). Therefore, when I refer to a "standard" form, I am referring to an ideological construct, rather than a form that exists in practice. Accordingly, when I refer to this form, I refer to it using quotations to remind the reader that a "standard" form is merely conceptual. The only time that I do not refer to "standard" in quotations is when it is used to describe the research condition; I do this for clarity when discussing the research methods and results.

correction, and a failure to adopt the "standard" form eventually leads to consequences in the classroom and beyond. Thus, many students with enormous potential are stifled by standard language ideology and the negative language attitudes it promotes.

Beyond the classroom, Southerners face a myriad of stereotypes that paint Southern culture as ignorant, unintelligent, and bigoted. These negative presentations contribute to the negative language attitudes that everyday Americans hold about Southern American English. Language attitudes are any opinion (positive, negative, or neutral) that any person holds about a language or dialect. The negative language attitudes towards Southern speech are widespread and even held by some Southerners. These attitudes are reinforced, in part, by standard language ideology as well as negative media portrayals of Southerners (Lippi-Green, 2012). Ultimately, these negative attitudes compound the standard language ideology found in the classroom to undermine the legitimacy of Southern speech and culture.

In light of (1) the education system's role in spreading standard language ideology and (2) prevalent negative language attitudes towards "non-standard" dialects, evaluating language attitudes in the classroom with regards to instructor grading practices is crucial for educational equity. More expressly, studying teacher grading practices could illuminate whether negative language attitudes modulate graded feedback provided by instructors. The current study explores whether teachers' prior experience with a student's accent influences grading outcomes. Specifically, I am interested in how negative language attitudes about Southern American English affect how teachers grade students who speak this variety as compared to "standard"-sounding students. The purpose of this study is to quantify this effect through teacher grading practices and determine whether standard language ideology taints the grading process.

Standard Language Ideology

Standard language ideology (SLI) describes a system of beliefs that promotes a "correct" way to use language pedaled as the "standard" form that should be adopted by all the language's users (Lippi-Green, 2012; Milroy & Milroy, 2012). In practice, a homogenous, idealized "standard" form cannot exist, however, because language varies to meet the needs of its speakers (Rickford, 1999; Tamasi & Antieau, 2015). Nevertheless, everyday folks and dominant institutions alike unquestioningly embrace SLI as the law of the land.

In effect, SLI teaches children, adults, native and non-native speakers that their language is sublinguistic if it does not follow the rules of the "standard" form. For that reason, this ideology operates in direct contradiction to the linguistic fact that speakers are experts in their native language. Thus, adoption of SLI requires a trade-off that often goes unquestioned: native speakers must exchange their own language authority for the superficial standards of SLI (Lippi-Green, 2012). This trade-off is the first step in Lippi-Green's (2012) language subordination model, which outlines how speakers of a language become indoctrinated by SLI. First, "language is mystified," thus becoming something that speakers only understand upon instruction from "experts." These experts, or "language authorities," claim the authority that everyday speakers surrender in their adoption of SLI. Language authorities teach in classrooms, write dictionaries and news columns, and appear on media broadcasts. They use their power to trivialize "non-standard" varieties and threaten all those who speak them. In turn, speakers of "non-standard" varieties quickly learn that they either conform to the "standard" or reap the threatened consequences of non-conformity: poor grades, poor reputation, and no future. Ultimately, adopting SLI requires speakers to reject their own language, which has intricate ties to their identity (Lippi-Green, 2012; Tamasi & Antieau, 2015). In essence, these authorities demand that

speakers cast aside their own language and thus deny a part of who they are—all in the name of SLI but under the guise of language instruction with the promise of a successful future.

Unlike countries like Spain and France, which have state-established language standardization (e.g., La Real Academia Española and Académie Française, resp.), the United States does not formally recognize a "standard" English. Rather, "Standard American English" eludes formal definition. In fact, it is more readily defined in terms of what it is not: Not Southern, not associated with Black, Indigenous, and People(s) of Color (BIPOC), not lower class, not foreign-accented (Lippi-Green 2012). Importantly, these varieties are not inherently "non-standard." Nevertheless, when relevant only as counter examples, dialects from these communities are stripped of their inherent dignity as valid language varieties. This denigration is not simply a by-product of SLI; these negative views must be upheld to perpetuate standard language ideology. Thus, SLI necessarily authors and enforces linguistic prejudice to maintain its power.

As a result, SLI negatively perverts the perception of many English dialects, such as African American Language, Southern American English, and several others, as well as dialects of other languages (Rickford, 1999; Preston, 1996; Preston, 2000: Cramer & Montgomery, 2016). To facilitate this perversion, SLI preys on language variation. Language variation describes the process where a language takes on different forms that may be grammatical in one dialect but not another. For example, *They was at the game last night* is a grammatical utterance in Southern American English but may not be grammatical (violating subject-verb agreement) in another dialect. Instead of acknowledging the legitimacy of this variant, SLI prescribes that the phrase is incorrect, should never be uttered, and thus denigrates anyone who uses phrases with that grammatical construction. In short, SLI assigns a negative value to the feature, thus marking

the feature as “non-standard.” The markedness of a feature increases its salience, making it readily identifiable as “other.” This value is then perpetuated over and over for each “non-standard” feature. Consequently, the dialect--Southern English in this example--is regarded as incorrect, uneducated, and unworthy to be spoken. To be clear, language varies, that is the same language can exist in different forms in different spaces and places for different speakers, and to expect every dialect of English to have the same rules undermines that unavoidable characteristic.

Language Attitudes

Language attitudes are any opinion (positive, negative, or neutral) that any person holds about a language or dialect. Language attitudes are complex—comprised of biases from personal experience and preference, standard language ideology, and so forth. Everyone has language attitudes, but how these attitudes potentially beget prejudice cannot be overlooked. First, everyday folks oftentimes do not realize that these language attitudes are unfounded. For example, Southern American English is as logical and grammatical as any other language or dialect. This fact, of course, contradicts what most Americans (and students around the globe) learn in the classroom: standard language ideology. As a result, language attitudes often reflect the narrative of standard language ideology—upholding one correct variety while demonizing other “non-standard” varieties.

Using various methods, many social scientists have researched the language attitudes that non-linguists hold about different languages varieties (Lambert, 1967; Giles et al., 1992; Preston, 1996; Preston, 1999; Purnell, Idsardi, & Baugh, 1999). Lambert (1967) found that respondents exhibited group biases towards speakers of different languages (here, English and French) using

a matched guise test, one of many ways to measure language attitudes. A matched guise test is a research method where the same speaker records an identical script using two different language varieties (both of which the speaker speaks fluently). This method was an important advancement in language attitude research as it provided controls for other characteristics such as gender, physical appearance, and age. In doing so, the listener's perceptions are more readily attributed to language.

Another valuable method for uncovering the language attitudes of non-experts is perceptual dialectology (Preston, 1996; Preston, 1999; Montgomery & Cramer, 2016). One popular method in perceptual dialectology is the Draw-a-Map task, where non-linguists use a non-labeled lined map of a specific country or region to indicate places where they believe their language to be spoken differently. Then, they rate these areas on scales of "correctness" and "pleasantness," which correspond to linguistic "status" and "solidarity" respectively. There are several variations of these methods (see Montgomery & Cramer, 2016 for an overview), but the purpose remains to measure non-experts' perceptions of language use.

Southern English as 'Other'

While all language varieties, including Southern American English (SAE), are rule-governed, Southern speech is largely seen as incorrect, incomprehensible, and even illogical. These negative stereotypes are reflected in everyday Americans' language attitudes, which have been recorded through studies in sociolinguistics and perceptual dialectology (Lippi-Green, 2012; Preston, 1996). Southerners face these negative language attitudes in everyday settings, such as school (Dunstan & Jaeger, 2015) and even while watching Disney movies (Lippi-Green, 2012). According to perceptual dialectology studies in the U.S., the South is the most highly

identified regions in the Draw-a-Map task by non-experts (including Southerners themselves). For example, Preston (1996) found that 94 percent of respondents from Michigan had drawn some version of a "linguistic South" while the next most frequently drawn region (the North) was only identified by 61 percent of respondents. Preston suggests that this distinctiveness is a result of negative language attitudes towards the Southern dialect. The data corroborates this claim as most respondents from Michigan and Indiana rate the South as the least correct English on their maps. Moreover, they also rate the South low on pleasantness. Both of these respondent groups tend to rate their own English as correct and pleasant; for example, Michiganders rate speech in Michigan as both most correct and most pleasant. On the other hand, Southerners show a different pattern: they do not rate their English as correct (though they do not rate it as low as Michiganders), but they do rate their English as most pleasant. Preston (1996) asserts that this pattern is a result of linguistic insecurity in that Southerners do not evaluate their own English as correct, thereby affirming its low status, but do maintain a sense of solidarity through rating it as highly pleasant. This pattern echoes the damaging effects of SLI on "non-standard" speech communities—it not only attacks the overall perception of the Southern dialect but undermines its speakers' view of their speech as well.

SLI in the Classroom

The education system is a language authority which perpetuates standard language ideology. In Lippi-Green's (2012) language subordination model, she demonstrates how speakers of a language become indoctrinated by SLI. A key step in this model is that "authority is claimed" (2012). Language authorities take many forms in this model, but the authority

pertinent to this research is that of the education system. The education system is a large source of prescriptive attitudes, and it privileges certain dialects over others (Lippi-Green, 2012).

According to the National Council of Teachers of English (NCTE), a major outcome of English teaching is to promote literacy that motivates responsible citizenship (1998). In promoting literacy, American English classrooms teach "academic English" to their students. This form, though idealized as a "standard" form, is not homogenous but depends heavily upon who is teaching and where they are teaching. Academic English most highly reflects the English of white, upper-middle class speakers. Consequently, students of this demographic are more likely to succeed in learning, or rather recognizing, its grammatical conventions. Conversely, students who speak different varieties may struggle with learning this new form (Charity Hudley & Mallison, 2011). This disparity fosters the opportunity for linguistic prejudice in the classroom.

In the English classroom, teachers instruct students how to use academic English in reading and writing. Yet, many teachers do not limit their instruction to the written medium. Consequently, they begin to police the speech of students who speak differently. This policing creates a linguistic power struggle: Students' who speak different Englishes become deviants who refuse to follow the rules instead of speakers of robust but different dialects (Lippi-Green, 2012). In effect, SLI demeans any dialect that deviates from the "standard" form, thus sterilizing the linguistic landscape and privileging students that can and do conform to the "standard." As a result, SLI has negative effects for students who speak stigmatized dialects.

To be clear, language instruction does not necessitate SLI. Teachers can instruct their students about linguistic forms without undermining the students' ability to speak their native variety of English, which may deviate from the "standard" form by few or many linguistic

features. In fact, Charity Hudley and Mallinson (2011) recognize the value of teaching students about "standardized forms of English." Moreover, Labov (1995) used linguistic data to create a reading program, The BRIDGE Program, that aids speakers of African American Language (AAL) to learn to read in their native dialect and then slowly transition to reading in "standard" English. These linguists acknowledge that many modes of reading and writing require a strong command of academic English. This recognition must be distinguished from an acceptance of so-called "standard" features as the only academic variety of English. The idea that one variety of English is more worthy of academia than another is a fruit of SLI.

Accordingly, many linguists challenge the exclusivity and ubiquity of "standardized forms" in the classroom, namely through critical language pedagogy (Alim, 2005; Godley et al., 2006) and the multicultural education movement (Charity Hudley and Mallinson, 2011). Critical language pedagogy seeks to challenge the systems that teach and enforce standard language ideology by equipping students with linguistic knowledge and encouraging them to be critical about the language attitudes that they encounter. With less emphasis on challenging systems, the multicultural education movement promotes dialectal differences and acknowledges the equality of all language varieties. In both practices, teachers would recognize, rather than demonize, dialectal variation in the classroom through methods such as grading for transfer and the explicit teaching of language variation. These methods have been shown to increase student's awareness of SLI and ability to challenge its harmful implications (Godley & Minnici, 2008). Use of these methods begin with teacher preparation through their own learning about dialectal differences, specifically learning about the linguistic structure of "non-standard" varieties of English (Alim, 2005; Godley et al., 2006; Charity Hudley and Mallison, 2011).

Yet, many linguists have observed that acquiring this knowledge does not necessarily lead to a change in teaching practices (Wheeler, 2019; Metz, 2019; MacSwan, 2020). Wheeler (2019) found that learning linguistically informed teaching strategies fostered in her students a new sense of empathy towards speakers of "non-standard" dialects. Even so, her students (both current and future educators) were resistant to implement new grading strategies designed to validate the home dialect while continuing to teach the school form. In other words, these educators were given the tools to grade for transfer—a method that recognizes the linguistic features of a "non-standard" dialect and in so doing offers a "translation to school English" rather than marking the feature as incorrect. Rather than implementing this practice, Wheeler finds that future educators overwhelmingly report that teaching grammar intimidates them, and that grading grammar is easier when they rely on their intuition. Metz (2019) examined teachers' language ideologies and found that though their personal opinions are more "counter-hegemonic" (or more linguistically-informed), they hold more hegemonic (or more SLI-informed) views when considering societal perceptions and the dominant school narrative. In other words, on the individual level, teachers tend to hold more linguistically informed opinions about the role of language in the classroom—specifically when it comes to valuing individual student dialects as well as considering their personal role as educators. On a macro level, however, teachers tend to subscribe to the standard language ideology that holds greater society captive to negative, and oftentimes prejudicial, views about more "vernacular" language use—recognizing and subsequently believing that a pivotal role in language education is to promote academic English as the correct and valid alternative to the "incorrect" forms of English that children bring into the classroom.

Studying teacher grading practices can reveal the impact of SLI on student success. A wide breadth of education research has demonstrated that student characteristics influence the grades that teachers administer (Malouff et al., 2013; Klapp Lekholm & Cliffordson, 2008; MacCann & Roberts, 2013). In fact, Malouff and colleagues (2013) found that instructors who have positive prior experience with a student are more likely to give that student a good grade as compared to a student with whom they have had negative prior experience. What's more, one way that social scientists have quantified language attitudes is looking at how instructors grade students who speak "non-standard" language varieties as compared to their "standard"-speaking peers (Crowl & MacGinitie, 1974; Shepard, 2011). Both Crowl and MacGinitie (1974) and Shepard (2011) found that non-white speakers were graded lower for identical oral responses when compared to white speakers when judged solely on auditory input. These findings suggests that educators demonstrate negative biases against students based solely on linguistic input. With similar methods, a recent study conducted by Fletcher (2020) explored instructor evaluations of students with Southern accents compared to "standard"-sounding students when providing "oral answers to academic questions." Overall, instructors very highly rated answers in both conditions, so there was not a statistically reliable difference between the two conditions. Though no statistically significant results were found, it is important to note that this question is of interest to researchers in the field of sociolinguistics and education.

The current study explores the effect of language attitudes in teacher grading practices. In particular, I examine whether instructors' language attitudes result in a systematic bias against "non-standard" dialects, specifically focusing on Southern American English, in grading practices. Specifically, I inquire whether students with Southern accents receive lower grades than students with "standard" accents, despite equal performance. Therefore, the participants

were randomly assigned to either the Southern accent condition or the “standard” accent condition and asked to evaluate two assignments: 1) an identically-scripted student presentation in which the student’s accent varies (either Southern or “standard”) and 2) a student essay that will be identical between conditions and written to approximate standard writing conventions.

If negative language attitudes bias the grading process, then I hypothesize that the Southern-speaking student will receive lower grades on the presentation despite an identical script shared with the “standard”-sounding presenter. I expect that this bias will also be observed in the participants’ evaluation of the student essay. The grader’s teaching experience, educational experience, place of origin, and experience with English will also be taken into account.

This research is the first to examine the grading practices of graduate students and their intersection with the Southern dialect. From this study, the academic community can test its inclusivity and expectations of its Southern students, specifically in the setting of a private Southern university.

Chapter Two: Methods

My research question is whether students with a Southern accent are systematically graded more negatively than students with a "standard"-sounding accent despite equal performance. To answer this question, participants were randomly assigned to one of two conditions: Southern accent or standard accent. Participants evaluated two student assignments: an oral presentation and a 3-page essay. First, participants viewed a student oral presentation with all variables held constant except the student's accent, which reflected the condition to which participants were assigned. Participants then provided feedback regarding the presentation. Then, all participants evaluated and graded a student essay (identical between conditions) on an unrelated topic, which the researcher explained to be the work of the presenter for another class. The feedback and grades of each assignment were then compared between conditions. I hypothesized that graders in the Southern condition would assign lower grades than graders in the standard condition.

Participants

I recruited a convenience sample of 15 graduate students currently enrolled in a private Southern university via their university affiliate email and through department listservs. All participants had to be at least 18 years of age, enrolled in graduate school, and have teaching/grading experience or preservice training. These requirements strengthened the external validity of the study as the participant pool consisted of experienced graders who work in academia. Upon recruitment, the participants were randomly assigned into the two test conditions: the Southern accent condition (60%, $n = 8$) and the standard accent condition (40%, $n = 7$).

Table 1 describes participants' teaching experience as well as their future plans for teaching. All participants reported that they had teaching experience before and/or during their program, so each participant had at least some teaching experience before participating in this study. Additionally, 14 participants reported that they were considering teaching or planning to teach after completion of their program and that this plan may include teaching writing. Therefore, this sample provides some insight into graduate students that plan to teach after their programs.

Table 1

Participant Experience with and Plans for Teaching

Enrollment Status	Sample [a]	
	n	%
Pre-program	11	73.3
Current program	13	86.7
Post-program plans	14	93.3
Student writing in post-program plans [a]	14	93.3

[a] Reflects the number and percentage of participants answering "yes" or "maybe" to this question out of $N = 15$.

Participants reported the department of their current program. Table 2 describes the department breakdown by whether the department is in the humanities or not. I collected this data because the student presentation and essay pertained to content more readily identifiable as humanities-based coursework (see Materials and Measures), and their department may mediate their ability and/or confidence in grading these assignments.

Table 2

Categorical Breakdown of Participant Department

Department	Full sample	
	n	%
Humanities	4	26.7
Non-humanities	11	73.3

Table 3 describes the language experience of each participant by mapping the area that the participants "call home" (see Appendix A) onto the regional dialect map of the United States from the Atlas of North American English Dialects (Labov, Ash, & Boberg, 2005; pp. 148). To be clear, this association cannot determine the dialects that the participants speak; rather, it paints a picture of the speech to which the participants have most likely been exposed. Additionally, participants who identified as international reported that their "primary" languages (see Appendix A) were Mandarin (n = 2), Portuguese (n = 1), and Turkish (n = 1). Ten of 11 participants that identified as domestic learned English as their "primary" language, and the majority learned English at home (n = 8; 80% of domestic participants) while the others first learned English at school (n = 2). The remaining participant that identified as a domestic student indicated that they learned Spanish as their primary language and then learned English at school.

Table 3

Predicted Language Experience of Participants by Region

Dialect Region	Full sample	
	n	%
North	6	40.0
South	2	13.3
West	1	6.6
Midland	2	13.3
International	4	26.7

Materials and Measures

Demographic Survey

Participants completed a demographic survey upon indicating verbal informed consent (see Appendix A). The survey comprised of questions regarding graduate school enrollment status, year in graduate program, department, age, gender, and race/ethnicity. Moreover, the survey included questions about participants' language background: where they call "home," how long they have lived in Atlanta, whether they identify as an international student, their dominant/native language, and where they learned English (i.e., home, school, or other). Finally, the survey contained questions to assess participants' past and current teaching experience. It also measured participants' intention to teach upon completion of their program; whether those plans involve student writing; and how confident they feel when grading student writing.

Student Presentation

The presentation discussed the life and work of Simone de Beauvoir, a 20th century philosopher, and was approximately two minutes in duration. The presentation was designed to approximate the level of a first-year university student in an introductory Philosophy course. The presentation was a PowerPoint file consisting of pre-recorded audio, visuals, and transitions. The student presenter was only identified through audio stimuli as to control other confounding variables.

The student presentation utilized a matched guise paradigm (e.g., Lambert, 1967) in order to present the independent variable: student accent. The matched guise paradigm eliminates interspeaker variation in testing language perception by employing a single speaker that speaks

both languages (or dialects) fluently. In holding all variables constant besides language use, the listener's perceptions are more readily attributed to language.

A 20-year-old white female from Alabama recorded three voice samples using a pre-determined script. I shared these voice samples with three linguists, who indicated how Southern the speaker sounded on a scale from 1 ("Not at all") to 5 ("Very much so") for each of the three recordings. There was a consensus on which recording sounded the most Southern and which sounded the least Southern. After narrowing down to two recordings, I performed a linguistic analysis of both audio recordings to measure the validity of dialectal differences between the Southern condition and the standard condition using type/token ratios. I examined three features of Southern American English: /ai/ monothongization, the pen-pin merger, and g-dropping. These features were chosen because they are three of the most marked and most noted features in Southern American English phonology.

Overall, the Southern recording had more Southern features than the standard recording, where the Southern recording had 75% of possible tokens while the standard condition had 46%. For /ai/ monothongization, the Southern recording had 50% of possible tokens while the standard recording had 7%. For g-dropping, the Southern recording had 50% of possible tokens while the standard recording had 0 instances of g-dropping. For the pen-pin merger, both recordings had 100% of possible tokens. Moreover, the Southern Vowel Shift was impressionistically recognized as being more prominent in the Southern recording than the standard recording. Because these features were chosen due to their notability as Southern features, the 100% use of the pin-pen merger in the standard condition recording may have interfered with the manipulation of the matched guise. Rather than having a "standard"-sounding recording, the recording may be better classified as less-Southern sounding. A small group of linguists did vote

that the standard recording was the least Southern; therefore, it may not be the best representative of a "standard" American English accent. Even so, the two recordings are different from one another, so finding an effect is still possible.

Student Presentation Feedback

Participants graded the student presentation using the student presentation feedback form (designed for this study; see Appendix B). Participants rated the student presentation on four dimensions to measure their impressions of the student presentation: content, clarity, delivery/timing, and audio/visual materials. Participants rated each dimension on a 5-point Likert scale from 1 ("Not at all") to 5 ("Very much so"), where higher scores indicated better student performance on a given dimension. Participants then gave an overall impression of the presentation by assigning a grade on a pass/fail basis. The feedback form also contained an assignment prompt in order to guide participants in the grading process.

Student Essay

Participants in both conditions read an identical 3-page close reading essay that examined "Response to Sir Filotea" by Sor Juan Ines de la Cruz. I wrote the essay in academic English approximating the level of a first-year college student writing for an introductory English class. Participants were informed that this essay was written by the same student who gave the presentation.

Student Essay Rubric

Participants graded the student essay using the student essay rubric (adapted for this study; see Appendix C). Participants rated the student essay on five dimensions to measure their impressions of the student essay: evidence, analysis, formality, organization, and mechanics. Participants rated each dimension on a 4-point Likert scale from 1 ("Needs Improvement") to 4 ("Excellent"), where higher scores indicated better student performance on a given dimension. Participants then gave an overall impression of the paper by assigning a letter grade corresponding to the 4.0 GPA scale. The essay included an assignment prompt before its heading in order to guide participants in the grading process.

Design

My research design was informed by research on the halo effect in grading (Malouff et al., 2013). The halo effect predicts that previous experience, positive or negative, will shape future interactions in the same manner. Resultantly, this effect would allow for bias in the subjective grading of open-ended assignments. In a study conducted by Malouff and colleagues (2013), university professors and teaching assistants who had just evaluated a "good" oral presentation subsequently graded an identical written assignment (attributed to the presenter) significantly higher than those who had just evaluated a "bad" oral presentation. These findings suggest that previous positive experience with students can bias grading in their favor while negative experience can hinder them from future success.

In tandem with standard language ideology, I decided to examine whether experience with a student's dialect produces the same effect. Consequently, I adapted the research design used in the Malouff et al. (2013) study to answer this question. The current research maintained a

between-subjects experimental design, but the student presentation varied on the basis of student accent (Southern versus "standard") rather than level of preparedness. An impressionistic linguistic analysis was conducted to demonstrate the difference between the Southern and "standard" accent (reported above). Importantly, graders still graded an identical, unrelated written assignment, which was attributed to the student presenter.

Procedure

The researcher recruited participants directly via their university affiliate email or through department listservs. Participants confirmed that they were 18 years of age and enrolled in graduate school prior to scheduling the video meeting. The researcher invited eligible recruits to a scheduled video meeting held via Zoom Video Communications, Inc. (San Jose, CA) to complete the study. The study lasted approximately 45 to 50 minutes. Participants verbally consented to the study prior to participating in the study. Additionally, the researcher provided participants with her e-mail address if any questions arose.

Participants reported personal demographic information, language background, and teaching experience. Participants then viewed a student presentation that corresponded to their randomly assigned condition (Southern accent or standard accent). The researcher assigned participants to a unique respondent number upon recruitment. This number was randomly assigned to one of the two conditions. Immediately after viewing the presentation, participants completed the student presentation feedback form, where they evaluated the presentation on four dimensions and assigned a pass/fail grade. Participants then read and evaluated a student essay on an unrelated topic, which was explicitly attributed to the previous student by verbal instruction. Participants evaluated the student essay using the student essay rubric, which

evaluated the essay on five dimensions. The researcher encouraged participants to grade the student essay as they would in the natural classroom environment, writing comments and providing justification when appropriate. After completing the rubric, participants scored the essay with a letter grade on a 4.0 scale. Finally, the researcher verbally debriefed participants, revealing that they had been randomly assigned to either the Southern or standard accent condition. Participants received a $10 Visa gift card for their participation in this study.

Data Analysis

I conducted quantitative analyses using Fisher's exact test to determine whether the grades (both by dimension and by overall grade) between conditions for both the student presentation and essay were significantly different. I also utilized thematic analysis when reflecting upon any comments made by the graders for both assignments. I expected significantly higher grades in the standard accent condition as compared to those in the Southern accent condition.

Chapter Three: Results

This sample consisted of 15 graduate students enrolled at a private university in the Southeast. Participants were randomly assigned to one of two conditions, the Southern condition ($n = 8$) or the standard condition ($n = 7$). No participants were excluded after recruitment; however, the analysis for the student essay is limited to 14 participants, where the Southern condition has one less observation ($n = 7$) than in the analysis for the student presentation. This difference in *n* is due to Participant C not grading the student essay and is noted when relevant. Before analyses were conducted, the alpha level for statistical significance was set at $\propto = 0.05$.

Student Presentation

Grader Ratings

I expected participants in the Southern condition to assign significantly lower presentation grades than those in the standard condition. To evaluate the student presentation, participants from each condition rated the assignment on a 5-point Likert scale for each of the following dimensions: content, clarity, delivery/timing, and audio/visual. I treated this data as ordinal (as opposed to interval), in line with more conservative interpretations of Likert scale data (Gaito, 1980). Additionally, participants assigned a pass/fail grade for their overall impressions; all participants assigned *pass* as their overall impression, so no further analyses were necessary for this dimension.

Due to a small sample size ($N = 15$), I ran a Fisher's Exact test for each dimension. The Fisher's Exact tests revealed that there were no statistically significant differences in the ratings between conditions for any of the five dimensions (see Table 4). These results provided evidence

against my hypothesis that participants in the Southern condition would assign lower presentation grades than those in the standard condition.

The lack of statistically different presentation grades may have resulted in an ineffective manipulation for the second part of my methodology. In other words, if the grades for assignment one are not statistically different, there is an even smaller likelihood of statistically different grades for assignment two. I will further discuss this idea and its implications in the discussion section of this thesis.

Despite the lack of statistically significant findings, the Southern condition had a consistently wider spread in grades than the standard condition. For each dimension, the standard presenter tended to receive more moderate ratings while the Southern presenter was rated either very well or fair. In other words, participants in the standard condition held more consensus in their grades than did the participants in the Southern condition.

Table 4

Student Presentation Ratings of Participants between Conditions

Feedback Dimension	*Not at all*		*2*		*3*		*4*		*Very much so*		*Fisher's Exact*
	n	%	*n*	%	*n*	%	*n*	%	*n*	%	*p*
Content											0.49
Southern	-	-	-	-	1	14.3	1	14.3	5	71.4	
standard	-	-	-	-	1	14.3	3	60.0	3	60.0	
Clarity											0.18
Southern	-	-	-	-	1	14.3	1	14.3	5	71.4	
standard	-	-	-	-	1	14.3	4	57.1	2	28.6	
Delivery/Timing											0.18
Southern	-	-	-	-	1	14.3	1	14.3	5	71.4	
standard	-	-	-	-	1	14.3	4	57.1	2	28.6	
Audio/Visual											0.73
Southern	-	-	2	28.6	3	60.0	1	14.3	1	14.3	
standard	-	-	2	28.6	3	60.0	-	-	2	28.6	

Note. N = 15, (Southern = 8; standard = 7). % reflects the percentage of participants within that condition.

Grader Comments

I conducted a thematic analysis of participant presentation comments to illuminate the findings of the quantitative presentation data. The option to leave "additional comments" about the student presentation was listed at the end of the student presentation feedback form (see Appendix B). Although participants were instructed to grade these assignments as they would in a course they teach, some comments were written in third-person. For example, Participant Q commented, "…however, *the student* could have done a better job when introducing the author." Additionally, Participant F stated, "I was glad to see *the student* mentioned both positive and problematic aspects associated with her subject." Overall, three participants made explicit reference to "the student" while four participants spoke directly to the student, using second-person "you." The remaining participants used passive voice in their comments, avoiding any direct reference to the student. This variation suggests that some participants viewed their grading as an abstract task rather than comparable to the grading that they perform for a class.

I conducted a thematic analysis on the comments using the dimensions in the student presentation feedback form: content, clarity, delivery/timing, and audio/visual materials. I coded each comment with the relevant dimension(s) and whether the comment was positive or negative (see Table 5).

Table 5

Examples of Student Presentation Comments by Dimension

Content	**Audio/Visual Material**
"I was glad to see the student mentioned both positive and problematic aspects associated with her subject. This allowed the audience to have a more accurate impression of the subject, which will be helpful when her work is discussed in class."	"However, the visual aspects of the presentation did not seem to always align with what the speaker was saying, which made for an awkward flow that was a bit difficult to follow." (**Negative**; *Participant O, standard condition*)

(**Positive**; *Participant F, Southern condition*)

Delivery/Timing
"She needs to improve her voice inflection…"
(**Negative**; *Participant D, standard condition)*

Clarity
"The presentation has a clear structure and very easy to follow. Good job!"
(**Positive**; *Participant M, Southern condition*)

Overall Impressions
"Overall great layout."
(**Positive**; *Participant A, standard condition*)

Overall, participants made 31 comments on the student presentation feedback form, where 61% were negative (n = 19) and 39% were positive (n =12). The audio/visual dimension received the most negative comments (n = 10, 53% of negative comments) while the content dimension received the most positive comments (n = 6, 50% of positive comments). Negative comments tended to be more specific in nature, pointing out more precise critiques, while positive comments tended to be more general. For example, Participant H commented "Great content, but the presentation could have benefitted from the use of images to supplement the bullet points." Here, the participant supplied a general compliment about the content but a specific critique about the audio/visual materials. While this pattern is not true for every comment, it is a regular pattern in the data.

Table 6 lists how many negative and positive comments that each participant made along with their overall average combined presentation rating. Using a Pearson's correlation test, I measured the relationship between average combined ratings and number of comments made. I found a small negative relationship that was not statistically significant (r = -0.21, p = 0.45). I ran another Pearson's correlation test measuring the relationship between participants' average combined presentation ratings and number of negative comments made (r = -0.46, p = 0.08) as well as number of positive comments made (r = 0.18, p = 0.51). Here, the difference in r values

between the overall number of comments made (r = -0.21) and the number of negative comments made (r = -0.46) suggests that negative comments may exhibit a stronger effect in participant grading habits than simply how many comments were made. Interestingly, this effect suggests that positive comments also mediate the average combined rating administered in that considering positive comments with negative comments weakens the negative relationship between number of comments and higher ratings. Ultimately, the number of negative comments made had the strongest relationship with participants' average combined presentation ratings, suggesting that negative comments may be the most useful for further exploring the ratings.

Table 5

Number of Student Presentation Comments per Participant

Participant	*Total Negative Comments*		*Total Positive Comments*		*Average Rating of Dimensions*
	n	%	*n*	%	
Southern					
C	0	0.0	0	0.0	4.75
F	1	5.3	1	8.3	4.75
H	1	5.3	1	8.3	4.5
J	4	21.1	1	8.3	3.5
L	1	5.3	4	33.3	4.5
M	0	0.0	1	8.3	5
P	2	10.5	0	0.0	4.25
R	1	5.3	0	0.0	3
standard					
A	3	15.8	1	8.3	3.75
B	1	5.3	0	0.0	3.5
D	2	10.5	1	8.3	3.5
G	0	0.0	0	0.0	5
I	0	0.0	0	0.0	3.75
O	1	5.3	1	8.3	3.75
Q	2	10.5	1	8.3	4.5

Note. N = 15. % reflects the percentage of comments within that column (either positive or negative).

For each dimension, I ran a Pearson's correlation test between the number of negative comments and the grade assigned to that dimension.

Content. Overall, 5 negative comments were made about content (26% of negative comments). There was a statistically significant negative correlation between the number of negative comments that a participant made about presentation content and the rating they assigned to content (r = -0.66, p = 0.008). In other words, the more negative comments that a participant made about the content of the presentation, the lower the grade they assigned to content.

Clarity. There was only one negative comment made about clarity (5% of negative comments). Participant J explained, "It was confusing to see the themes section at the bottom of your third slide when those themes did not apply to her earlier works (or you at least did not connect her first work to any of the listed themes)." Accordingly, Participant J assigned a *3* from for *clarity*. Interestingly, despite the lack of comments, ratings of clarity still ranged from *3* to *5*, not including Participant J's rating. This inconsistency demonstrates that the variation in the clarity ratings administered cannot be fully explained by grader comments.

Delivery/Timing. Of note, two participants (11% of negative comments) in the standard condition commented on delivery while there were no comments made about delivery in the Southern condition. Participant A commented, "Pacing yourself when speaking can also help in better enunciation." Participant D commented, "She needs to improve her voice inflection." I expected that delivery comments such as these would be directed toward the Southern presenter, but she received no such commentary. Although the two participants made comments about delivery, participant A assigned a *5* (the highest rating), and participant D assigned a *4* (still higher than *3*, which was the lowest grade assigned for delivery/timing). Despite the lack of

comments for this dimension, the assigned grades varied from *3* to *5*. This variation mirrors that of *clarity* but without the same number of comments. It is possible that these dimensions may measure similar aspects of the presentation. Another explanation is that delivery/timing seems more intuitive to the grader, so they believe issues on this dimension require less commentary.

Audio/Visual. This dimension received the most negative comments (n = 10; 53% of negative comments). There was a statistically significant negative correlation between the number of negative comments that a participant made about audio/visual materials and the rating they assigned to audio/visual (r = -0.60, p = 0.02). In other words, the more negative comments that a participant made about the audio/visual materials, the lower the grade they assigned to audio/visual was.

Positive Comments. The following dimensions received positive feedback from at least one grader: content, clarity, and overall. Audio/visual, which received the most negative feedback (n = 10, 32% of total presentation comments), received no positive comments. Content received the most positive feedback (n = 6, 19% of total presentation comments). Of note, comments about content were more specific than other positive dimensions. For example, Participant L comments, “You put together an effective introduction to Simone de Beauvoir. You addressed all key aspects of the prompt (education, career, and major literary works) and cited your sources appropriately.” Here, Participant L gives specific feedback regarding why the student succeeded on the dimension of content. This comment highly contrasts with those in the overall dimension (n = 4, 13% of total presentation comments) such as “Overall great work” and “Nice job!”

Student Essay

Grader Ratings

I expected participants in the Southern condition to assign significantly lower essay grades than those in the standard condition. To evaluate the student essay, participants from each condition rated the assignment on a 4-point Likert scale for each of the following dimensions: evidence, analysis, formality, organization, and mechanics. Additionally, participants assigned a letter grade based on the 4.0 GPA scale (with *A* as the highest grade and *F* as the lowest). Again, I treated this data as ordinal (see above explanation). One participant failed to grade the student essay, so their data was excluded from this analysis. Moreover, one participant did not assign a final grade (although they did grade the essay using the rubric). As such, 14 participants were included in this analysis, 7 from the Southern condition and 7 from the standard condition.

Due to a small sample size ($N = 14$), I ran a Fisher's Exact test for each dimension. The Fisher's Exact tests revealed that there were no statistically significant differences in the ratings between conditions (see Table 7). This result provided evidence against my hypothesis that participants in the Southern condition would assign lower essay grades than those in the standard condition.

Despite the lack of statistically significant findings, the Southern condition again had a consistently wider spread in grades than the standard condition. For each dimension, the standard condition graders tended to assign more "very good" ratings while the Southern condition graders assigned either "very good" or "excellent" grades. In other words, participants in the standard condition again held more consensus in their grades than did the participants in the Southern condition.

Table 7

Student Essay Ratings of Participants between Conditions

Prompt Dimension	*Needs Improvement*		*Fair*		*Very Good*		*Excellent*		*Fisher's Exact Test*
	n	%	*n*	%	*n*	%	*n*	%	*p*
Evidence									0.76
Southern	-	-	1	14.3	3	42.9	3	42.9	
standard	-	-	1	14.3	5	71.4	1	14.3	
Analysis									1.00
Southern	-	-	2	28.6	3	42.9	2	28.6	
standard	-	-	2	28.6	4	71.4	1	14.3	
Formality									1.00
Southern	-	-	1	14.3	2	28.6	4	71.4	
standard	-	-	-	-	3	42.9	4	71.4	
Organization									0.14
Southern	-	-	4	71.4	1	14.3	2	28.6	
standard	-	-	3	42.9	4	71.4	-	-	
Mechanics									1.00
Southern	-	-	1	14.3	3	42.9	3	42.9	
standard	-	-	1	14.3	4	71.4	2	28.6	

Note. $N = 14$ ($n = 7$ for each condition). % reflects the percentage of participants within that condition.

Grader Comments

I conducted a thematic analysis of participant essay comments to illuminate the findings of the quantitative essay data. I instructed participants to grade the student essay as they would for a class that they had instructed, and I encouraged them to leave comments throughout the essay. I conducted a thematic analysis on these comments using the dimensions in the student essay rubric (see Appendix C): evidence, analysis, formal tone and style, organization/transitions, and mechanics. I coded each comment with the relevant dimension(s) and whether the comment was positive or negative (see Table 8).

Table 8

Examples of Student Essay Comments by Dimension

Formal Tone & Style "Avoid using extreme adjectives like 'perfect.'" (**Negative**; *Participant M, Southern condition*)	**Overall Impressions** "Overall, the prompt is addressed relatively well." (**Positive**; *Participant R, Southern condition*)
Evidence "The fact that you have not included the passage you are analyzing and have not provided enough contextualization information to a general audience will have an impact on your final grade." (**Negative**; *Participant Q, standard condition*)	**Analysis** "Additionally, towards the end it wasn't quite clear how she used feminine identity to create the space mentioned in your thesis." (**Negative**; *Participant H, Southern condition*)
Mechanics (Spelling & Grammar) "This sentence is long and hard to follow as a reader." (**Negative**; *Participant F, Southern condition*)	**Organization & Transitions** "This paragraph was well-written and it was easy to follow your logic. Consider using this as an example for previous paragraphs." (**Positive**; *Participant O, standard condition*)

Overall, participants made 118 comments on the student essay, where 78% were negative (n = 92) and 22% were positive (n = 26). The evidence dimension received the most negative comments (n = 31, 34% of negative comments) while the overall dimension received the most positive comments (n = 14, 54% of positive comments). Again, negative comments tended to be more specific in nature, pointing out more precise critiques, while positive comments tended to be more general. For example, Participant O commented, "This paper touched on some interesting ideas but was difficult to follow at times. Consider revising the first several paragraphs to set a stronger foundation for your arguments that the reader can easily follow." Here, the participant supplied a general compliment about the content but a specific critique

about the organization—supporting their critique with ways to correct the error in the future while leaving the compliment without further explanation.

Overall, the frequency of the comments was similar between groups, where participants in both conditions made an average of ~8.0 comments. Table 9 lists how many negative and positive comments that each participant made along with the overall grade they administered. Using a Pearson's correlation test, I measured the relationship between the GPA value of the grade (on a 4.0 scale) and the number of comments made. I found a moderate negative relationship that was statistically significant ($r = -0.67$, $p = 0.01$). I ran another Pearson's correlation test measuring the relationship between participants' grade administered and number of negative comments made ($r = -0.87$, $p < 0.001$; see Figure 1) as well as number of positive comments made ($r = 0.15$, $p = 0.63$). Here, the difference in r values between the overall number of comments made ($r = -0.67$) and the number of negative comments made ($r = -0.87$) again suggests that negative comments may exhibit a stronger effect in participant grading habits than simply how many comments were made. Moreover, this effect suggests that positive comments also mediate the grade administered in that considering positive comments with negative comments weakens the negative relationship between number of comments and higher ratings. Accordingly, the number of negative comments made had the strongest relationship with the grade participants administered, suggesting that negative comments may be the most useful for further exploring the ratings by dimension.

Table 9

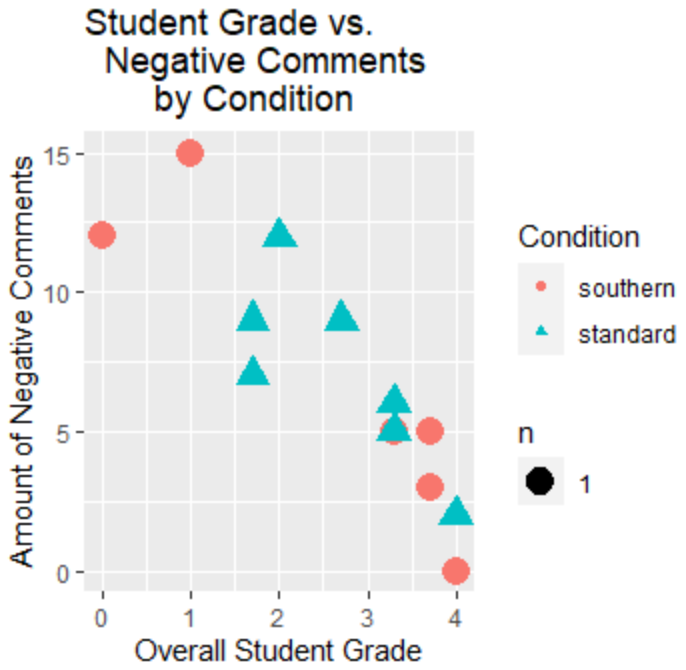
Number of Student Presentation Comments per Participant

Participant	*Total Negative Comments*		*Total Positive Comments*		*Letter Grade Assigned*
	n	%	*n*	%	
Southern					
F	12	13.0	0	0.0	F
H	2	2.2	4	15.4	*NA*
J	15	16.3	2	7.7	D
L	5	5.4	9	34.6	B+
M	5	5.4	1	3.8	A-
P	0	0.0	0	0.0	A
R	3	3.3	1	3.8	A-
standard					
A	9	9.8	0	0.0	C-
B	9	9.8	1	3.8	B-
D	2	2.2	1	3.8	A
G	7	7.6	0	0.0	C-
I	5	5.4	2	7.7	B+
O	12	13.0	4	15.4	C
Q	6	6.5	1	3.8	B+

Note. $N = 14$ ($n = 7$ for each condition). % reflects the percentage of comments within that column (either positive or negative).

Figure 1

Correlation between Negative Comments and Essay Grades



Note. See Table 9 for percentages of overall negative comment share.

For each dimension, I ran a Pearson's correlation test between the number of negative comments and the grades assigned to that dimension.

Evidence. Evidence received a 34% share of negative comments (n = 31). There was a statistically significant negative correlation between the number of negative comments that a participant made about evidence in the student essay and the rating they assigned to evidence (r = -0.56, p = 0.04). In other words, the more negative comments that a participant made about evidence, the lower the grade they assigned to evidence was.

Regarding evidence, graders in the Southern condition gave an average of approximately 2 negative comments (SD = 2.5) compared to an average of approximately 3 negative comments (SD = 1.8) made by participants in the standard condition. Only 3 (9% of total evidence

comments) positive comments were made by Southern conditions graders while standard condition graders made no positive comments.

Analysis. Analysis received a 24% share of negative comments (n = 22). There was a statistically significant negative correlation between the number of negative comments that a participant made about analysis in the student essay and the rating they assigned to analysis (r = -0.67, p = 0.009). In other words, the more negative comments that a participant made about analysis, the lower the grade they assigned to analysis was.

Regarding analysis, graders in the Southern condition gave an average of approximately 2 negative comments (SD = 1.8) compared to an average of approximately 1 negative comment (SD = 1.6) made by participants in the standard condition. Only 5 positive comments were made, with 4 (15% of total analysis comments) being made by Southern condition graders and 1 (4% of total analysis comments) being made by a standard condition grader.

Formal Tone & Style. Only two comments (2% of total comments) about tone and style were made, both of which were negative. One was made by a Southern condition grader, and one was made by a standard condition grader, yet they assigned a *4* and a *3* respectively (the lowest score being a *2*). Formality did receive the highest frequency of *4* ratings, but the scores still demonstrated a lack of consensus, ranging from *2* to *4*. Therefore, the comments (or lack thereof) do not shed much light on the range of grades.

Organization & Transitions. There was a small but statistically insignificant inverse relationship between negative comments about organization and transitions and the grades assigned to this dimension (r = -0.25, p = 0.38). Interestingly, organization and transitions received more low grades than any other dimensions. Moreover, negative comments about organization comprise 28% of negative comments (n = 26), coming in second behind *evidence*

for share of negative comments. Despite participants obvious attention to organization and transitions, the comments do not linearly align with the grading practice.

Regarding organization and transitions, graders in the Southern condition gave an average of approximately 6 negative comments (*SD* = 5.5) and 2 positive comments (*SD* = 3.2) compared to an average of approximately 7 negative comments (*SD* = 3.2) and 1 positive (*SD* = 1.4) comment made by participants in the standard condition.

Mechanics. Only 10% of negative comments pertained to mechanics (*n* = 9). Moreover, there was a small but statistically insignificant inverse relationship between negative comments about mechanics and the grades assigned to this dimension ($r = -0.28$, $p = 0.29$). Even still, there lacks a strong consensus among the graders about mechanics in the essay.

Regarding mechanics, 8 (89% of total mechanics comments) negative comments were made by graders in the standard condition while only 1 (11% of total mechanics comments) was made by a grader in the Southern condition. There were no positive comments regarding mechanics.

Positive Comments. The following dimensions received positive feedback from at least one grader: evidence, analysis, organization and transitions, and overall. Evidence, which received the most negative feedback (*n* = 31, 29% of total comments), was given 3 positive comments. All three of these comments were given at the end of the paper in a reflective feedback paragraph (rather than commented upon within the essay). As an example, Participant J wrote, “You give good examples.” This comment is another case of positive feedback being presented in general terms. The generality is not just demonstrated through the words themselves but also through the placement of the comment: it appeared at the end of the paper in a more general reflection of the student’s performance rather than a specific compliment about a given

example. *Overall* received the most positive feedback (*n* = 14, 12% of total essay comments). This category was created for comments that referred to more general compliments, rather than ones that could be associated with any particular dimension. For example, Participant R commented, "the prompt is addressed relatively well." What follows is a more specific critique, "...but there are points within the evidence given for the argument that are themselves ambiguous." While the Participant R does not elaborate on this ambiguity, he does give a specific area (evidence) upon which the student could improve—further emphasizing the difference in nature between positive and negative comments.

Language Experience

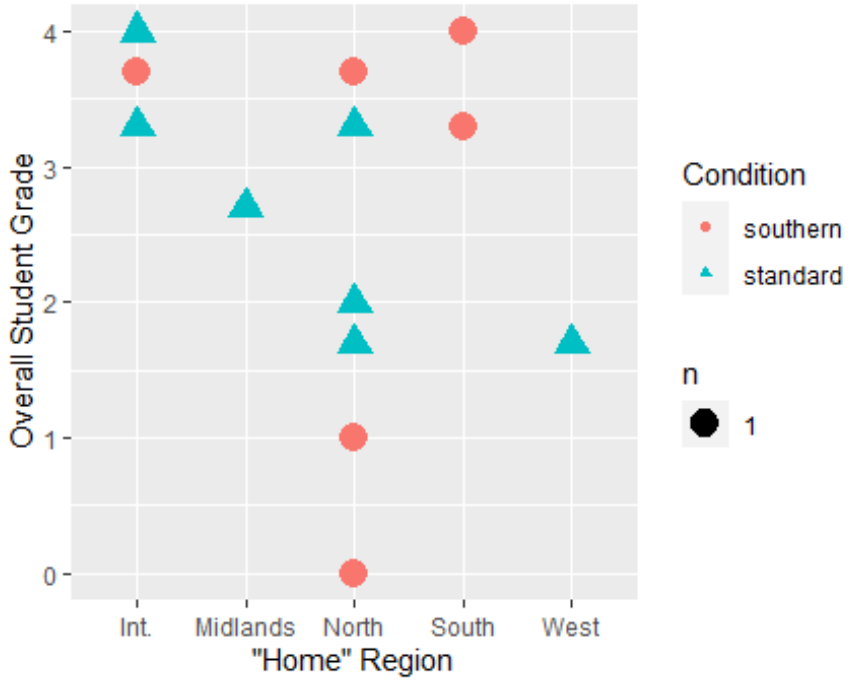
Although there was not a statistically significant difference for overall essay grades between the Southern and standard conditions, I was interested in measuring how participant language experience mapped onto this relationship. Figure 2 demonstrates the overall essay grades between conditions when considering language experience, or the dialect region that contains where the participant calls "home." Interestingly, the two lowest grades were assigned in the Southern condition by graders whose "home" is located in the Northern dialect region. In perceptual dialectology studies, Preston (1996) found that Michiganders, who are a part of the Northern regional dialect, tend to find Southern speech as "incorrect" and "unpleasant." In accordance with this finding, the two graders who call the North "home" may have exhibited this language bias in their grading. In turn, both Southerners in the sample graded the Southern student and assigned relatively higher grades as compared to the average standard condition grade. According to Preston's (1996) research, Southerners exhibit solidarity with those who speak similarly to them. Using perceptual dialectology findings to interpret this data, it is

possible that the graders who call the South “home” are exhibiting a bias in favor of the Southern student. Moreover, the graders who identify as international students graded very similarly between conditions, which could demonstrate a lack of language bias when compared to domestic students.

When considering this explanation, this pattern illuminates the wider variation in Southern grades as compared to standard grades. These data suggest that the larger variation is due to varying language attitudes regarding Southern accents. A larger, more representative sample size would be needed to further measure the accuracy of this claim.

Figure 2

Essay Grades between Conditions by Region



Academic Department

I examined whether participant department could further explain the relationship between essay grades and participant condition. Because of the limited sample size, I was unable to identify any suggestive patterns in the data as demonstrated in Figure 3.

Figure 3

Essay Grades between Conditions by Region and Department

Individual Grading Differences

Figure 4 demonstrates the relationship between a participant's combined average ratings of the student presentation and the overall grade that they assigned to the student essay. I ran a Pearson's correlation test and found a small but statistically insignificant relationship between the two scores (r = -0.16, p = 0.60). The data suggests that there is no correlation between the

student presentation grades and the student essay grades. As a result, any differences in grades between conditions or other factors are likely not a result of a participant who assigned higher or lower grades on average. Of note, the presentation grades were higher (*M* = 4.13/5.00), on average, than the essay grades (*M* = 2.65/4.00), so this difference in means may attribute to the lack of relationship between scores.

Figure 4

Individual Grading Differences between Assignments

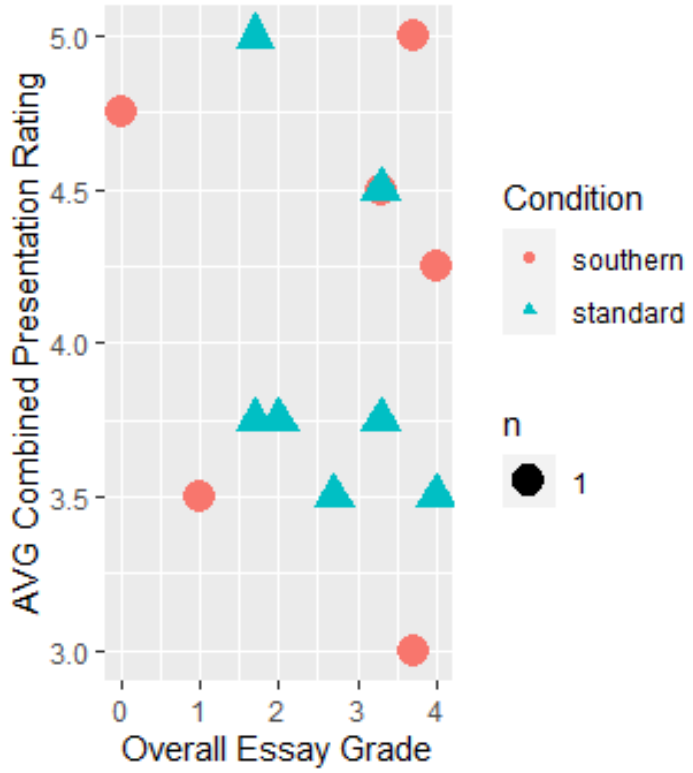


Table 10 lists the number of comments made by each participant between both assignments. Using a Pearson's correlation test, I measured whether the participants who commented on the student presentation were more likely to comment on the student essay. There was a moderate relationship between the number of presentation comments and the number of

essay comments ($r = 0.46$, $p = 0.099$). While this relationship is statistically unreliable at the predetermined alpha level, it does suggest the possibility that graders who highly comment on one assignment are more likely to highly comment on another assignment.

Table 10

Number of Comments per Participant between Assignments

Participant	*Total Presentation Comments*		*Total Essay Comments*	
	n	%	*n*	%
Southern				
C	0	0.0	*NA*	*NA*
F	2	6.5	12	10.0
H	2	6.5	6	5.1
J	5	16.1	17	14.4
L	5	16.1	14	11.9
M	1	3.2	6	5.1
P	2	6.5	0	0.0
R	1	3.2	4	3.4
standard				
A	4	12.9	9	7.6
B	1	3.2	10	8.5
D	3	9.7	3	2.5
G	0	0.0	7	5.9
I	0	0.0	7	5.9
O	2	6.5	16	13.6
Q	3	9.7	7	5.9

Note. % reflects the percentage of comments within that column (either presentation or essay).

Chapter Four: Discussion

This research sought to answer the question of whether teachers assign lower grades to Southern English speakers than to their "standard"-sounding peers. To answer this question, I conducted a study in which participants graded an identically-scripted student presentation—either given in a Southern accent or a "standard" accent—and then an identical student essay. I hypothesized that participants in the Southern condition would assign lower grades to both assignments than those in the standard condition due to language interference based on a standard language ideology.

Variation in Southern Grades

When measuring the average difference between ratings, the Southerner and "standard"-sounding student fared similarly. Yet, an average score between groups does not tell the whole story; evaluating the spread of scores reveals the variation in which Southerners are graded. When evaluating the variance between groups, a pattern arises that reveals the surety in which folks graded the "standard"-sounding student and the variation in which folks graded the Southern student. In other words, the variance reveals the lack of consensus between graders when grading the Southern student. As compared to the Southern student, most graders were able to agree on what grade the "standard"-sounding student deserved. This same deviation existed for the amount of comments that participants gave regarding the student essay.

Though participants gave approximately the same amount of comments by dimension, the graders in the Southern condition were more likely to vary upon how many times they commented. This pattern could potentially be attributed to individual grading style except for the fact that this variation did not occur in the standard condition. Moreover, there were a similar

amounts of comments between conditions on average, but these comments were more equally distributed between participants in the standard condition as compared to the Southern condition. In the context of sociolinguistic and language attitude research, this pattern is not surprising. In fact, this research provides support to the interpretation that Southern student grades vary more because of varying language attitudes.

Statistical analyses of the grades did not support my hypothesis as the scores between the conditions were not reliably different from one another. Even so, the distribution of Southern condition grades was wider than that of the standard condition for both the presentation and the essay. This variation demonstrates a lack of equality in the education system. While some educators may tolerate, or even appreciate, a student with a "non-standard" accent, this treatment is not guaranteed. The inconsistency of grades for the Southern student is best understood in light of standard language ideology. The graders who assigned higher grades to the Southern presenter identified as international students or Southerners. It makes sense that international students would not be privy to the language prejudices held against Southerners as they most likely were not indoctrinated with the standard language ideology of American English (in the same way as domestic students). What's more, Southerners graded the Southern student very favorably, perhaps demonstrating an act of linguistic solidarity with a fellow Southern academic.

In the case that the wider variation in Southern grades is due to normal variance, my results would suggest that instructor grades do not reflect any bias regarding Southern speech. Because I did not directly measure language attitudes, I cannot conclude whether lack of biased grading conflates to a lack of bias against Southern speakers. In fact, in accord with sociolinguistic and language attitude research (Lippi-Green, 2012; Preston, 1989; 1996), I expect that most (if not all) my participants hold at least some negative biases against Southern speech.

Even so, this bias was not captured in my study. While the most optimistic view may conclude that Southerners escape unscathed in teacher evaluations, this conclusion does not dismantle the powerful stronghold that standard language ideology maintains within the education system and beyond. Therefore, the following discussion will examine three possible explanations of why linguistic bias was not detected in the grading: 1) The manipulation of the student presentation, 2) The experience level of the graders, and 3) The confidence to grades across disciplines.

Student Presentation

The student presentation grades were relatively high across both conditions, with every participant assigning "satisfactory" (rather than "unsatisfactory") as their overall impression. Furthermore, the scores assigned to the different dimensions (content, clarity, delivery/timing, and audio/visual materials) were not reliably different between conditions. In other words, there was no measured bias elicited against the Southern speaker in this task. Importantly, the "standard" speaker used the pin-pen merger 100% of times possible in the recording, so the risk that participants registered the "standard" speaker as a Southern speaker is plausible. If this attribution did take place, then it would make sense why there was no difference in scores between conditions.

Despite the integrity of my matched guise manipulation, these results corroborate the findings of Fletcher (2020) in which Southern-speaking students were not graded reliably lower than their "standard"-speaking peers when providing "oral answers to academic questions." Yet, both Fletcher (2020) and the current study seemingly contradict the research on stigmatized versus "standard" varieties (Giles, 1992; Purnell, Idsardi, & Baugh, 1999) as well as perceptual dialectology studies (Preston, 1989; 1996). I argue, however, that this contradiction

illuminates the subjectivity in grading. In other words, grading bias is not limited to a singular bias, such as linguistic bias, but rather a complex combination of biases that informs grading practices. These biases can pertain to a variety of student characteristics, such as previous performance or physical appearance, as well as reflect the educator's personal background. Because such a range of biases can arise in the grading process, it becomes harder to detect a singular bias, such as a linguistic bias. While I was unable to detect a difference in ratings between conditions, the ratings administered still varied—demonstrating individual differences in grading. As such, some other biases must be at play. To measure whether instructors include negative language attitudes about Southern speech as one of these biases, future studies should increase their sample size to ultimately increase effect sizes that were potentially too small for statistical significance in this study.

Beyond measuring these effects with the assigned grades, I also analyzed the comments made by the graders. Interestingly, only two comments regarding speech manner were made by graders, and both graders were in the standard condition. One urged the presenter to work on "voice inflection" while the other instructed the speaker to "pace [herself] when speaking" in order to achieve "better enunciation." To my surprise, graders in the Southern condition did not provide any feedback about speech.

Grader Experience

The age and experience level of the sample may be more resistant to bias. According to Metz (2019), English teachers with more experience are more likely to hold "hegemonic views about the dominant school narrative." In other words, the longer that English teachers have been in the education system, the more that their views regarding educational institutions reflect

standard language ideology. In the current study, 73% of participants had some teaching experience prior to their participation, but the mean age of participants was approximately 27 (*SD* = 4.57). Therefore, using age as an estimate for teaching experience, the average participant tended to have limited experience. As a result, the possibility that this sample represents newer instructors who are more resistant to standard language ideology is plausible. Accordingly, this effect could explain the surprising lack of bias in the Southern condition. To further substantiate this claim, similar research should be conducted with more experienced graders; that is to say, the research would need to demonstrate my hypothesized effect within the more-experienced sample. If this effect were found, then the finding would further support the pervasive nature of standard language ideology in that its prevalence grows as one's exposure increases.

Grading across Disciplines

Participants' comfort with grading across disciplines could have impacted their attention to student accent. During the study, participants graded two student assignments: a presentation and an essay. The presentation was about the life and work of Simone de Beauvoir for an introductory Philosophy course, and the essay was a close reading of Sor Juana Ines de la Cruz's Response to Sor Filotea for an introductory English course. Both assignments were written for classes in the humanities. Approximately 29% (*n* = 4) of participants reported disciplines in the humanities while approximately 71% (*n* = 10) reported disciplines in either the social or natural sciences. During their participation, some participants expressed difficulty in grading the assignments because they were not familiar with the genre or discipline. If participants were not comfortable grading student work outside their discipline, then perhaps Southern bias would have been harder to detect. In future studies, stratifying the data by participant discipline could

serve to mitigate confusion and discomfort that detracts from the grading process. Moreover, a stratification method would lend itself to greater external validity as the participants would be grading work in their own disciplines.

Practical Implications

Grading can capture more than academic achievement. In this study, I demonstrated that a Southern student was graded with less consensus than her "standard"-sounding counterpart. This variation has major implications for student success. While average test scores aid in computing statistical significance, individual students have the same instructor for the duration of the semester. In reality, if a Southern student were to have an instructor who assigned lower grades all semester, those grades would not average out with the "standard"-sounding counterpart. In other words, most classes do not have multiple graders that ensure an average rating and consensus among more than one grader. Therefore, this variation is much more troubling in practical terms.

If educators are interested in identifying whether standard language ideology informs their grading practices, my research suggests that teacher language background is a factor in how SLI affects grading. In other words, if an educator speaks or grew up around a "non-standard" dialect, then that educator may exhibit more solidarity with "non-standard" speaking students in their grading practices. Even so, the education system is a key language authority in spreading standard language ideology (Lippi-Green, 2012). Additionally, while English teachers may hold linguistically informed views about their own role in English education, they tend to maintain SLI-informed views about the purpose of the education system (Metz, 2019). Moreover, many teachers do not align their grading practices with their linguistically informed views (Wheeler,

2019). Therefore, evaluating one's own biases in grading is only the first step in a longer process in ensuring the equality of grading practices.

References

Alim, H. S. (2005). Critical language awareness in the United States: Revisiting issues and revising pedagogies in a resegregated society. *Educational Researcher*, *34*(7), 24–31. https://doi.org/10.3102/0013189X034007024

Charity Hudley, A., and Mallinson, C. (2011). *Understanding English language variation in U.S. schools*. Teachers College Press.

Cramer, J., & Montgomery, C. (Eds.). (2016). *Cityscapes and perceptual dialectology: Global perspectives on non-linguists' knowledge of the dialect landscape*. De Gruyter Mouton.

Crowl, T. K., & MacGinitie, W. H. (1974). The influence of students' speech characteristics on teachers' evaluations of oral answers. *Journal of Educational Psychology*, *66*(3), 304–308. https://doi.org/10.1037/h0036498

Dunstan, S. B., & Jaeger, A. J. (2015). Dialect and influences on the academic experiences of college students. *The Journal of Higher Education*, *86*(5), 777--803.

Fletcher, M. K. (2020). Bias in the classroom: How that Southern twang could influence instructor evaluations. https://doi.org/10.17615/e5y5-7m73

Gaito, J. (1980). Measurement scales and statistics: Resurgence of an old misconception. *Psychological Bulletin*, *87*(3), 564--567.

Giles, H., Henwood, K., Coupland, N., Harriman, J., & Coupland, J. (1992). Language attitudes and cognitive mediation. *Human Communication Research*, *18*, 500--527.

Godley, A. J., & Minnici, A. (2008). Critical language pedagogy in an urban high school English class. *Urban Education*, *43*(3), 319–346. https://doi.org/10.1177/0042085907311801

Godley, A. J., Sweetland, J., Wheeler, R. S., Minnici, A., & Carpenter, B. D. (2006). Preparing teachers for dialectally diverse classrooms. *Educational Researcher*, *35*(8), 30–37. https://doi.org/10.3102/0013189X035008030

Klapp Lekholm, A., & Cliffordson, C. (2008). Discrepancies between school grades and test scores at individual and school level: Effects of gender and family background. *Educational Research and Evaluation*, *14*(2), 181--199.

Labov, W. (1995). Can reading failure be reversed? A linguistic approach to the question. *Literacy among African-American Youth: Issues in Learning, Teaching, and Schooling*, *39*, 68.

Labov, W., Ash, S., & Boberg, C. (2005). *The atlas of north American English: Phonetics, phonology and sound change*. De Gruyter Mouton.

Lambert, W. (1967). The social psychology of bilingualism. *Journal of Social Issues*, *23*, 91--109.

Lippi-Green, R. (2012). *English with an accent: Language, ideology, and discrimination in the United States.* Routledge.

MacCann, C., & Roberts, R. D. (2013). Just as smart but not as successful: Obese students obtain lower school grades but equivalent test scores to nonobese students. *International Journal of Obesity*, *37*(1), 40--46.

MacSwan, J. (2020). Academic English as standard language ideology: A renewed research agenda for asset-based language education. *Language Teaching Research*, *24*(1), 28–36. https://doi.org/10.1177/1362168818777540

Malouff, J. M., Emmerton, A. J., & Schutte, N. S. (2013). The risk of a halo bias as a reason to keep students anonymous during grading. *Teaching of Psychology*, *40*(3), 233--237.

Metz, M. (2019). Accommodating linguistic prejudice? Examining English teachers' language ideologies. *English Teaching: Practice & Critique, 18*(1), 18--35.

Milroy, J., & Milroy, L. (2012). *Authority in language: Investigating standard English*. Taylor and Francis.

National Council of Teachers of English. (1998). Mission statement. https://ncte.org/about/

Preston, D. R. (1996). Where the worst English is spoken. *Focus on the USA, 16*, 297.

Preston, D. R. (Ed.). (1999). *Handbook of perceptual dialectology* (Vol. 1). John Benjamins Publishing.

Preston, D. R. (1999). A language attitude approach to the perception of regional variety. In D. R. Preston (Ed.), *Handbook of perceptual dialectology* (pp. 359--373). John Benjamins Publishing.

Purnell, T., Idsardi, W., & Baugh, J. (1999). Perceptual and phonetic experiments on American English dialect identification. *Journal of Language and Social Psychology*, *18*, 10--30.

Rickford, R. (1999). *African American Vernacular English: Features and use, evolution, and educational implications*. Wiley-Blackwell.

Shepherd, M. A. (2011). Effects of ethnicity and gender on teachers' evaluation of students' spoken responses. *Urban Education*, *46*(5), 1011–1028. https://doi.org/10.1177/0042085911400325

Tamasi, S., & Antieau, L. (2015). *Language and linguistic diversity in the US: An introduction*. Routledge.

Wheeler, R. S. (2019). Attitude change is not enough: Disrupting deficit grading practices to disrupt dialect prejudice. *Proceedings of the Linguistic Society of America*, *4*(1), 10--1.

Appendix A: Demographic Survey

Are you currently enrolled in a graduate program at *Southern* University?
What year are you in your program?
In what department is your program?
What is your age?
What is your current gender identity?
Please indicate your race/ethnicity? (Select all that apply)

- American Indian/Alaska Native
- Asian
- Black/African American
- Hispanic/Latinx
- Native Hawaiian/Pacific Islander
- Caucasian (including European, Middle Eastern, or North African descent
- Other: (Fill in the blank)

Where are you from? i.e., Where do you consider to be "home"?
How long have you lived in Atlanta?
Are you an international student?
Is American English your primary or dominant language?
If not, which language is primary?
Where did you initially learn English?

Teaching Experience Questionnaire
Did you have any teaching/grading experience prior to entering your current program?
(This could include TA experience, K-12 teacher, or undergraduate course instructor.)

If so, please list those experiences.

Do you have any teaching/grading experience gained from your current program?
(This could include TA experience, K-12 teacher, or undergraduate course instructor.)

If so, please list those experiences.

Do you plan to teach upon completion of your program?
(This could include K-12 teaching, college professor, etc.)

Do your future teaching plans involve student writing?

How confident are you in your ability to assess student writing?
1 2 3 4 5
(Not at all) (Very Much So)

Appendix B: Student Presentation Feedback Form

Prepare a 2- to 3-minute presentation about the author of the assigned reading. You must introduce the author and present a brief summary of their education, career, and major literary works. This presentation is not meant to be comprehensive but rather a conversation starter for class discussion. This assignment is graded on an S/U basis and will be 5% of your overall grade.

Content

How well did the student answer the prompt (provided above)?

1 2 3 4 5
(Not at all) (Very much so)

Clarity

How well were you able to follow and understand the presentation?

1 2 3 4 5
(Not at all) (Very much so)

Delivery / Timing

How well did the student time and organize their presentation?

1 2 3 4 5
(Not at all) (Very much so)

Audio / Visual Materials

Did the audio/visual materials enhance the presentation?

1 2 3 4 5
(Not at all) (Very much so)

Overall Impressions Grade

☐ Pass / ☐ Fail

Additional comments:

Appendix C: Student Essay Rubric

Element	Needs Improvement 1	Fair 2	Very Good 3	Excellent 4
Evidence	No evidence provided.	Information is unclear, inaccurate, or lacks citations.	Provides evidence and includes necessary details with some citation errors.	Supports the paper with appropriate evidence, thorough details, and accurate citations.
Analysis	Contains little to no analysis of the information presented.	Attempts to analyze the information, but the analysis is unclear or inaccurate.	Analyzes most of the information presented.	Clearly, concisely, and thoroughly explains and analyzes the information presented.
Formal Tone and Style	Informal language present throughout.	Writing contains some informal elements.	Writing attempts to maintain a formal and objective tone.	Writing maintains a formal and objective tone throughout.
Organization & Transitions	Little to no attempt at organization.	Attempts to organize ideas, but transitional language is needed.	Organizes ideas in a logical way. Transitional language used.	Strong organization and transitional language used skillfully throughout.
Mechanics (Spelling & Grammar)	Distracting mechanical errors throughout.	Mechanical errors distract the reader at times.	A few errors present, but they largely do not distract the reader.	Mechanics reflect careful editing.

Essay Prompt:

ENG181: Writing about Literature

For your first formal essay, you will perform *a close reading* on a passage that we have not analyzed in class. Your close reading should be *thesis-driven* and written for a *general audience*. You should *not* consult outside sources. As you are writing, it will help you to consider the themes of our class. You should have a draft of your essay prepared for peer workshopping, which will occur a week before the official due date. Please note that this essay is worth **20% of your grade**. See the attached rubric for more information regarding how your paper will be evaluated. If you have any questions, feel free to email me or visit me during my office hours.

Appendix D: Tables

Appendix E: Figures