
David M. Johnson  
*Kennesaw State University*, djohnson@kennesaw.edu

Lewis VanBrackle  
*Kennesaw State University*, lvanbrac@kennesaw.edu

Recommended Citation  

Abstract
Raters of Georgia’s (USA) state-mandated college-level writing exam, which is intended to ensure a minimal university-level writing competency, are trained to grade holistically when assessing these exams. A guiding principle in holistic grading is to not focus exclusively on any one aspect of writing but rather to give equal weight to style, vocabulary, mechanics, content, and development. This study details how raters react to “errors” typical of African American English writers, of ESL writers, and of standard American English writers. Using a log-linear model to generate odds ratios for comparison of essays with these error types, results indicate linguistic discrimination against African American “errors” and a leniency for ESL errors in writing assessment.

1. Introduction

Several years ago, Michele Pfieffer portrayed a high school English teacher in the movie *Dangerous Minds*. The movie is set in an underachieving urban high school in the U.S. and focuses on the attempts of a novice but caring high school teacher who, through innovative techniques, is able to motivate students who had heretofore been uninterested in academics. However, more fascinating and more realistic is not the depiction of the main character but rather her more experienced and considerably more embittered writing teacher colleague. In a memorable scene, this character sits alone grading student essays. The camera focuses on him as he grades papers, and the audience sees his furrowed brow and listens as he mumbles his impressions of the student writers. “Idiot” and “stupid” are his adjectives of choice as he shakes his head and circles mistakes. Many writing teachers may identify with this character who is “irritated” by student writing errors. Raters of student writing react strongly to writing errors and those reactions range from irritation and exasperation to amusement and sympathy. We will argue that reactions to errors are a significant factor in the assessment of student writing on a standardized writing exam that is graded holistically. Most significantly, this study will demonstrate that raters react differently to “errors” typical of African-American (AAE) English
writers, of errors of English as a Second Language (ESL) writers, and of errors of Standard American English (SAE) writers and the end result is imbalanced assessment of essays that are identical except for eight different kinds of errors.

For this study, we limited our exploration of this linguistic discrimination on the part of raters to an examination of which kinds of errors raters consider more egregious; thus, we did not consider the other higher level writing features. As the data collection procedures will show, we were able to isolate raters’ reactions to errors exclusively and not consider reactions to other writing issues such as coherence, originality, or style.

We need to explain our use of quotes around *errors* when we reference these in regard to African American English. Linguists note that what mainstream writing teachers regard as errors might be reclassified as *features*. The African American dialect has many non-standard features (as does any dialect of English) which should be considered a *feature* from a descriptive linguistic standpoint. These contrast from ESL errors which result from a lack of complete language acquisition. We acknowledge this classification of features by non-linguists as errors, but at the same time our focus is not on a justification of a group’s right to their own dialect. Our focus is on how different kinds of features (or “errors”) influence writing assessors. Thus, all non-standard elements in the students essays will henceforth simply be called errors.

2. Previous Studies on Writing Errors

Previous research has examined rater reactions to errors. Connors and Lunsford (1988) conducted an historical and experimental study on rater reactions to errors. They concluded that educators in the early 20th century were more concerned about errors than contemporary educators due to the advent of process writing pedagogies. Paradoxically they found that while
contemporary educators claim to be less error-driven, they, in fact, mark errors more than they comment on other aspects (e.g. development or content) of essays. Most significantly, Connors and Lunsford (1998) contend that definitions of errors and views of what constitutes a severe error change, and that teachers are influenced by current trends. This last finding was confirmed by Lunsford and Lunsford (2008) in which the Connors and Lunsford study (1998) was replicated. Both of these research projects resulted in “top-twenty” essay error lists. Our research incorporates many of these top-twenty errors.

Santos (1988) and Marshall and Powers (1969) note that raters’ reactions to errors play a major role in the evaluative process. Santos (1998) notes that errors may irritate professors “…even when the message is comprehensible to them” (p. 70). James (1977) examines rater reaction to ESL errors and notes that while certain types of errors (verb morphology) are viewed more critically than other errors (lexical), judging the seriousness of error types remains highly subjective. He speculates that most raters do not “explicitly formulate...criteria” when considering seriousness of errors (James, 1977, p. 116).

Other studies have focused on a comparison of errors from ESL and SAE. These studies are somewhat contradictory. Vann, Meyer, and Frederick (1984) and Roberts and Cimasko (2007) concluded that raters react more negatively to errors typical of ESL writers. In contrast, Rubin and Williams-James (1997) and Janopoulos (1992) concluded that raters overcompensate for ESL writers’ errors and raters are frequently more critical of SAE writers than ESL writers because the SAE writer “should know better.” Their status as native English speakers gives them an advantage over ESL speakers, and thus, they should be able to edit their writing more carefully. The fact that errors remain indicates simple carelessness or laziness according to raters in Janopoulos’ study (1992).
Williams (1981) notes that writing errors invoke strong emotional reactions from teachers and he questions the ferocity with which many teachers look for errors. He notes that “…it is all very puzzling: (There is) great variation in our definition of error, great variation in our emotional investment in defining and condemning error, great variation in the perceived seriousness of individual errors” (Williams, 1981, p. 155).

Previous research on errors has also included how teachers respond to errors that stem from non-standard dialects. Hairston (1995) and Beason (2001) focused on errors relating to dialects which Hairston (1995) labeled as “status markers” (e.g. “When Mitchell moved, he *brung* his secretary with him.”) and which ones “bothered” (to use Hairston’s term) prospective employers. Results indicate that errors greatly influence employers’ opinions of writing competency and that raters reacted most negatively to errors related to non-standard dialects. Beason (2001) notes that raters view errors more harshly when they interfere with communication and, most significantly, when they give the evaluator an “image” of the writer as uneducated, which is a typical reaction to non-standard dialect features. He notes that business executives worry this image of the writer would then subsequently be projected onto the company. Lippi-Green (1997) notes that “pejorative attitudes” toward non-standard dialects are common. She notes “…complaints fall into two categories: targeted lexical items or grammatical features which cause immediate reaction” (Lippi-Green, 1997, p. 179).

Sloan (1979) argues that an increasingly oral culture contributes to an increasing number of errors in student writing. Though Sloan’s conclusion has been contested, his assertions are relevant for our study. First, many, but by no means all, of the errors we used to examine teacher reactions could be argued to have stemmed from influence from oral forms. Second, raters seem to notice oral-like errors more than other types of errors and accordingly punish these more
severely. And perhaps most significantly, some research indicates that it is non-standard oral forms stemming from non-standard dialects, as opposed to errors unrelated to dialects issues, which receive the swiftest correction from teachers (Delpit, 1988, p. 58).

3. Overview of African American English

Of all non-standard American English dialects, AAE has been the most discussed in academic and linguistic circles. AAE is well researched specifically because it elicits the responses noted by Hairston (1981) in the business world. In educational settings, AAE remains the most stigmatized dialect in general (Smitherman, 1977; Spears, 1987) and in writing (Crew, 1977; Ball and Lardner, 1997). A central issue for the current study is whether features of AAE appear in writing and thus receive harsher treatment. Whiteman (1981) and Smitherman (1977) contend that there are indeed AAE features that appear frequently in writing: verbal –s absence (He walk to school every day), plural –s absence (They walk down the street with the radio in their hand.), consonant cluster simplification (He miss the bus yesterday), and is and are absence (She so calm). These are well-known features of AAVE and rarely occur in the speech or writing of white speakers (Whiteman 1981). Many of these features were chosen to appear as errors in our data collection specifically because they are indentifiable as features of AAE. Whiteman (1981) is careful to not attribute all the above features to dialect influence, but she maintains that this is a significant influence.

4. The Study

The study that is presented here has three unique elements that differentiate it from previous studies on rater reactions to written errors. First, many previous studies employed a
questionnaire format in which respondents ranked sample errors (e.g. Vann, Meyer, and Lorenz 1984). Our study used actual essays which were manipulated in such a way that raters would, unknowingly, respond to kinds of errors. Freedman (1979) conducted a similar study in which she used manipulated essays to study rater reactions to four essay features: content, organization, sentence structure, and mechanics. Her study, however, did not insert errors typical of AAE nor ESL and did not focus exclusively on errors. Second, previous studies have compared student writing from AAE speakers to SAE speakers or ESL speakers to SAE speakers, but no studies that we are aware of have compared rater reactions to all three simultaneously. Third, we collected 358 ratings of manipulated essays on an actual college writing competency test which raters evaluated not as part of a study but as part of the regular testing procedure. Other studies asked participants to rate an essay that was provided (e.g. Santos 1988); thus, raters were aware they were part of these studies. The participants in our study were unaware of any data collection and thus simply graded these 358 essays as they would any other essay.

Before detailing the research design and the results, we need to give additional background. First, we need to explain our interest and experiences in this topic because it is relevant to the research design. Second, we need to define terms. And third, we need to explain the historical context of the assessment we used to collect data.

4.1 Anecdotal Evidence

First, our interest in this topic is both academic and anecdotal. Trying to understand why ESL writers delete copulas (and how to address that pedagogically) becomes as interesting and challenging as understanding why AAE speakers delete copulas (and how to address that pedagogically). Though the writing issue is the same for raters (deleted copulas), the linguistic
motivation and pedagogical reaction is quite different. Most raters are not fully aware of the linguistic motivations of these errors, and we believe most raters succumb to a societal view that regards non-standard dialect features in writing as simply substandard and careless writing. Anecdotally, we have heard many raters condemn non-standard dialect features (both AAE and southern American English) as “lazy” or “careless” mistakes and comment that they should result in a failing grade not because of the quantity but because of the severity and linguistic saliency of them. On the other hand, we have noted a tendency to excuse ESL errors because the writers are still learning English. This anecdotal evidence prompted the study.

4.2 Definitions

Second, we need to define key terms which include dialects, markedness, saliency, AAE, and surface errors. Dialects are varieties of a language that differ lexically, phonologically, and grammatically from the standard dialect. A universal law of sociolinguistics is that non-standard dialects will be denigrated in some way. Certainly some non-standard dialects receive covert prestige in various contexts. But by in large, non-standard dialects are denigrated in formal contexts such as academic writing. Milroy (1992) notes part of the condemnation of non-standard dialects is that they “…are often discounted as unstructured” (p. 3). This is similar to Lippi-Green’s (1997) observation that criticism of non-standard dialects includes an appeal to language purity which implies that the standard language has a more inherently logical structure. This view of non-standard dialects as “unstructured” explains in part the negative reaction of raters to non-standard dialect features on standardized English tests because raters look for conformity to the supposedly more logical structure of SAE. This lack of conformity due to the
existence of marked features (linguistic features that are not the dominant form) leads many people (including raters of essays) to discount the users of these features as uneducated.

We acknowledge that features of a dialect are rarely exclusive to that dialect. Distinct dialects may share linguistic features, but when looked at as a whole, dialects are distinguished from each other by a sufficient number of unique features thus constituting separate dialects. It is difficult to discern exactly how many distinctive features constitute separate dialects, but native speakers certainly are aware of different dialects based on only a few features.

A guiding principle for speakers to determine if a dialect is different from their own is saliency. If phonological, lexical, or grammatical features vary from one’s native dialect, which they view as neutral, then they note the differences as salient. For example, British and American English users view the opposing dialect as full of salient features because they view their own dialect as neutral or standard (at least in their local context), so that the other dialect is the one that has salient features. Raters note dialect features in writing that differ from standard English due to the features’ saliency. Linguistic saliency is defined as those linguistic features that are marked as prominent and conspicuous because they are not typical of the standard dialect.

Defining AAE is a somewhat problematic. It has been referred to in many ways in academic and popular settings: African American Vernacular English, Black English, Ebonics, and African American Language (AAL). We reject the term AAL because it feeds the popular rather than linguistic notion that AAE is a separate language. The other terms are somewhat dated, so AAE is the most widely accepted term presently. AAE became a dialect due to its unique history and it remains a distinctive dialect, like all non-standard dialects, due to sociopolitical pressure which motivates speakers to maintain it for identity purposes (Milroy
1992). It is important to note that AAE is not a monolithic whole in the U.S. (Fromkin, Rodman, Hyams, 2010, p. 443). As with all dialects, there is variation within the dialect, but there are sufficient common linguistic features to justify using a single designation. It is equally important to note that not all African Americans use this dialect. As with all dialects, there is a continuum of usage by speakers of that dialect. Despite publicity surrounding AAE, in large part due to the Oakland Resolution on Ebonics in 1994 and the plethora of education and linguistic articles on this issue before and since Labov’s seminal article “The Logic of Non-standard English” (1972), AAE remains a highly stigmatized dialect of American English.

The present study is built upon the notions of surface writing errors. We define surface errors as those non-standard writing (marked) features that do not interfere with communication but would be noted by most raters as non-conformity to the conventions of standard English. Surface errors include spelling, wrong verb forms, punctuation, and syntactic problems that do not obscure meaning (Connors and Lunsford, 1988).

4.3 Historical context

The history of the writing assessment we used to collect data is relevant to this study. The University System of Georgia (in the USA) uses a one-hour writing test, the Regent’s Writing Exam, that rising sophomores must pass in order to continue their university studies. The exam was instituted in 1971. Historically, there has been criticism of the Regents Exam for dialect and linguistic discrimination. There have been critics of the exam that maintain that the creation and institution of this exam was a subtle attempt to maintain racial segregation in the university system through linguistic means. The desegregating of colleges and universities in Georgia had only begun in 1961 with the first three African American students attending the
University of Georgia. The later 1960s saw most institutions of higher learning in Georgia become desegregated but not without considerable protests. Critics of the Regent’s Test maintained that the timing of the test’s institution was suspect and that the nature of the test would lend itself to segregation by linguistic means.

Crew alleges in 1977 that this segregation could be done by training raters to penalize non-standard grammar usages which would apply to “blacks and rednecks” (p. 708). He also maintains that the test was a “useful devise for eliminating a sizeable portion of the population…” (1977, p. 710). He accused that the test was a “…sophisticated collegiate instrument…” similar to literacy test from earlier decades used to deny African Americans the right to vote (1977, p. 710). The historical accusations of racism and discrimination inherent in the test can best be summarized by the charge that African American students were asked to change dialects and become bidialectal, a task few white students were asked to do. As Crew says it is prejudice to “…expect minority students to conform to the myriad of linguistic patterns peculiar to the majority” (Watters, 1979, p. 12).

A linguist from the University of Texas, called the exam “racist as well as irrelevant” (Watters, 1979, p. 10). In 1974, the three predominantly black colleges in the state had a dramatically lower passing rate (37%) in comparison to the University System as a whole (71.4%). This led the U.S. Department of Education to threaten the state of Georgia with a federal discriminatory suit.

But not everyone agreed that the origins of the exam were racist. Charles Nash, a former assistant vice chancellor for academic development for Georgia who helped create the exam, dismissed the criticism during the 1970s. Nash (an African American) noted, “The exams’ whole purpose is to determine strengths and weaknesses of students and then correct weaknesses.
If a large proportion of blacks need weaknesses corrected, then they should be. It would be far more racist not to do that and just perpetuate the mess.” (italics added, Watters, 1979, p. 10).

In 1971, the University System of Georgia did not have many ESL writers, so the issue was a dialect/ethnic issue. But currently the ESL population has risen dramatically. In 2007, the international student population was over 20,000. The Regents’ Writing exam currently has a sizeable number of ESL speakers as well AAE speakers. So rater reactions to non-SAE errors affects a large number of students.

The format of the Regent’s Test has changed little since its beginning in 1971. Currently, there are hundreds of topics used by test administrators and students are given a list of four topics (randomly chosen for each test booklet) to choose from and must write an essay on one topic. They receive one of three grades: 3 - high passing, 2 – passing, 1 - failing. The essays are graded holistically and 2 of the 3 raters must pass the essay or the student must retake the exam. Since it is a one-hour test, no outside research is required; however, students may use a dictionary the last fifteen minutes to check spelling.

Sample topics include:

- Discuss the influence that advertising has had on your life or the lives of your friends.
- If you were asked to make a fair evaluation of your teachers, what criteria or standards would you use for the evaluation? Discuss.
- What influence should students have in the determination of college policies? Explain.
- If your doctor told you that you had only a few months to live, how would you alter your way of life? Discuss.

The training of raters occurs in the following way. First raters are asked to read general descriptions of a THREE, a TWO, and a ONE essay. The complete descriptions are included in
Appendix A. The descriptions describe the three levels of essays in general terms and little mention is made about grammar mistakes except to say that a TWO essay may contain “a few” errors that do not interfere with communication. A ONE essay is described as having serious flaws at all writing levels including “numerous mechanical errors.” Please note that no specific number or type of errors is given. It up to the rater to interpret these “holistic” descriptions.

Before beginning rating, raters are asked to review two essays. The first sample essay is considered on the border between a ONE (failing) and a TWO (passing). The second sample essay is considered on the border between a TWO (passing) and a THREE (high passing). Two short descriptions are given that explain why these essays are border line. Raters are then required to judge actual essays as slightly better or worse than the model “on the border essays” and then rate the essays with a three, two, or one.

5. Methods and Materials

Our research questions were the following:

- Do essay raters penalize essays with AAE errors more than essays with SAE errors on a holistically scored writing assessment?
- Do essay raters penalize essays with ESL errors more than essays with SAE errors on a holistically scored writing assessment?
- Do essay raters penalize essays with ESL errors more than essays with AAE errors on a holistically scored writing assessment?

In order to investigate these questions, we used the following design. First, we used collected three sample student-written Regents’ essays from previous years. These were actual essays written by university students from previous years and graded holistically by Regents’ exam
raters. We used an essay that had failed, an essay that had passed, and an essay that had a high pass. We then corrected all the surface errors in the three exams, but left everything else unchanged. Subsequently, we created an AAE version, an ESL version, and a SAE version of each of the three essays (for a total of nine essays).

The three topics for this study are listed below.

Topic 1 (failing essay): *Do you believe it is the responsibility of the young to provide financial security for the aged? Why or why not?*

Topic 2 (passing essay): *What are most important factors in personal success?*

Topic 3 (high pass essay): *Is increased life expectancy a blessing or a curse?*

The three versions of each of the three essays differed only by the insertion of eight surface errors into each essay. The eight surface errors were errors typical of either AAE, ESL, or SAE writers.

Tables 1, 2, and 3 include samples of the errors that were inserted into the three essays. A complete list of errors is included in Appendix B.

*Table 1 inserted here*

*Table 2 inserted here*

*Table 3 inserted here*

It is important to note that each version of the three essays had exactly eight errors. One or two single errors in the versions would not be a reliable method for noting differences in rater reactions. However, eight was a number we determined (based on a previous pilot study) that caused readers to note errors were of three kinds: AAE, ESL, or SAE. Certainly many of the errors could cross boundaries. A comma splice and an inappropriate preposition are not, in and of themselves, indicative of one language group. However, we were careful to select linguistic
features and a predominance of errors that indicated a pattern that was consistent with AAE, ESL, or SAE. Our guiding principles for selecting errors were *linguistic saliency* and *preponderance*. We wanted to have essays that had linguistically salient errors typical of each language group and have a preponderance of those errors, so raters would sense a pattern of one of the three groups. At one level, this is very easy for linguists and language experts because the saliency of the linguistic features. But for non-linguists, the patterns are more intuitive which is exactly the point of the study. Do raters intuit a pattern of errors that they deem more serious because they are of a certain type?

The selection of errors was based on previous research of AAE, ESL, and SAE language use as well as anecdotal comments from fellow raters. We used Fromkin, Rodman, and Hymas (2010) and Green (2002) as sources for common AAE errors in writing. We consulted Ferris (2005) for common ESL errors. For SAE errors we consulted with fellow raters and avoided errors that could be considered AAE or ESL in origin. AAE included unconjugated “be,” *hisself* for *himself*, and lack of apostrophe –s. ESL errors include “…verb tense and aspect issues, the use of articles and other determiners, noun endings, errors in word form and word order” (Ferris, 2005, p. 41). SAE errors include spelling, punctuation, and subject-verb agreement.

The three essays are included in Appendix C. These are the “clean” versions of the three essays. Using previously rated essays from previous years allowed an objective process for selecting three different levels of essays. The rationale for having three levels of essays was two-fold. First, having three essays increased statistical reliability that raters were responding to different kinds of errors. Second, having three different level essays allowed an analysis of how raters responded to the various errors on low, intermediate and high essays.
Since 2005, the raters do the rating online. The hand-written essays are scanned and raters read a PDF version of each essay on their own computer and select a grade for each. In this study, the three versions of the low essay had the same hand-writing; the three versions of the passing essay had the same hand-writing; and the three versions of the high pass essay had the same handwriting. Thus, we controlled for any effect of handwriting might have on raters.

5.1 Participants

The raters for the USG Regents’ Writing Exam are required to have an M.A. in English or a graduate degree with 18 hours of graduate work in English. Typically there are several hundred raters across the state of Georgia who participate. Robert Barrier (personal communication, 2010) notes that the majority of raters are part-time writing instructors at colleges and universities, but there are also full-time English faculty who participate. We did not collect any demographic data on the participants themselves. The administration tracks who rates each essay, but we deemed it too intrusive to contact raters afterwards and collect any data since they did not know some of their ratings were part of a study.

5.2 Data Collection

Each rater grades the essays online and receives $1.25 for each essay. There are embedded norming essays that the raters must rate. These have been previously rated by the Regents’ testing office as pass or fail. Raters do not know when an embedded norming essay will appear. If a rater misses too many of these, the system will lock the rater out and they will not be permitted to continue rating. This system benefited this study since raters are motivated to read each essay carefully.
The administrator of the exam agreed to insert the nine sample essays. Each essay received between 43 and 54 ratings by different raters. We would like to emphasize again that these essays were inserted into the normal grading process. The raters would have no way of knowing they were rating sample essays for this research. A total of 430 ratings were collected.

There were 72 instances of an essay being rated more than once by the same rater. In 60 of these instances, the rater assigned the same rating to each multiple rating of the essay. In these cases, we replaced the multiple ratings by the one, consistent rating. In four of the twelve cases of inconsistent ratings given to the same essay by the same rater, there was a “majority” rating. We replaced the multiple ratings with the “majority” rating. For example if an essay received ratings of 1, 2 and 1, we replaced the three ratings with the rating of 1. In the remaining eight cases, there was no “majority” rating. For these essays, we randomly selected one of the ratings given to the essay. Because there were so few ratings of 3 (8 out of the 358 essays), we choose to replace the ratings by Pass (a rating of 2 or 3) and Fail (a rating of 1).

After replacing the multiply scored essays, there were 358 essay ratings. The following table indicates the number of ratings used in the data analysis for each version of the three essays.

Table 4 inserted here

6. Results

A log-linear model was used to generate odds ratios for each pairing of the nine essays. The odds ratios were used to compare two different essays with respect to the odds of failure. For example, the odds ratio of failure of a Low/AAE essay compared to a Low/ESL is 4.2. This indicates that a Low/AAE essay is 4.2 times as likely to fail as a Low/ESL essay.
The following table indicates the odds ratios for all pairings of the nine essays.

*Table 5 inserted here*

The number in each cell indicates how likely the row essay is to fail compared to the column essay. For example, the number 4.2 in the row labeled 1 and the column labeled 2 indicates that an AAE Low essay is 4.2 times more likely to fail than an ESL low essay. The number 0.3 in the row labeled 2 and the column labeled 4 indicates that an ESL Low essay is 0.3 times as likely to fail as an AAE Int. essay.

Each of the odds ratios was tested for statistical significance using an alpha level of 0.05. In the table, the statistically significant odds ratios are indicated by asterisks. Some of the significant odds ratios are to be expected. It makes sense that a AAE Low essay would be significantly more likely to fail than a SAE High essay (odds ratio of 15.6). Similarly, a SAE Low essay should be more likely to fail than a SAE High essay (odds ratio of 7.5).

However, there are some significant odds ratios that indicate that AAE essays are at a disadvantage compared to other essays of the same level of quality. An AAE Low essay is 4.2 times as likely to fail as an ESL Low essay. An AAE Intermediate essay is 7.6 times as likely to fail as an ESL Intermediate essay and 6.1 times as likely to fail as a SAE Intermediate essay. Finally, an AAE High essay is 4.2 times as likely to fail as a ESL High essay and 9.1 times as likely to fail as a SAE High essay.

There are some significant odds ratios that indicate that ESL essays have an advantage compared to AAE essays. An ESL Low essay is only 0.3 times as likely to fail as an AAE Int. essay, and an ESL Int. essay is 0.2 times as likely to fail as an AAE High essay.

We also offer the following tables as another way to view the data.
8. Limitations of the Study

It could be argued that raters penalized AAE errors more than SAE errors because the AAE errors were simply more egregious than the SAE errors. In other words, an SAE error such as a comma splice was viewed less critically than a AAE verb error. While this is possible, we reiterate that there are a total of eight errors and approximately 45 ratings of essay. Thus, the number of errors and the number of raters mitigate against this. Given the number of raters, it would be difficult to generalize that punctuation (which some raters consider very serious) was generally overlooked. A more plausible explanation is the types of errors (indicative of AAE, SAE, or SAE) were in fact the factor that caused a disparity in ratings. Also, the ESL errors could be considered as serious as AAE errors and a direct comparison of AAE and ESL essays showed a much higher failure rate for AAE errors.

We acknowledge that a single particular error taken in isolation (e.g. punctuation) might not be considered as serious as another single particular error (e.g. missing preposition or use of *hisself*) on a timed, holistically-scored writing test. The emphasis in the research design was not to investigate reactions to single errors, but to investigate rater reaction to types of errors taken collectively as they represent a non-standard dialect speaker, a non-native speaker, and a standard English speaker.

A second criticism could be the sample size. While we did not have hundreds of ratings, we consider having “ecological sound” or naturalistic data an answer to this criticism. We had
actual essays with actual raters and so despite what some may consider a small sample size, the study employed naturalistic data collection.

We do not consider the results definitively conclusive, but given the result design, the statistical analysis, and the lack of explicit instructions to raters in regard to errors (which caused raters to devise their own system), we do consider the results extremely provocative.

9. Discussion and Conclusion

The odds ratio table and the percentage failure rates above clearly indicate a bias against errors that contain AAE errors and they equally indicate that raters are not as critical of ESL errors. There are several possible interpretations as to why AAE errors, ESL errors, and SAE errors are viewed different by raters.

One interpretation is that AAE speakers are native speakers of English and thus the errors are viewed simply as carelessness and raters are more “annoyed” by them as Santos’s research indicates (1988). A more cynical view is that AAE has been denigrated (based on racial discrimination) and has explicitly and implicitly been used in education as a model for what not to do. Raters have simply absorbed the message that salient AAE features are anathema to good writing. Baugh (1995) acknowledges that educators “harbor stereotypes (about AAE)…often subconsciously” against AAE (p. 87). Raters intuit a native speaker of English who could (with various degrees of effort and various degrees of success) conform to SAE. The fact that these errors persist in essays is taken by raters as a perplexing unwillingness to conform to SAE or a frustrating refusal to write in SAE.

We would like to emphasize that it is not our intention to criticize raters. Raters must adhere to the criteria established for a passing and failing essay. But at the same time, the
discrimination against AAE features is pronounced when in comparison to essays that have an equal number of errors which, in the case of the ESL errors, are equal in gravity. Is racial discrimination, then, an explanation for the disparity in grades? Perhaps the discrimination involved is not an overt form but rather an internalized derision for AAE features, and this derision for these AAE features does have historical origins in discriminatory and segregationist thinking as Crew (1977) maintains.

Rubin and Williams-James (1997) in their study found linguistic bias as well but not necessarily overt racism. In their study, the raters, who knew the identity of essays writers (fabricated in their study as either Asian or Northern European), did not penalize them based on ethnicity. Rubin and Williams-James found raters being more “lenient” with Asian writers due to Asian languages’ “distance from English” which raters apparently considered when assigning grades. And Spears (1987) notes, despite growing linguistic evidence about AAE, most teachers are simply unaware of the linguistic analysis of AAE which specifically establishes AAE features as dialect features and not as errors. In other words, the disdain for AAE features continues (often unconsciously) in a way that it does not for other grammar errors due to long-standing stereotyping of these errors.

A more positive pedagogical interpretation is that raters may be simply trying to prepare students for the “real-world” where AAE errors will be less tolerated by potential employers, so raters fail them now in the hopes that the students will address these errors. This brings into question the pedagogical wisdom of giving students a “right” to their dialect. While sociolinguistic sound, this study indicates that this approach is not always the panacea some educators would like it to be. The “real-world” of standardized writing tests and job applications will most likely continue to penalize AAE features more harshly.
This discrimination against AAE features may explain, at least in part, the achievement gap on these standardized tests. Many AAE speakers score well below their SAE speaker counterparts particularly on assessments that have free responses and essay components. A bias against AAE could be a partial explanation. There has been much more discussions in educational settings of the unique linguistic needs of ESL students, and, as a consequence raters, are more sensitive to and less critical of ESL errors. Raters understand that native-like fluency in writing is not always a realistic goal for ESL students, so errors are more tolerated even on standardized tests, while no such leniency exists for AAE speakers.

Our results indicate and confirm a long-standing sociolinguistic principle: stigmatized varieties of a language suffer in formal educational settings. This empirical study demonstrates that AAE errors are viewed quite differently from ESL errors. We suspect that had this study been completely 40 years ago when the Regents’ Writing Exam was instituted, both AAE and ESL errors would have been viewed far more similarly. Education and dissemination of information about ESL issues have affected raters in regard to ESL writing errors. While there has been more discussion about AAE in recent years, it remains more controversial and discriminated against than ESL writing issues in university settings. It is possible that a better understanding of AAE on the part of raters would change the results of this study. The perplexing pedagogical question of how to teach SAE without denigrating non-standard dialects remains an open issue. Prescriptive attitudes toward errors are a hallmark of rater practices. These prescriptive attitudes “…(are) undoubtedly a transfer from attitudes toward less educated native speakers as well as attitudes ingrained after years of prescriptive education” (Santos, 1988, p. 84). This study and its findings support this notion.
Based on previous research, it is not surprising that our results indicate linguistic discrimination, but what is surprising and provocative is the extent of the discrimination against AAE errors.

Acknowledgements

We would like to thank Robert Barrier, Leslie Caldwell, and Lissa Small for their assistance in conducting the research. We would like to thank Cherif Diop and Chris Palmer for their comments on earlier drafts.
References


Appendix A Instructions for Raters

INSTRUCTIONS FOR SCORING REGENTS’ TESTING PROGRAM ESSAYS ©

DESCRIPTION OF ESSAY SCORING PROCEDURE

Raters should read each essay quickly to gain a general impression of its quality. This approach, holistic rating, contrasts with the analytic grading commonly used in essay evaluation.

The essays are rated on a three-point scale in which “1” is a failing score and “2” and “3” are passing scores. The model essays represent borderline cases; each essay to be rated must, by definition, fall above or below a model. One model essay represents each dividing line.

<table>
<thead>
<tr>
<th>RATINGS</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>MODELS</td>
<td>2/1</td>
<td>3/2</td>
<td></td>
</tr>
</tbody>
</table>

Raters should compare the essays they read with the models. They should not rate in terms of their usual grading standards or some abstract standard.

The most important task for a rater is to determine whether an essay is better than the 2/1 model essay. An essay worse than the “2/1” model receives a failing score of “1.” An essay better than the “2/1” model but not as good as the “3/2” model receives a grade of “2.” An essay better than the “3/2” model receives a grade of “3.”

Raters should keep in mind that students have one hour to compose an essay on a choice of assigned topics. The essay should not be evaluated as a final product that the student might be able to produce after additional time for reflection and revision. The model essays are chosen based on reasonable expectations for writing samples produced in one hour.

The Testing Subcommittee of the University System Academic Committee on English attempts to choose models by using the following definitions of competency, although it realizes that these definitions are by no means exhaustive.

3: The “3” essay shows distinction. It meets all and exceeds most of the criteria for a “2.” The ideas are expressed freshly and vividly, and the essay arouses the reader’s interest to a greater extent than the typical “2” essay.

2: The “2” essay meets the basic criteria. It has a central idea related directly to the assigned topic and presented with sufficient clarity that the reader is aware of the writer’s purpose. The organization is clear enough for the reader to perceive the writer’s plan. The paragraphs coherently present some evidence or details to substantiate the points. The writer uses ordinary, everyday words accurately and idiomatically and generally avoids both the monotony created by series of choppy, simple sentences and the incoherence caused by long, tangled sentences. Although the essay may contain a few serious grammatical errors and several mechanical errors, they are not of sufficient severity or frequency to obscure the sense of what the writer is saying.
The “1” essay fails to demonstrate competence. It has any one of the following problems to an extraordinary degree or it has several to a limited degree: it lacks a central idea; it lacks a clear organizational plan; it does not develop its points or develops them in a repetitious, incoherent, or illogical way; it does not relate directly to the assigned topic; it contains several serious grammatical errors; it contains numerous mechanical errors; ordinary, everyday words are used inaccurately and unidiomatically; it contains a limited vocabulary so that the words chosen frequently do not serve the writer’s purpose; syntax is frequently rudimentary or tangled; or the essay is so brief that the rater cannot make an accurate judgment of the writer’s ability.

2/1 MODEL 2/1

TOPIC: WHY WOULD YOU LIKE OR DISLIKE OWNING YOUR OWN BUSINESS?

Going out of Business Sale! Signs of this nature can be seen everywhere. Today opening up a business can be scary, because of the extensive risk, high cost, and extreme stress.

The chief reason I would not want to start my own business is the great risk of failure. Today statistics show that four out of every six businesses fail within the first year. Those are not very good odds for one just starting his or her own business.

The second reason not to start my own business is the high cost of starting a business. Businesses take a great deal of money to get started, and for that matter to keep running. The first thing one has to do is find a place to put the business. Lots are very expensive. Then a building has to be built, and merchandise to fill the building has to be purchased.

Finally owning a business can be stressful. Being one's own boss can be stressful to her or him by the way of having to make all of the important decisions, or can cause stress at home. The stress at home can be very detrimental to the marriage, or even the family as a whole.

Concluding this owning a business is just one big headache. On the other hand some people are very successful, and they got that way by taking the risk of owning their own business. I personally don’t think that owning a business is worth the risk, when working for someone else is a lot safer.

3/2 MODEL 3/2

TOPIC: WHAT ANIMAL DO YOU LIKE (OR DISLIKE) THE MOST? WHY?

In my family there are numerous animal lovers, including me. Though I love all of God's creatures, I must say that I am lacking in emotions for cats. It is not that I hate the creatures. I have had bad experiences with them in the past.

The first feline that I owned was a pure-bred Persian. A beautiful cat to say the least, but its beauty was deceiving. It constantly shed its fur all over the house. No where could I run to find a haven from my newly discovered allergies.

The second feline that I owned was a tabby. This cat was a fun animal to own. It was constantly playing and batting things around. All was well until it matured. When it got its claws in, my furniture became
the clawing areas and when it went forth into the surrounding neighborhood it would bring back to me a rabbit, squirrel, bird, etc. that it would politely leave on the porch. I suppose that it was a gift, but I could not make the cat understand that I did not want it.

Now that I had owned two cats, I was not happy when my best friend got me another for my birthday. I immediately began trying to get rid of it, but nobody wanted it. I was reluctant to take it to a shelter because it might end up being put to sleep. I wanted to know that it would have a home, but not mine.

As my luck goes with cats, it would meow in the most horrific scream every night. After two nights of this torture I took it to my best-friend's house and left it with her parents. I told them to give it to her as a gift. She kept it and still has it. I am glad that she kept it and did not take it back to where she got it.

As for me, I prefer to own dogs, fish, parrots, and my horse Trigger. Why? Because none of these animals has ever destroyed my furniture, made me sick, brought dead animals to my house, or kept me up all night for no reason. My dogs are excellent guards. My horse Trigger is an old friend that I have grown up with. My fish are silent, beautiful creatures of the deep and my parrot is a most intelligent bird. They all keep me company, and they all make me happy.
Appendix B

Low Essay

_AAE Version Features_
1) The reasons are that different culture’s has…(agreement)
2) The reasons are that different culture’s has…(apostrophe – s in place of plural -s)
3) …is the theory that with the growth of the baby boomers in the industrialize areas…
   (consonant cluster simplification)
4) …the country population will become…(apostrophe – s deletion)
5) …population will become dependent on the government to pay for they needs (they for their)
6) The propose solution…(consonant cluster simplification)
7) If the idea of more lifelong workers was introduced…(lack of subjunctive format)
8) There no one solution… (deleted copula)

_ESL Version Errors_
1) The question of should the young provide financial security to the elderly is a difficult
   question for answer. (preposition error)
2) The reasons are that different cultures have varying values with regard to the older members
   of population. (article deletion)
3) …elderly are highly honored members of society and in the United States, is welfare to pay
   for the needs of the elderly. (it deletion)
4) The money for welfare comes from taxes paid for by younger population. (article deletion)
5) If the young had continue paying for the needs of the older population…(infinitival to
deletion).
6) … to the problem of welfare are more workers and more training to the older population on
   new technology. (preposition error)
7) … If the idea of more lifelong workers were introduced in the western nations that will be
   experiencing an increase in the amount in the older population, could help the problem
   because (it deletion)
8) If the government encouraged more lifelong workers, the economy would not become
   slowing. (word choice)

_SAE Version Errors_
1) The reasons are that different cultures has varying values with regard to … (subject-verb
   agreement)
2) _If the young had to continue paying for the needs of the older population_. (fragment)
3) The _suggestion_ of the welfare state is the theory that with the growth of the baby boomers in
   the industrialized areas in the world…(word choice)
4) This would mean higher taxes to pay for the financial needs of the elderly, this would also
   cause fewer jobs to be available for the young…(comma splice)
5) All this could happen if the young _was_ to continue paying for the needs of the elderly. _was
   for were_
6) If the idea of more lifelong workers _was_ introduced in the western nations that will be
   experiencing an increase in the amount in the older population…(lack of subjunctive format)
7) Their is no one solution to the problem of should the young provide for the needs of the
growing aged population. (their for there)
8) The leaders and governments of the world (spelling)

Intermediate Essay

AAE Version Features
1) Being successful start with wanting to be successful. (third-person –s deletion)
2) The individual have to want it. (subject-verb agreement)
3) Only when someone pushes hisself does he see how disciplined he is. (hisself for himself)
4) It important that you are disciplined enough to use the time that you have allowed yourself. (contractable be deleted)
5) The peace of mind will keep you focus on your goals. (consontant cluster simplification)
6) The kinds of activities you do in your free time is important. (subject-verb agreement)
7) For example, my brother friends have no goals and bad habits. (apostrophe –s deleted)
8) I believe it play the biggest part. (third-person –s deletion)

ESL Version Errors
1) Being successful starts with wanting be successful. (infinitival to missing)
2) Only when you push yourself do you see how disciplined are you. (word order)
3) The time set aside should be free for distractions. (wrong preposition)
4) You should this time use wisely to achieve the task for the day. (word order)
5) Is important that you are disciplined enough to use the time…(It deletion)
6) Not using time wisely will set up you for failure. (word order)
7) The peace of mind will keep you focused on some goals. (some instead or your)
8) I believe plays the biggest part. (It deletion)

SAE Version Errors
1) Being successful starts with wanting to be successful the individual has to want it. (run-on sentence
2) A great plan will give you a foundation to begin with. (ending s sentence with a preposition)*
3) The time set aside should be free from distractions. (spelling)
4) …where you spend your free time, and the friends you surround yourself with are all a key to your success. (agreement)
5) Living in a clean organized area, will give you peace of mind. (unnecessary comma)
6) The kinds of activities you do in your free time are important, these activities should be positive influences. (comma splice)
7) To achieve goals, manage time, and have a positive atmosfere, self-discipline is needed. (spelling)
8) Adding a little self-discipline in your life can be the difference between success, and failure. (unnecessary comma)

*Many writing teachers still view ending a sentence with a preposition as an error.
**High Essay**

**AAE Version Features**
1) As modern technology improve, so do our life expectancy. (deletion of third-person –s)
2) As modern technology improve, so do our life expectancy. (do for does)
3) There be many reasons why it could be either one, but...(unconjugated copula)
4) Study’s have shown that a low fat…(apostrophe –s instead of plural)
5) … but will also increase longevity for hisself. (hisself for himself)
6) Even though our government save social security benefits for each of us…(deletion of third-person –s)
7) My father company was able to help with this retirement planning. (deletion of apostrophe–s)
8) Some might argue that he don’t want to live to be ninety or one hundred. (don’t for doesn’t)

**ESL Version Errors**
1) As modern technology improves, so our life expectancy…(deleted does)
2) Every year people lives longer due to the knowledge that medical science has gained. (subject-verb agreement)
3) Some might wonder this gain is a blessing or a curse. (omitted if)
4) An important factor living long life is eating and living heathy…(omitted preposition)
5) Walking just thirty minutes a day can vastly improve your cardiovascular system, which is essential for to living a long life. (preposition error)
6) Even though government saves social security benefits for each of us (deleted article)
7) Even though government saves social security benefits for each of us, is barely enough to survive. (deleted it)
8) …they will also see many joyous births of new beautiful babies that will enrich their lives and fill them of joy and happiness. (preposition error)

**SAE Version**
1) As modern technology improves so does our life expectancy. (punctuation error)
2) Every year; people live longer due to the knowledge that medical science has gained. (misuse of semi-colon)
3) An important factor to living a long life is eating and living heathy. (spelling)
4) …will not only make you look and feel healthy, but will also increase your longevity. (punctuation error)
5) Walking just thirty minutes a day can vastly improve you cardiovascular system, which is essential to living a long life. (you for your)
6) These people must remember that even though he must witness many deaths of loved ones… (wrong pronoun)
7) Some might argue that they are worried that when they get extremely old, their body will deteriorate, they are scared that their vision will go bad, as well as their hearing. (comma splice)
8) If you want to live a long, healthy life, be sure to plan for it properly by taking care of your finances, enricing you mind (spelling)
Appendix C

Non-error version of Low Essay

Do you believe it is the responsibility of the young to provide financial security for the aged? Why or why not?

The question of whether the young should provide financial security to the elderly is a difficult question to answer. The reasons are that different cultures have varying values with regard to the older members of their population. In Asian cultures, the elderly are highly honored members of society and in the United States, there is welfare to pay for the needs of the elderly. The money for welfare comes from taxes paid for by the younger population. If the young had to continue paying for the needs of the older population, there could be some difficulties in the future. These include the idea of the welfare state and a slower economy.

The idea of the welfare state is the theory that with the growth of the baby boomers in the industrialized areas in the world, the country’s population will become dependent on the government to pay for their needs. This would mean higher taxes to pay for the financial needs of the elderly. This would also cause fewer jobs to be available for the young, and the economy would be slower because of the lesser amount of people working. The sales of goods like homes and cars would decrease and new companies will have a harder time succeeding in the slower economy. All this could happen if the young were to continue paying for the needs of the elderly.

The proposed solutions to the problem of welfare are more workers and more training for the older population on new technology.

If the idea of more lifelong workers were introduced in the western nations that will be experiencing an increase in the amount in the older population, it could help the problem because if more people remained in the workforce taxes would not have to raised to pay for the financial needs of the older population. If the government encouraged more lifelong workers, the economy would not become slower. Also, if the idea of lifelong workers were to succeed, more training in new technologies and new job fields would have to be made.

Overall, if the young had to continue paying for the needs of the elderly, the idea of the welfare state could happen and the economy could be worse because of this. There is no one solution to the problem of whether the young should provide for the needs of the growing aged population. The leaders and governments of the world should research the best methods of dealing with the problem. Only through this can good solutions be made and the welfare state avoided.

Non-error version of Intermediate Essay

What are most important factors in personal success?

Self-discipline is the most important ingredient for success. Being successful starts with wanting to be successful. The individual has to want it. Developing a plan is the first key to success. A great plan will give you a foundation on where to begin. You have to set goals throughout the plan. Reaching those goals are a great measure of how disciplined you are. The goals need to be obtainable, but not too obtainable. If the goals are being reached with ease, you
might want to restructure your goals. Only when you push yourself do you see how disciplined you are. The goals cannot be out of reach. If the goals can never be achieved, you will become discouraged.

Management of time is another important factor. You need to set aside time to work on goals. The time set aside should be free from distractions. You should use this time wisely to achieve the task for the day. It is important that you are disciplined enough to use the time that you have allowed yourself. Not using time wisely will set you up for failure.

Surrounding yourself in the correct atmosphere is a very important key to success. The place you live, where you spend your free time, and the friends you surround yourself with are all keys to your success. Living in a clean organized area will give you peace of mind. The peace of mind will keep you focused on your goals. The kinds of activities you do in your free time are important. These activities should be positive influences. The friends you surround yourself with should have similar goals and habits. Having many friends with no goals and bad habits will bring you down.

Self-discipline plays a part in success. I believe it plays the biggest part. To achieve goals, manage time, and have a positive atmosphere, self-discipline is needed. Adding a little self-discipline in your life can be the difference between success and failure.

Non-error version of High Essay

Is increased life expectancy a blessing or a curse?

As modern technology improves, so does our life expectancy. Every year people live longer due to the knowledge that medical science has gained. Some might wonder if this gain is a blessing or a curse. There are many reasons why it could be either one, but with proper planning, one could live a long and enjoyable lifetime.

An important factor to living a long life is eating and living healthy. Studies have shown that a low fat, high fiber diet full of plenty of fruits and vegetables will not only make you look and feel healthy, but will also increase your longevity. Walking just thirty minutes a day can vastly improve your cardiovascular system, which is essential to living a long life.

Another important step to take if you want to comfortably live to a ripe old age is to plan financially. Even though our government saves social security benefits for each of us, it is barely enough to survive. One must begin planning for retirement as early as their twenties with 401 K’s and savings plans. My father’s company was able to help with this retirement planning.

Some might argue that they don’t want to live to be ninety or one hundred. They don’t want to watch all their friends and family die around them. These people must remember that even though they must witness many deaths of loved ones, they will also see many joyous births of new beautiful babies that will enrich their lives and fill them with joy and happiness. They will also be fortunate enough to experience a lot of history, and if they planned well financially, see a lot of the world as well. These experiences will make them wise and worldly.

Some might argue that they are worried that when they get extremely old, their body will deteriorate. They are scared that their vision will go bad, as well as their hearing. They are worried their bones and muscles will crack and ache. They are afraid their mobility will be challenged. For these people, I suggest practicing yoga or another form of daily stretching to
keep their body in top shape. Make sure and eat plenty of vitamins and drink lots of water every day. You should not smoke, and you should keep alcohol consumption to a minimum.

A long life can be a wonderful blessing. Throughout a long lifetime, one can witness many great historical moments and learn many great things. If you want to live a long, healthy life, be sure to plan for it properly by taking care of your finances, enriching you mind, and most importantly, taking care of your body and soul.