Writing Differences in Teacher Performance Assessments: An Investigation of African American Language and Edited American English

1MICHELLE Y. SZPARA and 2E. CAROLINE WYLIE
1Long Island University in New York, 2Educational Testing Service, Princeton, USA

Differential performance results occur when a specific population subgroup achieves a passing rate which is significantly lower than that of the normative reference group. African Americans do less well, in general, on all types of assessments, including constructed-response tests. The present study examined the writing styles of African American test takers in a National Board for Professional Teaching Standards (NBPTS) portfolio entry from the Middle Childhood/Generalist Certificate, as compared with the writing styles of European American test takers. The use of features other than Edited American English was examined as a possible source of construct-irrelevant variance in assessors’ scoring judgments. Thirty-two written commentaries, 18 from African American candidates and 14 from European American candidates, were coded for grammatical, lexical, and discourse features. The coding frame identified features of African American Language (AAL) and Speech Code Errors (SCE). Instances of AAL were fewer than instances of SCE and clustered according to potential users of AAL.

INTRODUCTION

Due to increased scrutiny on student achievement, teacher assessment is also being emphasized, from pre-service to master teaching. These assessments cover basic skills, curriculum, and pedagogy, and utilize multiple-choice, constructed-response, observation, and portfolios to measure what teachers know and are able to do.

While debates over the advantages and disadvantages of each assessment form are numerous, they share one attribute—the differential performance of minority teachers. African Americans do less well, in general, on all assessment types, including constructed-response; this issue must be addressed by test makers and users of test scores (Ladson-Billings 1998; Mickelson 2003). This study investigated potential writing differences between African American and European American candidates in a National Board for Professional Teaching Standards (NBPTS) portfolio entry from the Middle Childhood/Generalist Certificate (MC/Generalist).
In the assessment under study, qualitative evidence (Onafowora 1998) suggested that some candidates might be using forms of English other than Edited American English, including African American Language and speech code errors (grammar and spelling). As part of the training for assessors who score the portfolio, there are multiple activities focused on bias reduction, including specific attention to influence of writing style. Assessors, like all individuals, hold preferences for specific qualities in writing. It is possible that, without conscious attention to the scoring rubric, assessors could allow perceptions of a candidate’s writing style to influence their evaluation of content. This necessitates examination of the candidates’ writing style and form in order to explore this potential source for observed differential performance.

This study occurs in the US context of national testing for educational accountability and continued disparities in social and economic opportunities for low-income and minority racial/ethnic groups. Under the federal No Child Left Behind Act, public school students in grades 3–8 are expected to take annual assessments in key subjects. Results are made publicly available, and schools’ funding may be affected in accordance with students’ progress. Individual states require various tests for teacher certification, while NBPTS provides voluntary national testing to certify accomplished teachers. Differential performance is seen across these assessments for groups which are economically disadvantaged or are racial/ethnic minorities (Bennett et al. 2006). One possible explanation includes continuing covert racism and classism which has hampered the educational and career progress of some groups in the USA (Stanley 2006). This study examines language differences in the portfolio writings of specific racial/ethnic groups, in order to establish groundwork for future studies of possible differences in assessors’ evaluations of these writings.

DIFFERENTIAL PERFORMANCE OF AFRICAN AMERICAN TEST TAKERS

The differential performance of African American test takers is seen across all types of constructed-response tests (Jencks and Phillips 1998), both within educational assessments (McIntosh and Norwood 2004) and in other professional assessments (Ogbu 2002). While the gap was decreasing over time (Jones 1984), it may now be increasing (Lee 2002). Differential performance may be due to factors within the assessment or scoring design, or with scorers (Bond 1998a, 1998b). The causes may also lie outside the assessment, specifically in educational opportunities provided to test takers in grade school, college, and post-graduate professional development (Farkas 2003). Equity of access to resources such as test preparation materials, release time from work, and mentoring may also play a role (Ladson-Billings 1998). Finally, researchers have explored interactional factors, including
stereotype threat (Steele 1997) and use of culturally and linguistically distinct modes of expression in assessments (Irvine and Fraser 1998).

FOCUS FOR THE RESEARCH STUDY

This study employs a portfolio-entry assessment to examine more closely the writing of African American candidates—specifically language choice, rhetorical style, and organization. For the NBPTS MC/Generalist certificate, candidates submit multiple components including portfolio entries which focus on classroom performance. The validity of this assessment has been documented through a content-related validity study (Benson and Impara 1997), demonstrating appropriateness for identifying evidence of accomplished teaching. Studies have also documented the value of NBPTS certification in recognizing the skills of accomplished teachers, and specifically the added value that these teachers bring to the classroom (Bond et al. 2000). In previous studies attempting to examine sources for differential performance on this assessment, there has been no evidence of discernible biases within the content of the assessment, test directions, or test questions (Bond 1998a, 1998b).

The MC/Generalist certificate was selected, because it has the greatest number of African American candidates (NBPTS 2006a). The specific entry Writing: Thinking Through the Process was selected, as it did not include a videotape (NBPTS 2006c). It was desirable to select an entry where a candidate’s race/ethnicity would not necessarily be obvious. This entry requires a candidate to select two assignments/prompts to demonstrate how they use writing to develop students’ thinking and writing skills for different audiences or purposes. In a twelve-page written commentary, the candidate describes the teaching context, assignments, and how they fit into the overall writing curriculum. In addition, the candidate is required to analyze two students’ writing and reflect on the success of using both assignments.

Portfolio responses are scored holistically by teacher-peers, who have completed a rigorous four-day training. The training emphasizes recognizing and screening out personal and societal biases. The assessors complete four bias-reduction exercises; one focuses on minimizing the impact of potential writing biases on the scoring process.

TERMINOLOGY

The terminology describing the language used by some African Americans has changed over the last three decades. While recognition of the differences in language between African Americans and other English speakers dates to the eighteenth century, the term ‘Ebonics’ is relatively young (Smitherman 2001). It was first coined at a conference on language and African American children in 1973. From there, just as descriptors for African Americans
have evolved, so too have terms used to refer to Ebonics. Following the writing of one eminent researcher whose work spans several decades, one can observe these shifts in terminology. Smitherman (1973), in an earlier published work, uses the term Black English. In a 1997 article (Smitherman 1997a), she charts how Black English changed to Black Vernacular English, and then to African American English. In an article later that year (Smitherman 1997b), the term experienced a further shift to African American Language (AAL).

In consultation with a national panel of linguists experienced in the study of African American Language, it was agreed that the term AAL would be used. Similarly, it was agreed that equanimity in naming would best be achieved through the use of ‘African American’ and ‘European American’, rather than ‘Black’ and ‘White’.

The final set of terms involves the categories into which different language patterns were divided. These include African American Language (AAL), Edited American English (EAE), Southern White English (SWE), and Speech Code Errors (SCE). The use of EAE emphasizes the edited nature of formal prose, written under constraints of the assessment directions. The phrase ‘written edited American English’ appears in a 1974 resolution by the National Council of Teachers of English entitled, ‘On the Students’ Right to Their Own Language’. The term ‘Standard American English’ was avoided, in part, because it privileges American English over other forms of English, including AAL. Southern White English refers to a dialect of English specific to the Southeastern United States, which developed in parallel with African American Language (Feagin 1997). Speech Code Errors encompass errors made by either users of AAL or EAE that do not conform to consistent grammatical structures.

African American Language

The existence of AAL as a systematic, rule-governed, oral language is well documented (Labov 1969; Baugh 1983; LSA 1997). Writings by authors such as Geneva Smitherman and Toni Morrison, to name only two, incorporate AAL in their works, but overall less research has been conducted on written forms of AAL. Exceptions to this are studies of AAL in hip-hop and rap music (Smitherman 1997b) and in student writing, discussed below.

Qualitative evidence has suggested the presence of AAL features and use of African American Cultural Indicators (Onafowora 1998) in a subset of written assessments produced by African American teachers. The use of different linguistic codes, specifically AAL, has been well documented in the writing of grade school and college students of color (Ball 1992; Blumenthal and Hildenbrand 1993). This use of linguistic codes and cultural references distinct from EAE and majority-group cultural references (European American and middle class) has not been explored extensively in performance-based teacher assessments.
It is important to note that, while there are distinct patterns of AAL usage within texts that consist of predominantly Edited American English, the patterns exist at a very low frequency (Smitherman 1992). Even at a minimal level, AAL usage holds potential to impact the reader and have the undesired effect of producing stigmatization of the writer (Mohamed 2002). Although a reader may attempt to overlook grammatical and syntactical features inconsistent with EAE, the presence of such features may still cause the reader to devalue content (Santos 1988).

Smitherman (1992) conducted a longitudinal study of African American student writing. Among the features identified were the absence of ‘ed’ morpheme in past tense verbs, perfect tense verbs, verbal adjectives, and passives; the absence of ‘s’ morpheme in plural nouns, possessive nouns, and third-person singular verbs; hypercorrection; zero copula; subject–verb agreement; multiple negation; and pronominal apposition. These features formed the basis of the present study’s coding frame, later supplemented and refined by the national expert panel.

Fogel and Ehri (2000) conducted a study focused on six syntactic features that differed according to the form of spoken English used by students. The authors examined syntax since ‘unlike phonological features, nonstandard syntactic forms tend to stand out and, as a result, can be stigmatizing (Burling 1973)’. The features included possessive ‘s’, past tense ‘ed’, third-person present-tense singular ‘s’, plural ‘s’, indefinite article, and subject–verb agreement.

Ball (1998) analyzed a student paper written by an African American Vernacular English-speaking (AAVE) high-school sophomore. She identified a number of features common to many AAVE speakers, such as absence of third-person singular present tense ‘s’, and a range of semantic and phonological patterns. One pattern of interest was the student’s use of ‘willn’t’ instead of ‘won’t’. Ball noted this as a possible example of what Wolfman and Whiteman (1971) identified as hypercorrection.

Hypercorrection results from the writer’s attempt to produce academic patterns with which he or she is not completely familiar. Because of this unfamiliarity, the writer incorporates items not only where they are appropriate, but in inappropriate places as well. This phenomenon appears in spelling, vocabulary choices and syntactic structures, and is not exclusive to AAVE speakers. (Ball 1998: 235)

Smitherman (1994) and Ball (1995), among others, draw a distinction between syntactic and discourse features present in the writings of African American students. At the discourse level, typical features are ‘use of rhythmic language, anecdotes, parables, and patterns of repetition and call and response’ (Ball 1995: 259; Bohn 2003). Within the portfolio under study, given that the format is highly structured, candidates tended to present fewer of these features. They were included in the coding
frame to examine frequency and provide a more complete picture of any features that might contribute to construct-irrelevant variance in assessors’ scoring.

Smitherman (1994) noted that some students used what she described as ‘conversational tone’—an essay that read like ‘recorded language or a conversation’ (Chaplin 1987: 37). This was identified as ‘use of oral language in written form’ in this study.

Southern White English

AAL is spoken by a subset of the African American population that has been exposed to it in the home/community. Smitherman (2001) estimates that ‘90% of the African American community uses one or more aspects of the language some of the time’. AAL is distributed across the USA and is not limited by socioeconomic boundaries. In comparison, Southern White English is a localized dialect of American English which is largely constrained to the southeastern USA (Bailey 1997; Feagin 1997). SWE has been extensively documented in references such as the Bibliography of Southern American English (McMillan and Montgomery 1989) and the Dictionary of American Regional English (Cassidy and Hall 1996), and demonstrates distinctive patterns in grammar, pronunciation, vocabulary, naming practices, and word-play. SWE shares many features in common with AAL.

While the developmental histories of SWE and AAL are debated among linguists, it is generally agreed upon that the two languages developed in parallel due to cross-exposure between speakers (Winford 1998). As part of maintaining European-American dominance during the period of slavery, enslaved Africans were deliberately placed in mixed-language groups. In addition, it was illegal for the enslaved Africans to learn to read or write. Since most enslaved individuals had limited contact with English speakers, they were forced to develop a working communication system referred to as a pidgin (Escure and Schwegler 2004; McWhorter 2005). This pidginized language was a combination of West African languages and English. Over time this language was passed down to their children, became creolized, and has been maintained as part of the cultural identity of African Americans today (Gilbert 2002).

It is hypothesized that enslaved individuals who cared for young children of European American plantation owners had a strong effect on the children’s initial language development, bringing the influence of the nascent African American Language to bear on SWE. Evidence also exists in post-slavery times that the intermixing of African Americans and European Americans can lead in certain cases to the altering of what is defined as ‘Standard American English’ (Bailey 2001).
IMPACT OF AAL AND NON-EDITED AMERICAN ENGLISH

Little work has been conducted regarding impact on the writer, but the fact that many writers have been striving to have AAL recognized as a language and not just as ‘poor grammar’ indicates an underlying assumption of negative connotations. Smitherman (1992) states that ‘the question of the relationship between writing scores and distributions of BEV [Black English Vernacular] in writing is a logical issue given assumed negative teacher attitudes and the stigma attached to features of Black English in writing’. Ball (1998) discusses strategies for teachers working with students who use AAVE (see also Ball and Lardner 2005). This lack of understanding is not limited to European American teachers, as Ball (1998) notes that ‘most teachers, including many African American teachers, are unaware of the subtle features that characterize AAVE as a distinct linguistic system’. Erickson (1984) describes it as a ‘low-prestige’ language, and Anderson (1981) notes that employers used non-mainstream English as a criterion for eliminating otherwise well-qualified candidates applying for positions that did not require proficiency in mainstream/academic English.

Research in teaching English language learners (ELL) provides insight into the stigmatization of non-standard English writing. Researchers are divided on whether readers can, with training, overlook language errors and focus on content. According to Leki (1991), faculty are able to overlook local errors—those that do not interfere with comprehension. However, this ‘allowance’ may not be extended to native speakers’ writing; errors were perceived as indications of ‘lack of maturity of thought and of rhetorical style’. Albrechtson et al. (1980) found that the specific type of error was of less consequence; what counted was the frequency. Two of the most irritating and least acceptable errors were double negatives and subject–verb agreement, both of which have parallels in AAL. ELL research has shown that, due to lack of awareness about second-language learning, faculty mistakenly assume that some common errors are ‘careless’, and therefore they may feel justified in penalizing students (Kroll 1991). Since many teachers are unaware of the existence of AAL, they may assume that AAL patterns are representative of poor English grammar.

RESEARCH QUESTION

This research addresses the question of whether there is systematic evidence of the use of African American Language in written commentaries of teaching practice prepared by African American teachers. The research sought to systematically look for language patterns by candidates’ race/ethnicity and score. In other words, do language patterns differ for African American and European American candidates in this portfolio entry? If differences in language patterns are found, do they vary according to high and low scores on the portfolio entry?
The present study is the first phase of a larger project examining differences in the writing of African American and European American candidates. This report focuses on grammatical and syntactic features. In the second phase, discourse level features and cultural indicators, such as the use of culturally relevant pedagogy, will be examined.

It is important to note that this study, in isolation, cannot answer the question of whether there is a causal relationship between language features and scores. In other words, even if AAL features do cluster for low scoring African American candidates, it would not be possible to identify whether low scores were awarded because of rubric-based decisions (evidence of poor pedagogy) or because the writing influenced assessors inappropriately (Szpara and Wylie 2005). Further research is required.

METHODOLOGY

In order to develop a coding frame for systematically documenting the presence or absence of AAL features, a national panel of experts was assembled. A potential list of names was compiled from publications and presentations on AAL by professionals in education, literacy, and linguistics. These individuals were contacted with a request to participate. If they were unable to participate, they were asked to identify other qualified professionals for consideration. Seven experts, including five who were African American, agreed to participate in an initial review of MC/Generalist responses.

The panelists collaborated in face-to-face meetings and online consultations to develop the coding frame. In the meetings, the authors presented an initial list of AAL features based on current literature and sample candidate responses. Over multiple sessions, the panel developed an exhaustive list of all features of the written commentaries that were distinct from EAE. These were arranged into three categories:

1. African American Language
2. African American Cultural Indicators

The AAL category included many features common to the oral variety (Labov 1969), such as absence of ‘s’ when plurality is marked elsewhere, absence of ‘ed’ when aspect is indicated by other syntax, and absence of ‘s’ on third-person singular verbs when meaning was contextually clear. More salient features of spoken AAL, including double negatives and zero copula (‘to be’), were rare or absent in the written texts.

The African American Cultural Indicators (AACI) included features such as culturally relevant pedagogy (Ladson-Billings 1998), socially-conscious curriculum, and ‘high talk’ (Smitherman 1977). Culturally relevant pedagogy involves utilizing students’ own culture in school to maintain it and counteract the potential negative effects of a dominant non-minority
culture (Helling 2006). Teachers employing the minority cultures of their students in their lessons might make cultural references unfamiliar to non-minority readers. Socially-conscious curriculum focuses on social justice throughout all subject areas, including resources used, viewpoints expressed, and assignments given. Socially-conscious curriculum intersects with culturally relevant teaching, but extends beyond to cover issues such as global citizenship, environmental awareness, and poverty reduction (Goodman and Kuzmic 1997). While socially-conscious curricula, or education to create global citizens, would ideally be the norm, it currently represents the ‘marked’ case for the majority of US schools (Banks 2003).

Distinct from forms of pedagogy or curriculum which might be employed by teachers of all backgrounds, ‘high talk’ is a speech register ascribed to African Americans as part of AAL. It involves the use of exaggerated language to discuss accomplishments of self or other (Smitherman 1995). This dramatic form of expression might not be considered appropriate to the genre of professional portfolios, by unininitiated readers.

While these features might also be identified in the writing of non-African American candidates, the panel felt that these, alone or combined with AAL features, might trigger biases in assessors who were less familiar with African American cultural styles of teaching and communicating. The AACI codes will be discussed in a future report.

Speech Code Errors (SCE) represent a category that developed unexpectedly from the panel’s findings. SCE include errors in grammar and spelling. It became apparent that many written commentaries from both African American and European American candidates had examples of SCE. It was decided to investigate whether there is a relationship between increased presence of SCE and lower scores across all candidates, regardless of race/ethnicity.

From the national panel’s findings, an initial coding frame was developed. In order to test the coding frame, the panelists individually applied the codes to a set of portfolio entries. Based on areas of disagreement among the coders, the coding frame was refined and tested again. Table 1 provides a complete list of AAL and SCE features examined in this study.

Identification and training of coders

In order to identify language features across the selected set of written commentaries, which included both African American and European American passing and non-passing candidates, an appropriate group of coders was identified. Masters-level students were chosen who were familiar with linguistics, English grammar, AAL, and African American culture, specifically cultural styles of African American teachers. Six coders self-identified as African American, four as European American, one as Native American, and two as biracial.
Table 1: Summary of SCE and AAL codes

<table>
<thead>
<tr>
<th>SCE codes—sentence level</th>
<th>AAL codes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unnecessary apostrophe in plurals</td>
<td>Regular plural noun, absence of -s</td>
</tr>
<tr>
<td>No apostrophe in possessives</td>
<td>Possessive noun, absence of ‘-s/-s’</td>
</tr>
<tr>
<td>Incorrect apostrophe in possessives</td>
<td>Regular verb, simple past tense, absence of -ed</td>
</tr>
<tr>
<td>Irregular verb, simple past tense, non-EAE form</td>
<td>Regular verb, present/past/future perfect tense, absence of -ed</td>
</tr>
<tr>
<td>Irregular verb, present/past/future perfect tense, non-EAE form</td>
<td>Regular verbal adjective, absence of -ed</td>
</tr>
<tr>
<td>Irregular verbal adjective, non-EAE form</td>
<td>Regular passive voice, absence of -ed</td>
</tr>
<tr>
<td>Irregular passive voice, non-EAE form</td>
<td>Regular 3rd person singular present tense, absence of -s</td>
</tr>
<tr>
<td>Irregular 3rd person singular present tense, non-EAE form</td>
<td>Zero copula</td>
</tr>
<tr>
<td>Conjugated be, non-EAE form</td>
<td>Aspectual or invariant be</td>
</tr>
<tr>
<td>Subject–verb inversion</td>
<td>Deletion of word-final consonant(s)</td>
</tr>
<tr>
<td>Subject–verb discord</td>
<td></td>
</tr>
<tr>
<td>Other non-EAE verb forms</td>
<td></td>
</tr>
<tr>
<td>Other spelling errors</td>
<td></td>
</tr>
<tr>
<td>Missing or incorrect preposition</td>
<td></td>
</tr>
<tr>
<td>Other missing words</td>
<td></td>
</tr>
<tr>
<td>Incorrect use of connecting words</td>
<td></td>
</tr>
<tr>
<td>Noun–pronoun discord</td>
<td></td>
</tr>
<tr>
<td>Pronominal apposition</td>
<td></td>
</tr>
<tr>
<td>Homophones</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SCE codes—discourse level</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypercorrection (with error)</td>
<td></td>
</tr>
<tr>
<td>Hypercorrection (without error)</td>
<td></td>
</tr>
<tr>
<td>Oral language</td>
<td></td>
</tr>
</tbody>
</table>

The training consisted of three days of intensive review of AAL and EAE grammar, syntax, and rhetorical features, with two major goals: (1) creating a shared understanding of the definition of each feature, and (2) developing proficiency in identifying features in written commentaries.

Upon conclusion of the training, coders completed an independent coding of a single set of written commentaries. Within the categories of grammatical and syntactic-level features in both AAL and SCE, the coders demonstrated
strong agreement. For discourse-level SCE codes (hypercorrection and oral language), coders exhibited lower levels of inter-rater reliability. The application of inter-rater reliability rates and triangulation of results across coders will be discussed below.

Assignment of written commentaries to coders

For the development of the coding frame and training of coders, candidate responses from the 1998–99 cohort were used. For the study itself, candidate responses were selected from the 1999–2000 cohort. Entry directions did not change from the 1998–99 to 1999–2000 administration. All categorizations by race/ethnicity were made on the basis of candidate self-identification.

Written commentaries by African American and European American candidates were matched on gender, geographic locale, and scores. The study’s focus was on language usage of African American candidates, but it was also important to have written commentaries from European American candidates to provide a basis for comparison. Each written commentary was randomly assigned to three coders, who were not all African American or all European American.

Triangulation of results across coders

As mentioned earlier, the coders achieved a high level of inter-rater reliability for codes with high frequencies, such as EAE plural nouns with ‘s’. Depending on the specific code, reliability was at least .80. For codes with lower frequencies, such as instances of oral language, inter-rater reliability was lower. The researchers noted that across a given group of three coders, they were able to identify all or nearly all instances of a particular low-frequency feature, but each coder might have missed an instance. In order to add robustness to the coding process, a determination was made to apply triangulation of results across coders (Patton 2002).

Once the three sets of codes were completed for a particular commentary, the data were reviewed for outliers or discrepancies. For high frequency codes, the average of the three coders’ results was taken. For example, the instances of EAE plural nouns with ‘s’ in a particular candidate’s response were coded as occurring 327 times, 340 times, and 337 times, respectively across three coders. Given the high frequency, the average of the three numbers was used: 335.

For codes with low frequency occurrences (less than ten instances in a given commentary), every instance was re-examined by the researchers. For absent ‘ed’ on regular past tense verbs (AAL), one researcher would assemble a list of sentences containing all coded instances of this feature. Each instance would be examined for accuracy. When a researcher identified the mis-application of a code, a correction was made to the data. In order to ensure consistency in this second stage of coding, inter-rater reliability for the two researchers was .90. This resolution process served to reduce instances
of miscoding while still maintaining the integrity of the triangulated data. In no instance did a single coder determine the nature of a particular linguistic feature in the data.

**Data analysis**

Thirty-two written commentaries with three sets of codes were used for this analysis. Of the candidates who completed the written commentaries, 18 self-identified as African American, and 14 as European American. Two African American candidates and one European American were male. Table 2 provides the breakdown by candidates’ race/ethnicity and entry score.

The score scale used for scoring responses to all NBPTS entries ranges from 0.75 to 4.25. A 2.75 represents the lowest level of accomplished performance (NBPTS 2006b). For this study, the score scale was divided in half; low scores range from 0.75 to 2.74, and high scores from 2.75 to 4.25. This division is artificial, and performances with adjacent scores will not be qualitatively different from each other. Figure 1 shows the score distribution for both groups of candidates.

The initial data analysis consisted of a series of cross-tabulations of the frequency of each code. This initial analysis resulted in anomalies that were difficult to explain. In particular, there were multiple instances of European American candidates who used language features considered to be AAL. Evaluating overall patterns, four European American candidates in particular were identified. These four were all teaching in Southeastern states (Mississippi, North Carolina, and South Carolina). In order to understand their language patterns, these four candidates were designated as potential users of Southern White English, which has substantial overlap with AAL. It is important to note that, just as all African Americans do not use AAL, not everyone from the southeastern states is a SWE speaker. There were five other European American candidates from the same states who did not exhibit SWE.

Having identified these four candidates as potential users of SWE, the data were then disaggregated by race/ethnicity, and the category of European

<table>
<thead>
<tr>
<th>Race/Ethnicity</th>
<th>Entry score</th>
<th>Number of candidates</th>
</tr>
</thead>
<tbody>
<tr>
<td>African American</td>
<td>&lt;2.75</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>≥2.75</td>
<td>8</td>
</tr>
<tr>
<td>European American</td>
<td>&lt;2.75</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>≥2.75</td>
<td>5</td>
</tr>
</tbody>
</table>
American was further subdivided to separate users of SWE. Given that only two candidates in the SWE group scored below 2.75 and two scored at or above 2.75, care should be taken not to over-interpret the results for this group.

RESULTS

The primary research question is to investigate whether there is systematic evidence of African American Language in a subset of written commentaries prepared by African American teachers, and the degree to which all candidates exhibit speech code errors. Since the research focuses on whether language features demonstrate patterns by race/ethnicity or entry score, the data are disaggregated by race/ethnicity (African American and European American, with European American further sub-divided by patterns of SWE use) and entry score (divided into less than 2.75 and greater than or equal to 2.75).

In order to illustrate the types of language use that could potentially impact an assessor, examples from candidates’ actual written commentaries are presented, organized into nouns, verbs, lexical items, and discourse features. The key findings are then presented.

Nouns were coded for features typical to speakers of AAL, as well as common SCE features identified by the expert panel. Table 3 presents an example of each code. Plural nouns with ‘s’ absent and possessive nouns with ‘s’/apostrophe absent were coded as possible examples of AAL. Plural nouns with an unnecessary apostrophe and possessive nouns with a missing/incorrect apostrophe were coded as SCE.

Verb features identified by the expert panel as significant were absence of ‘ed’ in past tense verbs, non-Edited American English (non-EAE) forms of past tense irregular verbs, absence of ‘s’ in third-person singular present tense, and subject–verb discords (Table 4). A feature common in
Table 3: Examples of AAL and SCE noun constructions

<table>
<thead>
<tr>
<th>AAL</th>
<th>SCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>noun, regular plural, -s absent</td>
<td>‘I may touch on sentences that go on and on as well as looking at various way_ to end of piece.’</td>
</tr>
<tr>
<td>noun, regular plural, apostrophe added</td>
<td>‘Good writer’s research their topic’</td>
</tr>
<tr>
<td>noun, singular/plural, possessive, -s apostrophe absent</td>
<td>‘On the personal narrative, Nickie___ work demonstrated similar abilities.’</td>
</tr>
<tr>
<td>noun, singular/plural, possessive,</td>
<td>‘students ability . . . students experiences . . . persons life . . . students understanding’</td>
</tr>
<tr>
<td>apostrophe dropped</td>
<td>‘The one girls’ coat got stolen last night.’</td>
</tr>
<tr>
<td>noun, singular/plural, possessive,</td>
<td></td>
</tr>
<tr>
<td>apostrophe not used according</td>
<td></td>
</tr>
<tr>
<td>to EAE rules</td>
<td></td>
</tr>
</tbody>
</table>

Table 4: Examples of AAL and SCE verb constructions

<table>
<thead>
<tr>
<th>AAL</th>
<th>SCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>regular verb, simple past tense, -ed</td>
<td>‘Step two, she design__ her drafting her story (composing). Step three, she used editing to refine what she had written.’</td>
</tr>
<tr>
<td>abs. ed absent (AAL)</td>
<td></td>
</tr>
<tr>
<td>irregular verb, simple past tense,</td>
<td>‘Yesterday the students run out of the classroom, despite the teacher’s request not to.’</td>
</tr>
<tr>
<td>non-EAE (SCE)</td>
<td></td>
</tr>
<tr>
<td>regular verb, perfect tense, -ed</td>
<td>‘We have revise_ the models we wrote together...’</td>
</tr>
<tr>
<td>abs. ed absent (AAL)</td>
<td></td>
</tr>
<tr>
<td>irregular verb, perfect tense,</td>
<td>‘I wish Brian could have showed supporting details...’</td>
</tr>
<tr>
<td>non-EAE (SCE)</td>
<td></td>
</tr>
<tr>
<td>regular verbal adjective, -ed</td>
<td>‘When David is asked to add more detail, he gets discourage__ because he does not know what to write...’</td>
</tr>
<tr>
<td>abs. ed absent (AAL)</td>
<td></td>
</tr>
<tr>
<td>irregular verbal adjective,</td>
<td>‘I asked Matt if I could see the re-wrote paper from yesterday.’</td>
</tr>
<tr>
<td>non-EAE (SCE)</td>
<td></td>
</tr>
<tr>
<td>regular verb, passive voice, -ed</td>
<td>‘The students are expect__ to articulate the message that they are trying to relay...’</td>
</tr>
<tr>
<td>abs. ed absent (AAL)</td>
<td></td>
</tr>
<tr>
<td>irregular verb, passive voice,</td>
<td>‘The paper was wrote by a pair of students.’</td>
</tr>
<tr>
<td>non-EAE (SCE)</td>
<td></td>
</tr>
<tr>
<td>regular verb, present tense, 3rd person</td>
<td>‘Administration usually keep_ the second graders with the same teacher for third grade.’</td>
</tr>
<tr>
<td>singular, -s absent (AAL)</td>
<td></td>
</tr>
<tr>
<td>irregular verb, present tense, 3rd</td>
<td>‘He don’t understand how to do this problem.’</td>
</tr>
<tr>
<td>person singular, -s absent (AAL)</td>
<td></td>
</tr>
<tr>
<td>Other non-EAE verb forms, excluding</td>
<td>‘I realize he is a writer that has being responsible for his own learning.’</td>
</tr>
<tr>
<td>S-V concord (SCE)</td>
<td></td>
</tr>
</tbody>
</table>
spoken AAL is indicating past tense for regular verbs without the EAE ‘ed’. This was examined across simple past tense, perfect tense, verbal adjectives, and passive voice. The literature on spoken AAL does not identify a specific pattern for irregular verbs (Rickford and Rickford 2000). For this analysis, irregular verbs in non-EAE form are designated as speech code errors. The subject–verb discord marked the absence of ‘s’ on 3rd person singular verbs in present tense as being a potential indicator of AAL when the subject and verb were separated by no more than 1–2 words. Lengthy separation between subject and verb could lead to a lack of noticing a subject–verb discord; these were counted as SCE.

A series of codes were gathered under the composite heading of lexical features: word-final consonant deletion; other spelling errors; missing or incorrect prepositions; other missing words; and incorrect use of connecting words. Table 5 presents examples of these.

The expert panel identified three discourse-level linguistic features. Hypercorrection (error producing) occurs when the writer over-applies a grammatical rule where it is not required, leading to an error. Hypercorrection (without errors) occurs when a writer is overly prescriptive with grammatical rules, but is technically accurate, resulting in disfluency. The final feature is identified as ‘oral language’. This occurs when a sentence or phrase is not grammatically correct in formal writing, and yet if read aloud, the meaning is completely clear. Smitherman (1994) refers to this as ‘conversational tone’. Many instances of oral language could be classified as the use of colloquialisms, independent clauses not joined according to EAE rules, or dialogue without punctuation. Table 6 presents examples of these.

<table>
<thead>
<tr>
<th>Table 5: Examples of AAL and SCE lexical features</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Word-final consonant deletion (AAL)</strong>, e.g., ‘an(d)’</td>
</tr>
<tr>
<td><strong>Spelling errors (SCE)</strong></td>
</tr>
<tr>
<td><strong>Missing/incorrect preposition (SCE)</strong></td>
</tr>
<tr>
<td><strong>Missing words (SCE)</strong></td>
</tr>
<tr>
<td><strong>Incorrect connectors (SCE)</strong>, e.g., yet, but, and, so, etc.</td>
</tr>
</tbody>
</table>
Key findings

Looking across multiple coding categories, it is evident that there is a clustering of codes for African American candidates. The frequency at which African American candidates use both AAL and SCE is greater than for European American candidates. It is important to note that clustering is not associated with high or low scores. Specific findings were as follows:

- In keeping with the literature (Smitherman 2001), usage of AAL in formal, written, academic prose was relatively low.
- Evidence of AAL appeared consistently across both high-scoring and low-scoring African American candidates.
- The four European American candidates identified as not systematically using Edited American English showed linguistic patterns similar to AAL, hypothesized to be Southern White English.
- Evidence for Speech Code Errors was low across all groups, as can be expected given the formal academic assessment.
- African American candidates demonstrated greater evidence of SCE than European Americans who employed Edited American English.

These findings are tempered by limitations of small sample size and the challenges of working with human coders. In the case of the latter, multiple checks and balances were used to minimize the possibility of missed codes. These included background screening of coders, training and assessment, monitoring and feedback during the coding process, resolution of discrepancies through a standardized process, and triangulation of final results.

Results by language category

Table 7 (available online to subscribers at http://applij.oxfordjournals.org/) presents the results in summary form, organized into nouns, verbs, and other lexical features. While many codes had low frequencies, when combined,
they contribute to the overall impression given to a reader. Several of these codes, specifically noun-pronoun discord, pronominal apposition, homophones, non-EAE forms of copula, and subject–verb inversion were not discussed individually above, due to their relative scarcity in the data pool, but they are included in Table 7.

Although there are only two candidates classified as users of SWE who scored less than 2.75 and two who scored greater than or equal to 2.75, their data are included in Table 7. These four were grouped separately to make sense of anomalies in the data collected from European American candidates. However, comparisons between users of SWE and other candidates are limited, given the small sample sizes.

The use of AAL noun features clustered among African American candidates, both high and low scoring, and also among European American users of SWE. However, high-scoring European American candidates who used Edited American English showed no evidence of AAL features. Specific findings include the following: when indicating plurality of nouns, absent ‘s’ (AAL) was found among more than half of African American candidates, both high- and low-scoring. When indicating possession, however, absent apostrophe/’s’ (AAL) was rarely seen—used by only one African American candidate in each group (high- and low-scoring). Speech Code Errors of pluralization with unnecessary apostrophe and possession with missing/incorrect apostrophe were seen among African American candidates, European American candidates who used SWE, and low-scoring European American candidates who used Edited American English. In only one instance did a high-scoring European American candidate demonstrate SCE in nouns.

Speech Code Errors conform to neither the grammatical structure of EAE, nor that of AAL. As anticipated, both African American and European American candidates exhibited SCE. However, the rate at which European American candidates display these features is lower than for African American candidates. Combining all SCE and AAL noun codes shows some clustering among African American candidates, in particular among those candidates with high scores.

It is interesting to note that for African American candidates, the use of AAL verb forms was much less frequent than the use of AAL noun forms. While there were a few instances of European American candidates exhibiting noun features classified as AAL, none of this group used any AAL verb forms.

When indicating past tense, low-scoring African American candidates exhibited more non-EAE verb usage than high-scoring African American candidates—one-half of low-scoring African American candidates versus one-fourth of high-scoring candidates. Absent ‘s’ in AAL verb conjugation was found predominantly among African American candidates—seven of ten low-scoring candidates and three of eight high-scoring candidates. No European American candidates who used EAE showed the absent ‘s’ feature.
In reviewing overall use of AAL verb forms, instances were few, but consistent clustering patterns were seen among African American candidates. Both African American and European American candidates displayed verb forms categorized as SCE.

The incorrect use of connecting words is found predominantly in the African American candidates’ writing, and missing/incorrect prepositions also occurred more frequently among African American candidates. Two candidates with the greatest number of oral language instances were ranked first and third for greatest number of grammatical errors—both high-scoring African American candidates.

Overall there were consistent patterns of non-EAE usage by the majority of African American candidates—both high- and low-scoring—and a small group of low-scoring European American candidates who used predominantly Edited American English.

DISCUSSION AND CONCLUSIONS

This study contributes further insight into research on African American Language, by focusing on written usage, rather than spoken. It specifically examines the presence of AAL in the context of a formal, professional development activity for teachers—a portfolio entry in NBPTS assessments. This research contributes to the continuing debate over what factors lead to the significant differential performance seen between African American and European American candidates in all forms of assessments. The study lays the foundation for examining the relationship between AAL and differential performance outcomes; this is done by providing evidence for the existence of AAL in a context where it has the potential to affect assessors’ judgments, in the form of construct-irrelevant variance in the portfolio scoring process. Furthermore, this study offers additional levels of complexity to consider in research on AAL: (1) the possibility of non-African American speakers/writers using AAL-forms in their language production, including in professional contexts requiring Edited American English, and (2) the presence and possible effects of Speech Code Errors, which occur in teachers’ writing regardless of race/ethnicity.

The authors seek to contribute to the ongoing discussion on standardized assessments, which adversely impact individuals of particular racial, ethnic, and language backgrounds. A close reading of the NBPTS Scoring Handbook shows that the focus in scoring remains consistently on the candidate’s provision of appropriate evidence for teaching quality:

Especially if your score is at the lower end…did you submit a Written Commentary that was only two or three pages long…. You would not have been penalized automatically for not submitting all the student work or submitting a short Written
Commentary, but submitting very little material makes it much less likely that you will include evidence of meeting the standards assessed by the entry.... Did you provide evidence that you were able to reflect on your teaching and determine how you could modify your teaching of the instructional sequence in the future? (NBPTS 2006b: 8)

Writers are judged on the teaching evidence presented and not on the form of presentation. This paper lays the groundwork for future studies in the use of ‘non-prestige language varieties’ by identifying the presence of AAL in this particular writing context (Kasper 2006, personal communication). More will be said about possible future studies below.

This study examined a small sample of portfolio entries written by African American and European American candidates. Evidence was found for limited use of AAL features by a subgroup of African American candidates, which aligns with Smitherman’s research and others: up to 90 percent of African Americans use aspects of AAL some of the time. A pattern of features similar to AAL was also found among a subgroup of European American candidates, who identified themselves as being from the southeastern USA; it was postulated that this might be evidence of Southern White English. The European American candidates who used American English (as opposed to SWE) tended not to have employed any AAL features. All candidates, to varying degrees, demonstrated evidence of SCE. High scoring European American candidates who used American English had the fewest SCE.

Evidence for the use of AAL was found among African American candidates at all score levels, which lends support to the effectiveness of NBPTS bias-reduction exercises. That is, the existence of AAL features at all score levels suggests that assessors may have experienced some success in screening out personal prejudices against this writing form. The same could be said for the evidence of SCE found at all score levels. The assessor training explicitly addresses the need to identify and monitor writing biases, and not to factor writing style into score judgments.

This study did not consider assessors’ race or ethnicity, nor were judgments made about the quality of candidates’ work. Attempts were not made to identify whether there were responses awarded low scores in spite of high quality work or responses awarded high scores in spite of deficiencies. If AAL usage had been concentrated in low-scoring responses, then their content would have been examined as well.

However, research with English as a second language and various American dialects, and some limited studies on AAL, indicate the potential for negative evaluations by readers of texts containing non-EAE features. Given that both high- and low-scoring African American responses demonstrated evidence of AAL, it is reasonable to ask whether the presence of AAL has a small negative impact on assessors’ judgments across the full spectrum of the score range, causing a possible dampening effect at
all score levels. This might also hold true for the presence of SCE, which has been shown to cause reader irritation.

For the present cohort of candidates, an approximately half-point difference is observed in the mean scores between African American and European American groups. This is consistent with results from previous cohorts. It is possible that some of this differential performance could be due to the construct-irrelevant effect of the writing features used.

A follow-up study has been proposed to explore this possibility. Selected commentaries which display AAL features could be ‘re-written’ to employ only EAE. Similarly, selected commentaries written in EAE could be re-written with the help of linguists, to employ some AAL features. These modified responses could be re-scored by trained assessors, who would not be aware of the changes made. Scores from the altered responses could then be compared to the original scores.

A variation on this study would be to select a single commentary to ‘re-write’ according to specific parameters. These might include the placement of AAL features in specific locations (earlier in the document versus later), or varying the frequencies and/or types of specific AAL features. These altered versions could then be scored by trained assessors. Applying the NBPTS scoring rubric, there should be no significant variation in the scores assigned across different versions.

Additional questions stemming from this study include why the occurrence of SCE is less frequent for the European American candidates than for African American candidates. One possible hypothesis is that greater training and experience in formal academic writing for European American candidates could be a contributing factor. It has been hypothesized that African American (and Latino) children in the USA are more likely to receive a less adequate K–12 education, given that they are more likely to attend under-funded urban and rural schools and to be taught by teachers working outside their certification area (NCES 1999; Ascher and Branch-Smith 2005).

To explore the hypothesis that African American candidates who used more non-EAE features had less experience in formal academic writing, a reasonable correlate was sought in the limited background data available. Candidates provide information on their highest degree earned. An assumption could be made that candidates with a Bachelors degree would have less formal, academic writing experience than candidates with a Masters or beyond. However, when the data were examined, 12 of 18 African American candidates had a Masters or higher, compared to 7 of 14 European American candidates. For the African American candidates who had high scores on this particular entry, and yet who had significant numbers of SCE, 7 of 8 had Masters.

Looking at the highest degree earned, therefore, as a measure of formal writing experience does not explain the data. Another avenue for exploration could involve the effect of learning two languages (American English and AAL or SWE), mediated by access to high quality language education and
consistent opportunities for practice in both target languages. More background data would need to be collected from candidates to pursue this. There may also be an interaction between audience/purpose and the presence of non-EAE features. The pressure of preparing the portfolio may result in increased language errors and decreased attention to error elimination. The process of language production becomes even more complex for bidialectal speakers. A further avenue for study might be to examine candidates’ production of academic writing under less pressured circumstances.

A final consideration in understanding the results involves demands on candidates’ time versus formal support for preparing the portfolio. In other words, those candidates who have access to district and school support, mentors, or other NBPTS certified teachers may perform better than those who do not have such support. The limited background data available in this study did not allow for exploration of this hypothesis.

Differential performance for African American test takers has been documented across all assessment types for several decades. ‘There are urgent needs in this country for a systematic cultural-linguistic review of testing and assessment devices to be used with African-American[s]’ (Hilliard 2002). While the differential performance gap is slowly decreasing, more work must be done to uncover sources of this difference and to reduce their effects. This includes research into language use, cultural dissonance between candidates and assessors, educational preparedness of candidates, access to preparation courses and mentors, types of schools and student bodies where teachers practice, and the stereotype threat phenomenon.

Final version received July 2006

SUPPLEMENTARY DATA
Supplementary data mentioned in the text is available online to subscribers at http://applij.oxfordjournals.org/.

DISCLOSURE
The data in this research was provided by NBPTS. All conclusions are those of the authors, and no endorsement by NBPTS should be implied.

ACKNOWLEDGEMENTS
The authors wish to extend thanks to the members of the national expert panel, including Dr Carolyn Adger and Dr Judith Diamondstone, who guided the development of the coding frame and reviewed early drafts of this report.
REFERENCES


REFERENCES

Bennett, C., L. McWhorter and Ball, A. Ball, A. Ball, A. Ball, A. and L. Odell (eds): Evaluating Writing. Urbana, IL: NCTE.


Irvin e, J. and J. Fraser. 1998. ‘Warm demanders,’ Education Week, 13 May.


