View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by University of Bath Research Portal

Citation for published version:
Terry, JM, Hendrick, R, Evangelou, E & Smith, RL 2010, 'Variable dialect switching among African American children: inferences about working memory', Lingua, vol. 120, no. 10, pp. 2463-2475. https://doi.org/10.1016/j.lingua.2010.04.013

DOI:
10.1016/j.lingua.2010.04.013

Publication date:
2010

Document Version
Early version, also known as pre-print

Link to publication

University of Bath

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 12. May. 2019

Variable Dialect Switching Among African American Children: Inferences about Working Memory☆

J. M. Terry*, R. Hendrick*, E. Evangelou**, R. L. Smith**

University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA

☆This paper could not have been written without the generosity of the members of the AAE working group at the UNC Frank Porter Graham Child Development Center - most especially Joanne Roberts and Susan Zeisel - who shared their data and insight. Likewise, it would not be the same paper without the thoughtful advice we received from Lynn Frazier, and William Labov, both of whom commented on earlier drafts. Of course any errors are the authors' responsibility alone.

*Department of Linguistics

**Department of Statistics and Operational Research

Variable Dialect Switching Among African American Children: Inferences about Working Memory☆

Abstract

This paper presents evidence that dialect switching can pose a variable cognitive load that modulates success in verbally mediated tasks. A Bayesian Markov Chain Monte Carlo model is used to explore and confirm the hypothesis that the morphosyntactic organization of African American English (AAE) has significant, variable effects on second grade African American students' performance on mathematical reasoning tests conducted orally in Mainstream American English (MAE). These effects correlate with students' productions of AAE. Neither measures of spatial reasoning nor span measures of children's working memory correlated with this aspect of test performance, but certain types of representational mismatches did. These findings are consistent with other work suggesting that mathematical reasoning and language draw from a common working memory store, and that processing difficulties are linked to manipulating representations rather than limits on storage capacity.

Key words: working memory, African American English, dialect switching, Markov Chain Monte Carlo model

1. Introduction

This paper follows from the conjecture that bilingual and bi-dialectal speakers, to the extent that they are required to switch between linguistic codes in verbally mediated tasks, have an added cognitive load that can have variable and measurable effects.

Guided by this expectation, we test the hypothesis that the morphosyntactic organization of African American English (AAE), where it contrasts with Mainstream American English (MAE), has significant effects on second grade AAE speaking students' performance on mathematical reasoning tasks. Morphosyntactic organization is meant here in its broadest sense so

that having different morphemic inventories constitutes a different organization. This hypothesis draws superficial plausibility from the fact that high levels of AAE use have been shown to correlate with poor academic performance, with those students who can successfully switch between AAE and MAE performing better than those who cannot; see Craig and Washington (2006).[1] It is also broadly consistent with recent work by Fedorenko et al. (2007) showing that linguistic complexity interacts with mathematical tasks like addition, but not visuo-spatial tasks like image rotation. Our hypothesis leads us to expect that AAE speaking students who are better able to manipulate the mismatches between the linguistic representations of MAE and AAE will do better on verbally mediated mathematical reasoning tasks than those who cannot.

Three predictions emerge from this general hypothesis. First, because they involve representational differences in how sound and meaning are paired, morphemic mismatches will be the most likely locus for observing the effects of such mismatches. These mismatches may involve distinct semantic values of the two morphemes in question, but they need not. Mismatches can arise between an overt morpheme and a morpheme that is phonologically or semantically null as well. Any formal mismatch has the potential to pose a cognitive load according to our general hypothesis. It is an empirical question whether the hypothesis is true in its broadest form or whether it will need to be narrowed in light of empirical findings about distinct sub-classes of mismatches. We expect any depression in scores tracking morphological divergences to interact with problem difficulty, so that the harder the mathematical problem, the more the effects of linguistic mismatches will matter to the students overall test score. Second, we predict a cumulative effect of such mismatches. That is, the more mismatches there are in a mathematical word problem, the worse we expect students to perform. Third, we expect to find effects of the representation of morphemes. Mismatches that involve one to one relations between morphemes in the two linguistic codes will be easier to manipulate than other relations that require substantial cognitive work to infer and maintain.

The organization of this paper is as follows. In section 2 we present the

[1]This hypothesis is not meant to exclude other factors, such as socio-economic status, from also contributing to the complex problem of poor academic performance. The goal of our hypothesis is simply to outline the specific mechanics of one contributing factor.

data set that we will use to explore our hypothesis and a Markov Chain Monte Carlo (MCMC) model that we will use to evaluate the above predictions. Our primary goal in this section is to establish that specific morphological features of MAE (those that diverge from AAE) have a negative effect on mathematical reasoning in African American students who produce higher amounts of AAE. In essence we find that the features representing these divergences constitute a cognitive load that serves to depress performance on the verbally mediated mathematical reasoning tasks in which those features occur. In section 3 we pursue the hypothesis that that the principal effect isolated in section 2 stems from a cognitive load incurred when switching between representations of some morphemes in mainstream English with their corresponding representations in AAE. Section 4 summarizes our argument and offers ideas about how our hypothesis could be tested further.

2. Mathematical Reasoning and Linguistic Ability Interact

A dialect of English spoken by many, but not all, African Americans, AAE has proven to be a rule-governed linguistic system whose phonology, syntax and semantics are related to, yet in many ways significantly different from, those of more standard varieties of English; see Labov (1972); Wolfram (1974); Wolfram and Fasold (1974); Mufwene et al. (1998); Craig and Washington (2004); Green (2002).[2] We believe that one reason for the poor performance of AAE speaking school children relative to their MAE speaking peers is that, when instructed and tested in mainstream English, AAE speaking children bear the burden of keeping separate these and many other structures in which the same or very similar grammatical pieces are used with different meanings. From this perspective, the management of the mismatch constitutes a processing load that speakers of AAE incur during MAE language comprehension tasks.

In an attempt to test whether complex linguistic performance was insulated from other cognitive tasks we undertook a study of how young African

[2]Take, for example, the AAE sentence *Mary be studying.* Many mainstream English speakers erroneously assume that this sentence means Mary is studying when, in fact, it means something closer to Mary always studies or Mary studies all the time Green (2000). It is impossible to express the AAE meaning in mainstream English without the use of an adverb. One confusion that the mainstream English speaker faces interpreting this example is that the form of *Mary be studying* is so similar to the form of *Mary is studying* that it is easy to mistake the one for a degenerate form of the other.

American children performed on verbally mediated mathematical reasoning tasks. The group of children we studied exhibited variation in the degree to which they participated in the dialect of AAE. Mathematical reasoning tasks presented in MAE thus presented students with a complex linguistic task of variable degree as they managed the mismatch between AAE and MAE. We hypothesized that such dialect switching would have an effect on students' performance on the mathematical reasoning tasks.

2.1. Performance Data on Woodcock-Johnson-R Applied Problems Subtest

Eighty seven African American students were recruited from North Carolina community based child care centers to participate in a longitudinal study of children's health and development Roberts et al. (1995).[3] As a part of this study, at regular intervals, language samples were taken from the students, and they were administered a series of diagnostic test to assess their linguistic and other cognitive abilities. Of central importance to this paper, these tests included both the Calculation and Applied Problems sub tests from the Woodcock-Johnson-R (WJ-R) Psychoeducational Battery. Applied Problems is a subtest that assesses skill in analyzing and solving verbal math problems - the familiar 'word problems' - as distinct from the Calculation subtest, which tests accuracy of calculation. For example, *If you have seven pennies and you spend three of them, how many pennies would you have left?* is a typical Applied Problems question. We cannot reproduce the entire list of test questions for copyright reasons. In addition to the Applied Problems and Calculation subtest of the WJ-R, other diagnostic tests included the WJ-R Letter-Word Identification subtest, which assesses reading identification sills in identifying isolated letters in words (e.g. *What is the name of this letter?* while pointing to the letter "O"), and the Clinical Evaluation of Language Fundamentals (CELF-3), an instrument designed to measure overall receptive and expressive language ability. In addition to these tests narrative language samples were collected from each of the study participants. The administrator facilitated three narrative situations: telling a story about bears after viewing slides on a personal slide viewer, describing a picture of a circus, and responding to prompts about situations such as telling about having lost

[3]In subsequent work we attempt to leverage the longitudinal character of this data to assess the trajectory of the effects of the linguistic properties we document here for the same children at subsequent points of development and to address issues relating to strength of representations that are raised at the conclusion of this paper.

a tooth. Each was transcribed using Systematic Analysis of Language Transcripts (SALT) from Miller and Chapman (2000). Study participants were also screened to identify any hearing loss. All tests were administered by 1 of 7 trained examiners with expertise in speech and language assessment. The tests were given at the Frank Porter Graham Child Development Center, a university research facility.

The WJ-R sub tests were given orally at kindergarten entry and at the end of each grade year beginning with first grade. The data analyzed in the current paper include the individual responses of 75 of the original study participants at second grade. The mean age of the students at the time of testing was 8.32 years. Standard scores, called W scores, were calculated from the students' results. W scores are based on the Rasch ability scale and are centered on a value of 500, which is the approximate average performance of a beginning fifth grader. The 60 questions on the Applied Problems subtest increase in difficulty and are divided into pages. Students were 'ceiling tested' by complete pages until the six highest-numbered items were failed, or until the last test item was answered.

In addition to data concerning individual students' oral performance on each test question, three members of our team coded each test question for a range of linguistic properties. Reliability between coders was established over the last ten questions, as linguistically speaking, the last questions are the most complex, and therefore most likely to reveal any coding inconsistencies. Subsequently six morphological features were chosen for further statistical analysis. The first five, past tense *-ed*, the past participle *-en*, the past tense copula *was/were*, the auxiliary *have*, and third person singular *-s*, were all chosen because they have been identified as points of divergence between AAE and MAE; see Green (2002),Craig and Washington (2006). The final feature, the counter factual conditional *if + -ed*, was selected as a point of reference because of its importance to reasoning tasks and because, all indications are that it behaves the same in AAE as it does in MAE (although there is the possibility of interaction with *-ed*). All six features are listed along with MAE examples in Table 1.

This data set differs from the kind of data gathered in response to controlled experimental stimuli in psycholinguistic laboratories that are artificially constructed to test a hypothesis. We did not construct the stimuli of our data set, nor did we control or make detailed observations of all aspects of the testing environment. The primary disadvantage of our data set is that we have no direct observation of the time course of comprehension with which

to test our main hypothesis. Our hypothesis makes testable claims in this domain, but our data set does not shed light on such questions. The primary advantage of our data set is that we are able to test whether the hypothesized cognitive load has any significance in important real world tasks as opposed to artificial tasks in laboratory settings. In most experimental psycholinguistic experiments, the inference from results in laboratory settings to real world tasks is left unexplored. In this respect, our data set resembles more closely the kind used in climate research than the kind generated in controlled experiments on, for example, fluid dynamics. Both areas of research explore broadly similar natural processes, and each has strengths and weakness, depending on one's purposes. Certainly, both can be conducted with typical scientific standards, and we will proceed here to try to understand our data set and the mechanisms that could plausibly have produced it.

To test our first prediction - that morphemic mismatches between MAE and AAE have an effect on student performance - we seek, then, to determine whether the performance of these African American children on the WJ-R Applied Problems subtest correlates with those mismatches between mainstream English and AAE that are represented by the linguistic features in Table 1. Simply being African American, however, does not guarantee that one is a speaker of AAE. Further, there is substantial variation in the use of AAE features among AAE speakers. To establish the AAE speaking status of each student and to measure the variation in students' use of AAE, we calculated dialect density measures (DDMs) from unscripted narratives produced by the students to their mothers. This measurement used the inventory of AAE features given in Craig and Washington (2004). It expresses the rate of dialect feature production calculated as a ratio of number of dialect features to number of words Craig and Washington (2004). The mean DDM score for the students in the study is 0.168 with a standard deviation of 0.124. In interpreting these numbers it is important to recognize that the vast majority of AAE speech overlaps with MAE and other varieties of English.

2.2. An MCMC Model of the Linguistic Effect

Our goal in this subsection is to determine whether the features in Table 1 influence students' performance on the Applied Problems subtest. To this end, we model whether or not a student answers a given problem correctly as a function of that student's general mathematical ability, the level of difficulty of the problem itself, and the presence or non-presence of any of

the linguistic features in Table 1. So that we can evaluate our second prediction, that the linguistic effects we expect to find will be cumulative, our model sums the individual instances of each feature. We employ a Bayesian Markov Chain Monte Carlo Method (MCMC) to estimate the unknown parameters, including the effect of the features on a student. Details of the model, including a measure of its fit to the data, are given in the Appendix. Basically our approach is to treat the model parameters (the student's general mathematical ability, the overall difficulty of the question, and the extent to which the student is affected by a linguistics feature) as random effects. The Bayesian-MCMC approach to fitting the model is to define prior distributions for these parameters and use Gibbs and Metropolis sampling to construct posterior distribution for all the unknowns; see Young and Smith (2005) or Baayen (2008). The posterior distributions are then used to determine which, if any, linguistic features influence students' scores. If one treats the influence of the linguistic features as fixed rather than random effects, (i.e. if one assumes the influence of a particular linguistic feature is the same for all students), it is possible to estimate the model by standard logistic regression. However, not only is the number of unknown parameters too large to assume fixed effects, but doing so prevents testing whether the amount of AAE that a student uses correlates with the effect that some AAE features have on test performance, an important facet of our hypothesis that is pursued here.

Because they involve differences in how sound and meaning are paired, we hypothesized that morphemic mismatches between MAE and AAE, would affect student performance in learning and testing situations. For each linguistic feature we examined, Table 2 provides the correlation between a student's total score on the Woodcock Johnson Applied Problems subtest, and $\alpha^M_{i,k}$, the measure of the influence of that feature on the student provided by the model. If there is no correlation between a student's score and the linguistic feature, we expect the correlation to be 0. A high positive or negative value indicates that the affect of the feature is high. In the case of a positive correlation, students who are to a great degree negatively affected by the linguistic feature in question have worse than average scores, while students who show a high positive effect, tend to have strong scores. The reverse is true in the case of a negative correlation. Table 2 also provides the p-value for the null hypothesis that there is no correlation between a student's score and the linguistic feature; the lower and upper bound for the 95% confidence interval for the correlation is also reported.

The data in Table 2 suggest that the linguistic features we examined do influence students' overall scores. Further, as it is built into the model, the model's fit suggests that the effects are cumulative. Of all the features examined, third person singular -s appears to have the greatest effect, while the past tense copula and counterfactual conditional appear to have the least effect.

Similar to Table 2, Table 3 shows, for each linguistic feature, the value of $\alpha_{i,k}^{M}$, its standard deviation and coefficient of variation.

While Table 2 shows that third person singular -s has the greatest effect on students' overall scores, in Table 3 we are able to see that it also exhibits a high degree of variation. This tells us that there are some students who are highly affected by the presence of this particular feature.[4]. This finding is consistent with the important work of deVilliers and Johnson (2007) which showed that AAE speaking children did not comprehend third singular -s and that this insensitivity was independent of the general language development in the children. It also is broadly consistent with accuracy and eye tracking of listening comprehension reported in Beyer and Kam (2009). Our results complement this earlier work by showing that the insensitivity observed in these studies may in fact pose a cognitive load that depresses performance in mathematical reasoning tests.

2.3. Correlation with Variation in AAE Productions

Our hypothesis the dialect switching between AAE and MAE incurs a cognitive load predicts that African American children who can be strongly identified as AAE speakers should on average show more of an effect of the mismatches between the dialects than African American children who can be weakly identified as AAE speakers. As an approximate measure of how strongly we should identify children as AAE speakers we measured the amount of AAE usage (coded in terms of Craig and Washington's DDM) in spontaneous narrative produced by children in a task that involved spontaneous narratives, as described in section 2.1. We introduced this measure as an additional parameter in our MCMC model. The result of including this variable of AAE usage is presented in Table 2.3.

When we take into account students AAE productions the correlation of students' performance with the present 3rd person singular -s increases

[4]Our model allows us to estimate the size of the effect on average students, a topic we take up in Terry et al. (2009)

strongly. This increased correlation is consistent with our hypothesis that AAE speakers should exhibit a greater effect of the linguistic features.

The DDM is a broad measure of AAE usage that includes phonological, morphological and syntactic features. As a more stringent test of our hypothesis, we could explore that relation between children's spontaneous productions of third singular *-s* and the effect of AAE features on test performance. However, some children produced narratives that contained did not have any such observations, unlike the DDM that was calculated for each student. The story telling task that children performed yielded some narratives that contain no occurrences of third person singular *-s* in MAE while others had a high number of such occurrences. We would be interested in the proportion of third person singular *-s* that children spontaneously produced that would occur in MAE, recognizing that there may be no value for some children's narratives. If we evaluate our model with the more limited data on children's' spontaneous production of third person singular *-s*, the results are consistent with the evaluation using the DDM measure./footnoteIn this calculation, children who never produced third person singular *-s* when it would be used in MAE were scored as 0.00, as distinguished from children who showed no evidence of third person singular *-s* but did not produce utterances that would have third person singular *-s* in MAE. This later group were assigned .50. The result of including this variable of AAE usage is presented in Table 2.3.

2.4. Correlation with Spatial Reasoning

It is possible that the correlation between performance on test questions and properties of AAE is driven somehow by the students' general intelligence rather than the organization or the use of their linguistic knowledge. Although its existence is debated, the case for general intelligence is based largely on the fact that there is a correlation between scores on tests of demonstrably distinct domains cognition. For example, while a great deal of evidence shows that verbal memory (and ability) is distinct from spatial memory (and ability), research shows a correlation between subjects' scores on tests of the two; see Wake et al. (1996). If general intelligence or access to general working memory is the hidden driver behind the linguistic effects and groupings that we have identified, we expect that introducing a measure of spatial reasoning ability - an ability distinct from linguistic abilities but presumably driven by general intelligence - into our MCMC algorithm should lessen the apparent effect of the linguistic features. Using the Block

Design subtest from the third edition of the Wechsler Intelligence Scale for Children, as such a measure, we reran our MCMC algorithm. This subtest involves copying small geometric designs with four or nine larger plastic cubes. As can be seen in Table 6, our rerun of the algorithm showed no significant difference in the $\alpha_{i,k}^{M}$ and students' total test performance from the results of our algorithm that ignored spatial reasoning abilities. The independence of spatial reasoning abilities from the linguistic effects indicates that those linguistic effects are driven by more linguistically specific factors than general intelligence.

2.5. Correlation with Variation in Working Memory

If one conceptualized working memory as a mental resource with a limited capacity that was expended as tasks of increasing complexity were undertaken, our hypothesis that dialect switching between AAE and MAE poses a cognitive load would lead one to look for a correlation between performance on test questions that required dialect switching and independent measures of children's working memory. To evaluate this aspect of our hypothesis we introduced a measure of working memory as a variable in our MCMC model. The measure we used was success in a word list repetition task performed during a concomitant sentence comprehension task. This is the same span test developed by Daneman and Carpenter (1980) to measure the capacity of working memory.

As can be seen in Table 8, we do not observe any increase in correlation between $\alpha_{i,k}^{M}$ and test performance when the span of the word list task is used as a measure of working memory capacity. The ability of our model to correlate performance on test questions with, for example, third person singular *-s*, is not improved by adding information about students' performance on the memory load task used by Daneman and Carpenter (1980) to measure the capacity of working memory.

2.6. Discussion

It has been widely believed since Miller and Chomsky (1963) that memory interacts in important ways with linguistic information in many kinds of verbal tasks, and that, following Baddeley (1986), there is some independence of how verbal processing and visuo-spatial processing tap working memory. Beyond this common view, there is considerable debate about the mechanisms that underlie and explain such interactions. Baddeley suggested that working memory was composed of a central executive responsible for

general cognitive tasks and two slave systems: a phonological loop and a visuo-spatial sketchpad. Subsequent work on sentence processing divides as to whether such processing is a central executive function or if there is a dedicated subsystem, like the phonological loop, that carries it out. This debate takes on more general interest in the cognitive sciences when it is couched in terms of whether working memory is modularized (in the sense of Fodor (1983)) for the purposes of language processing.

One side of this debate theorizes that sentence processing is conducted in working memory as a homogeneous resource pool for cognitive processing generally (Baddeley's central executive) and that this resource has capacity limitations (in the spirit of Miller (1956)). A specific instantiation of this tack holds that activation drives both cognitive operations that are employed in language comprehension and the maintenance of cognitive representations of information; there is thus a trade off between processing and storage because they compete for the same resource. This view leads to the hypothesis that individual differences between experimental subjects in their working memory capacity correlate with performance on processing complex sentences (see for example Just and Carpenter (1992); King and Just (1991)). Subjects' performance on tasks involving complex syntactic structure (subject vs object extracted relative clauses) are claimed to correlate with an independent classification of subjects as high or low span readers Daneman and Carpenter (1980).

1. `The cat that ___ chased the squirrel climbed a tree.`
2. `The squirrel that the cat chased ___ climbed a tree.`

The more complex syntactic structure (an object extracted relative clause like 2 exhausts the limited capacity of working memory more quickly for low span readers than the less complex structure (a subject extracted relative clause like 1).

A second side in the debate conceptualizes working memory as containing a module dedicated to sentence processing separate from the central executive used for general cognitive tasks Waters and Caplan (2001). This sentence processing module is automatic (or reflexive) and, because it is modularized, is insulated from interactions with other cognitive tasks. Proponents of this view make use of dual task experiments to probe the relative independence of different types of tasks involving working memory as evidence for the automatic, modularized character of sentence processing. Processing complex sentences like 1 and 2 are predicted to be uninfluenced by a concurrent

memory load such as a span of digits or list of words, which is held to be a consciously controlled mechanism. These studies suggest that, while working memory may have a limited capacity that can be exceeded by complex tasks drawing on general, conscious functions of working memory, processing of complex sentences is done by a separate, dedicated subsystem that is insulated from such effects. They contest reports in other studies claiming to have found effects of length of digit or word lists and sentence complexity (Waters and Caplan).

A third position in the debate holds that working memory is not a pooled resource of limited capacity, but that differences on experimental tasks are the result of the processing nature of the tasks involved and the characteristics of information representations manipulated. This position is motivated by experimental findings that an interaction of concurrent memory load with sentence complexity can be observed when one manipulates the similarity of the complex sentence and the load on verbal memory rather than the span of the list comprising the memory load; see Gordon et al. (2002), Fedorenko et al. (2006), and Lewis et al. (2006). This general view is provided with further support by Fedorenko et al. (2007) which showed that linguistic complexity (as manipulated by relative clause extraction type in 1 and 2) interacted with verbally mediated tasks such as mathematical addition, but not visuo-spatial tasks such as image rotation.

The results summarized in section 2.3 seem inconsistent with the strict modularist view defended by Caplan and Waters. On the assumption that understanding test questions is an activity that involves unconscious linguistic processes, in conjunction with the assumption that the mathematical task requires conscious mental activity, Caplan and Waters would not expect an interaction between the two. However, the findings do not give us any reason to embrace the capacity explanation advocated by Just and Carpenter (1992) and King and Just (1991). When we looked for a correlation with a measure of working memory that was based on a memory span task, we were disappointed.

In the next section we further examine performance on the WJ-R Applied Problems Subtest to determine whether there is reason to prefer the representational account suggested by Gordon et al. (2002) and Fedorenko et al. (2007).

3. Mathematical Reasoning and Linguistic Representations Interact

3.1. Switching Between Types of Representations

The results of the MCMC model presented in 2.3 support our hypothesis that mismatches between MAE and AAE would depress performance on mathematical reasoning tests. We also found that the depression in test score tracks specific morphological divergences to different degrees, and in this section we explore why that would be true.

Our leading hypothesis from section 1 is that manipulating some types of linguistic representations is difficult, requiring more cognitive resources and diverting them from solving difficult mathematical problems. This explanation has the benefit of being consistent with our MCMC results from section 2. It also sheds some light on the finding reported by Craig and Washington (2004) that dialect switching between AAE and MAE is typically accompanied by reduced sentence complexity on the part of the speaker.

A key part of our hypothesis is that representational mismatches, specifically those that go beyond simple one-to-one relations between morphemes in the two dialects, will require greater cognitive resources than those simple relations. Our reasoning is that many-to-one relations entail two similar morphemic forms in one of the dialects. They are therefore subject to representational 'confusion' in the sense of Gordon et al. (2002) and Fedorenko et al. (2006). The mismatch in question does not need to be semantic in nature. In that sense, the difficulty posed by the third person singular -s could be that there was a semantic mismatch between MAE and AAE present tense marking, or it could be a mismatch between a MAE morpheme that simply has no counterpart in AAE.

To evaluate this aspect of our hypothesis, we returned to the data set outlined in section 2. AAE has mismatches with MAE other than those outlined in Table 1, and we examined mismatches between morphemes in MAE that were systematically absent in AAE as observationally clear instances of a failure of one-to-one mappings between the two dialects. We investigated the possessive *-'s* and the two contractible auxiliary verbs (*be* and *have*) in MAE. While there is certainly variation among individuals (and perhaps regions as well), generally speaking, there is no overt marking of the possessive in attributive position in the adult AAE grammar. Sentences such as *We went to John house yesterday* are perfectly grammatical and preferred to the MAE influenced *We went to John's house yesterday.* While an overt mark-

ing of the possessive is required in absolute final position (e.g. *This book is yours/hers/mines/John's*), these instances may very well be treated as special pronouns forms within the grammar, leaving the productive possessive morpheme systematically absent from AAE. Further, there is evidence that in child AAE grammar, overt marking of the possessive completely absent. Green (2009) finds little evidence of possessive marking in either attributive or absolute position in language samples of AAE speaking children aged 4-5. The picture that emerges is one in which the possessive 's is systematically absent in AAE.

In contrast the contractible auxiliary verbs have clear contexts of use in both AAE and MAE, but have different distributions in the two dialects.[5] This led us to expect that we would find that possessive *-s* in test questions would correlate with performance on test questions. At the same time, on the assumption that the contractible auxiliary verbs were syntactically present in both ME and AAE and subject to phonological deletion in AAE in essentially the way suggested by Labov (1969), we would not expect the appearance of contractible auxiliary verbs in test questions to similarly correlate with test performance.

3.2. Possessives and Auxiliary Verbs

The children's test performance is drawn from the same data set described in section 2. We returned to the Calculation and Applied Problems sub tests from the Woodcock-Johnson-R (WJ-R) Psychoeducational Battery and coded how many instances of possessive *-s*, contractible *be* and contractible *have* occurred in each question. We subsequently ran our MCMC model to understand if any of these variables correlated with students' performance on test questions.

For each linguistic feature, Table 9 shows the value of $\alpha_{i,k}^{M}$ averaged across students, its standard deviation and coefficient of variation.

Table 10 shows that the presence of possessive *-s* in test questions correlates with students' performance. Table 10 indicates that it also exhibits

[5]We assume the analysis of the English possessive offered in Abney (1987) in which it is classed as a determiner D that heads a phrase of the same type, DP, with a nominal specifier that it attaches to prosodically. In the absence of the possessive morpheme, there is no overt evidence to force the presence of such a D (or DP) and the possessor can be syntactically generated as the specifier of the possessed N.

a high degree of variation, suggesting that some students are highly affected by the presence of this particular feature.

3.3. Discussion

The contrast between the possessive pronouns and the possessive *-s* indicates that it is not the semantic mapping from AAE to MAE that has an effect on test performance. This inference is provided with further support from the fact that the auxiliary verbs *be* and *have* only show a weak effect on test performance despite the fact that these auxiliaries appear to have different semantic functions in AAE and MAE (see for example the discussion in Rickford and Thrberge-Rafal (1996) of preterit uses of AAE *had*). What seems to distinguish the possessive *-s* from the other forms in Table 10 is that it is a semantic relation that has an overt morphemic representation in MAE but never has one in AAE. For example, a proper name in AAE will stand in a one-to-many relation with its MAE possessive and non-possessive counterparts. Because of their similarity these MAE counterparts will be susceptible to representational confusion. This is a difference in how the possessive relation is represented in the two dialects. Taken together these observations suggest that it is not the mismatches between AAE and ME generally that pose a cognitive burden for the children in our study. Instead the penalty results from the type of mismatch involved, specifically whether the two dialects have mismatched representations.

The implications of this line of inquiry are quite broad because representations can be mismatched in a number of ways. We have focused on mismatches between phonologically overt morphemes and their phonologically null counterparts and inferred that such mismatches pose a greater cognitive load than mismatches between phonologically overt morphemes that have two distinct semantic values. Of course, it is conceptually possible for phonologically overt morphemes to be mismatched in situations where one has a semantic value but the other does not. It is important to ask whether such situations would pose a cognitive load comparable to the one we have found.

In fact the important work of Johnson et al. (2005) speculates that third person singular *-s* poses a difficulty for children for precisely just such a reason. Their view is that third person singular *-s* only codes formal syntactic features that are not semantically interpreted at all; this view is motivated by a desire to explain an asymmetry between production and comprehension of the morpheme as an agreement marker of plural subjects in children 3-4

years of age. From such a perspective, it is the lack of semantic values in third person singular *-s* rather than a mismatch in its form that poses the cognitive burden. We have not adopted that point of view here because it does not appear to offer a straightforward explanation for the effect of the possessive morpheme outlined in this section.

4. Conclusion

The results of our MCMC model show that linguistic features can have a significant impact on tests of mathematical reasoning. That impact can be facilitative, as in the case of the counterfactual conditional *if + -ed*, or it can be inhibitory, as in the case of third person singular *-s*. The impact is independent of individuals' abilities in spatial reasoning, consistent with other work suggesting that mathematical reasoning and language draw from a common working memory store that is nonetheless distinct from that used in spatial reasoning. The impact that we have documented provides support for our initial hypothesis that dialect switching between AAE and mainstream English poses a cognitive load that affects verbally mediated tasks generally. The character of this load is significant and deserves further exploration. We have suggested here that the nature of the representations that must be manipulated is a critical factor. Specifically, we have suggested that mismatches in the morphological inventory of the two dialects in question have significant effects when one morpheme is phonologically overt and the other is phonologically null and we have tentatively attributed this effect to processing difficulty associated with representational similarity.

Whether other types of mismatches have measurable but weaker effects remains an open empirical question. It is also quite possible that strength of representation rather than the similarity of the representations themselves is the critical issue. Mismatches between home and school dialects may leave AAE speaking children without consistent data as far as the syntactic distribution of morphemes such as those we have investigated here. Such highly variable input data might reasonably be thought to make resulting representations about morphemes weaker than if the input data were highly consistent and regularly activating the posited morphemic representation. Morphemes that are systematically absent in one dialect but present in the other would presumably pose the greatest problems for learning their distributions and meanings, thus leading to weak representations. Future work will focus on finding evidence for that would lead us to prefer one of these two different

representational hypotheses. Our guiding intuition for this future inquiry is to correlate load effects with representations that for different reasons are difficult to infer and maintain from overt physical properties of the speech stream and its context.

A. APPENDIX

A.1. The Model

Let z_{ij} denote the score of student i on question j; z_{ij} takes the value 1 for a correct answer and the value 0 for an incorrect answer. We ignore missing values. Define y_{ij} to be a measure of how well the student knows the answer. y_{ij} is an unobserved random variable such that $z_{ij} = 1$ if $y_{ij} > 0$ and $z_{ij} = 0$ if $y_{ij} \leq 0$. We are interested in the effect of feature k on student i; symbolized by α_{ik}. Besides the six linguistic features that we study, some other factors that might affect the student's answer could be the effect determined by the overall ability of the student (η_i) and the effect of the question driven by the overall difficulty of the question (β_j). For the influence of factor k on question j (x_{jk}) we used how many times the linguistic feature k appears on question j.
The model we use is

$$y_{ij} = \eta_i + \beta_j + \sum_{k=1}^{6} \alpha_{ik} x_{jk} + \epsilon_{ij} \tag{1}$$

for $i = 1, \ldots, 75$ and $j = 1, \ldots, 60$.
Here ϵ_{ij} represents the error of the model which is independent from the other variables (such as socio-economic status or properties of the students' home environment) that we have not taken into account.

A.2. Assumptions

It seems logical to treat all of η_i, β_j, and α_{ik} as random effects because the students and the questions were chosen randomly from a bigger group of students and questions.

Let us assume that the random effects have a normal distribution and that they are independent from each other:

$$\eta_i \sim N(\mu_\eta, \sigma_\eta^2), \; i = 1, \ldots, 75,$$

$$
\begin{aligned}
\beta_j &\sim N(\mu_\beta, \sigma_\beta^2),\ j = 1, \ldots, 60, \\
\epsilon_{ij} &\sim N(0, \sigma_\epsilon^2),\ i = 1, \ldots, 75,\ j = 1, \ldots, 60, \\
\alpha_{ik} &\sim N(\nu_k, \tau_k^2),\ i = 1, \ldots, 75,\ k = 1, \ldots, 6.
\end{aligned}
$$

where μ_η, μ_β, ν_k, σ_η^2, σ_β^2, σ_ϵ^2 and τ_k^2 are unknown and, together with η_i, β_j and α_{ik}, need to be estimated from the data.

A.3. Approach

We will apply a Bayesian-MCMC method to estimate the unknown parameters.

A.3.1. Simplification of the model

Looking at the original assumptions, we can simplify the model:

- The model is only affected by the difference between μ_η and μ_β and not by their individual values. So let us set $\mu_\beta = 0$.
- If we multiply all the y_{ij}'s by the same positive constant, the values of z_{ij} don't change so we can set $\sigma_\epsilon^2 = 1$.

The model becomes:

$$
\begin{aligned}
\eta_i &\sim N(\mu, \kappa_\eta^{-1}), \\
\beta_j &\sim N(0, \kappa_\beta^{-1}), \\
\epsilon_{ij} &\sim N(0, 1), \\
\alpha_{ik} &\sim N(\nu_k, \lambda_k^{-1}), \\
y_{ij} &= \eta_i + \beta_j + \sum_{k=1}^{6} \alpha_{ik} x_{jk} + \epsilon_{ij}, \\
z_{ij} &= \begin{cases} 1 & \text{if } y_{ij} > 0, \\ 0 & \text{if } y_{ij} \leq 0. \end{cases}
\end{aligned}
$$

Note that now we write the variances as κ_η^{-1}, κ_β^{-1} and λ_k^{-1} instead of σ_η^2, σ_β^2 and τ_k^2 to make calculations easier[6].

[6]This is because using a *Gamma* distribution for the prior of κ_η, κ_β and λ_k gives a *Gamma* posterior distribution for those variables, yielding a distribution in a closed form

A.3.2. Priors for the hyperparameters

The parameters μ, ν_k, κ_η, κ_β and λ_k are called hyperparameters.

We used:

$$\begin{aligned}
\mu &\sim U(-\infty,\infty),\\
\nu_k &\sim U(-\infty,\infty),\\
\kappa_\eta &\sim Gamma(a,b),\\
\kappa_\beta &\sim Gamma(a,b),\\
\lambda_k &\sim Gamma(a,b),
\end{aligned}$$

where U denotes the uniform distribution and $Gamma$[7] the gamma distribution and $a = b = 0.01$[8].

We assume that the hyperparameters are independent from each other and also they are independent from the effects α, β and η and from the error ϵ.

A.4. The MCMC algorithm

We will use I for the number of students ($I = 75$), J for the number of questions ($J = 60$) and K for the number of linguistic factors ($K = 6$).

The joint density of $(\kappa_\eta, \kappa_\beta, \lambda_k, \mu, \nu_k, \eta_i, \beta_j, \alpha_{ik}, y_{ij}, z_{ij})$ is proportional to

$$\kappa_\eta^{a-1}e^{-b\kappa_\eta}\cdot\kappa_\beta^{a-1}e^{-b\kappa_\beta}\cdot\prod_{k=1}^{K}\left\{\lambda_k^{a-1}e^{-b\lambda_k}\right\}\cdot\prod_{i=1}^{I}\kappa_\eta^{\frac{1}{2}}e^{-\frac{1}{2}\kappa_\eta(\eta_i-\mu)^2}\cdot\prod_{j=1}^{J}\kappa_\beta^{\frac{1}{2}}e^{-\frac{1}{2}\kappa_\beta\beta_j^2}\cdot$$
$$\prod_{i=1}^{I}\prod_{k=1}^{K}\lambda_k^{\frac{1}{2}}e^{-\frac{1}{2}\lambda_k(\alpha_{ik}-\nu_k)^2}\cdot\prod_{i=1}^{I}\prod_{j=1}^{J}e^{-\frac{1}{2}(y_{ij}-\eta_i-\beta_j-\sum_{k=1}^{K}\alpha_{ik}x_{jk})^2}\cdot Q(y_{ij},z_{ij}) \qquad (2)$$

where

$$Q(y,z)=\begin{cases}1 & \text{if } y>0 \text{ and } z=1,\\ 1 & \text{if } y\le 0 \text{ and } z=0,\\ 0 & \text{otherwise.}\end{cases}$$

All the variables in (2) are unknown except z_{ij}. The Bayesian solution is to construct the conditional density of $(\kappa_\eta, \kappa_\beta, \lambda_k, \mu, \nu_k, \eta_i, \beta_j, \alpha_{ik}, y_{ij})$ given all the z_{ij}. The basic idea of MCMC sampling is to construct a Monte Carlo sample from the joint density (2) by successively updating each of the unknown random variables.

[7]The density for the $Gamma(a,b)$ distribution is $f(x) = \frac{1}{\Gamma(a)}b^a x^{a-1}e^{-bx}$

[8]This is a typical choice for the MCMC

A.5. Updating scale parameters

The scale parameters are κ_η, κ_β and λ_k, $k = 1, \ldots K$

Updating the scale parameters consists of a random sample of one observation from the $Gamma(a', b')$ distribution, where:

- for κ_η, $a' = a + \frac{1}{2}I$ and $b' = b + \frac{1}{2}\sum_i(\eta_i - \mu)^2$
- for κ_β, $a' = a + \frac{1}{2}J$ and $b' = b + \frac{1}{2}\sum_j \beta_j^2$
- for λ_k, $a' = a + \frac{1}{2}I$ and $b' = b + \frac{1}{2}\sum_i(\alpha_{ik} - \nu_k)^2$

A.6. Updating location parameters

The location parameters are μ, ν_k, η_i, β_j and α_{ik}

Updating the location parameters consists of a random sample of one observation from the $N(\frac{B}{A}, \frac{1}{A})$, where:

- for μ, $A = I\kappa_\eta$, $B = \kappa_\eta \sum_i \eta_i$
- for ν_k, $A = I\lambda_k$, $B = \lambda_k \sum_i \alpha_{ik}$
- for η_i, $A = \kappa_\eta + J$, $B = \mu\kappa_\eta + \sum_j(y_{ij} - \beta_j - \sum_k \alpha_{ik}x_{jk})$
- for β_j $A = \kappa_\beta + I$, $B = \sum_i(y_{ij} - \eta_i - \sum_k \alpha_{ik}x_{jk})$
- for α_{ik}, $A = \lambda_k + \sum_j x_{jk}^2$, $B = \lambda_k\nu_k + \sum_j x_{jk}(y_{ij} - \eta_i - \beta_j - \sum_{k' \neq k} \alpha_{ik'}x_{jk'})$

A.7. Updating y_{ij}

The conditional distribution of y_{ij} given all the other unknowns is $N(\eta_i + \beta_j + \sum_k \alpha_{ik}x_{jk}, 1)$ (including the condition $Q(y_{ij}, z_{ij}) = 1$). We use rejection sampling to sample y: we generated consecutive values from the conditional distribution until the condition $Q(y_{ij}, z_{ij}) = 1$ is satisfied.

A.8. Implementation

For starting values, we set $y_{ij} = 1$ when $z_{ij} = 1$ and $y_{ij} = -1$ when $z_{ij} = 0$. We set all the location parameters equal to 0 and all the scale parameters equal to 1. We then ran 10,000 iterations as "burn in" updating all the unknowns. The results were discarded. This is done so that the starting values that we chose for the first step do not affect the results. From the next 100,000 iterations, we kept every 100th value and we ended up with

Figure 1: Plot of $\hat{\bar{z}}_{ij}^{[l]}$'s against $\bar{z}_{ij}^{[l]}$'s

a sample size of 1000 from the posterior distributions of the unknowns. We symbolize the nth observation in the sample with the superscript $^{(n)}$, for example $\alpha_{3,10}^{(45)}$ means the 45th observation in our sample for the parameter $\alpha_{3,10}$.

A.9. Checking the fit of the model

We can use at least two different methods to check how well our model explains the data.

A.9.1. Using the estimated values, $\hat{z}_{ij}$ compared to the original values of z_{ij}

Using the simulated data, we calculated the values $\hat{z}_{ij}^{(n)}$, where $\hat{z}_{ij}^{(n)}$ is the estimated value of z_{ij} for the nth observation.

First we calculate $\hat{y}_{ij}^{(n)}$ by:

$$\hat{y}_{ij}^{(n)} = \eta_i^{(n)} + \beta_j^{(n)} + \sum_{k=1}^{6} \alpha_{ik}^{(n)} x_{jk}$$

where $\eta_i^{(n)}$, $\beta_j^{(n)}$ and $\alpha_{ik}^{(n)}$ refer to the nth observation in the sample. Then we set $\hat{z}_{ij}^{(n)} = 1$ if $\hat{y}_{ij}^{(n)} > 0$ and $\hat{z}_{ij}^{(n)} = 0$ if $\hat{y}_{ij}^{(n)} \leq 0$. This is done for each n. For each pair (i, j) we calculate the sample mean $\hat{\bar{z}}_{ij}$ of $\hat{z}_{ij}^{(n)}$, $\hat{\bar{z}}_{ij} = \frac{1}{1000} \sum_n \hat{z}_{ij}^{(n)}$. This is a number between 0 and 1. We then divide the interval $[0, 1]$ into $L = 10$ equally spaced subintervals: $[0.0, 0.1), [0.1, 0.2), \ldots, [0.9, 1.0]$ and take the average $\hat{\bar{z}}_{ij}^{[l]}$ of the $\hat{\bar{z}}_{ij}$'s that belong to the subinterval l, $l = 1, \ldots, L$. This defines a set of pairs (i, j). We also take the average $\bar{z}_{ij}^{[l]}$, of the observed z_{ij}'s, for those (i, j)'s.

We expect that if we plot the $\hat{\bar{z}}_{ij}^{[l]}$'s against the $\bar{z}_{ij}^{[l]}$'s, then we will get a straight line. The plot is shown in figure 1. The correlation is 0.8702.

A.9.2. Plots of the median of β_j's against the proportion of correct answers for question j

The overall difficulty of question j is estimated by β_j. We expect an increasing pattern between the median of β_j's and the average number of

Figure 2: Plot of β against question average correct answers.

correct answers for each question[9].

The plot is shown in figure 2. The increasing relationship is more obvious here: the correlation is 0.9818.

We can infer that the original and simulated data are explained sufficiently well by the model that we can proceed to analyze the simulated data with some confidence.

References

Abney, S., 1987. The english noun phrase in its sentential aspect. Ph.D. thesis, MIT.

Baayen, R. H., 2008. Analyzing Linguistic Data: A Practical Introduction to Statistics. Cambridge University Press, Cambridge, UK.

Baddeley, A., 1986. Working Memory. Oxford University Press, Oxford, UK.

Beyer, T., Kam, C. H., 2009. Some cues are stronger than others: The (non)interpretation of third person present s as a tense marker by 6 and 7 year olds. First Language 29, 208–222.

Craig, H. K., Washington, J. A., 2004. Grade-related changes in the production of african american english. Journal of Speech, Language and Hearing Research 47, 450–463.

Craig, H. K., Washington, J. A., 2006. Malik goes to school: Examining the language skills of African American Students from preschool –5th grade. Lawrence Erlbaum, Mahwah, NJ.

Daneman, M., Carpenter, P., 1980. Individual difference in working memory and reading. Journal of Verbal Learning and Verbal Behavior 19, 450–466.

[9]We only used averages for questions 13 to 39 because there were not enough answers for the other questions

deVilliers, J., Johnson, V., 2007. The information in third-person /s/: Acqusition across dialects of american english. Journal of Child Language 34, 113–158.

Fedorenko, E., Gibson, E., Rohde, D., 2006. The nature of working memory capacity in sentence comprehension: Evidence against domain-specific working memory resources. Journal of Memory and Language 54.

Fedorenko, E., Gibson, E., Rohde, D., 2007. The nature of working memory in linguistic, arithmetic and spatial integration processes. Journal of Memory and Language 56, 246–269.

Fodor, J., 1983. Modularity of Mind. MIT Press, Cambridge, MA.

Gordon, P. C., Hendrick, R., Levine, W., 2002. Memory-load interference in syntactic processing. Psychological Science 13, 425–430.

Green, L., 2000. Aspectual be-type constructions and coercion in african american english. Natural Language Semantics 8, 1–25.

Green, L., 2002. African American English: A linguistic introduction. Cambridge University Press, Cambridge, UK.

Green, L., 2009. Language and the African American Child [in preparation]. Cambridge University Press, Cambridge,UK.

Johnson, V. E., Devilliers, J. G., Seymour, H. N., 2005. Agreement without understanding? thecase of third person singular /s/. First Language.

Just, M. A., Carpenter, P. A., 1992. A capacity theory of comprehension: individual differences in working memory. Psychological Review 99, 122–149.

King, J., Just, M. A., 1991. Individual differences in syntactic processing: the role of working memory. Journal of Memory and Language 30, 580–602.

Labov, W., 1969. Contraction, deletion and inherent variablity of the english copula. Language.

Labov, W., 1972. Language in the Inner City. Studies in the Black English Vernacular. University of Pennsylvania Press, Philadelphia, PA.

Lewis, R., Vasishth, S., Dyke, J. A. V., 2006. Computational principle of working memory in sentence comprehension. Trends in Cognitive Science 10, 447–454.

Miller, G., 1956. The magical number seven, plus or minus two: Some limits on our capacity for processing information. Psychological Review 63, 81–97.

Miller, G., Chomsky, N., 1963. Finitary models of language users. In: Luce, R., Bush, R. R., Galanter, E. (Eds.), Handbook of mathematical psychology. Vol. II. Wiley, New York, pp. 419–491.

Miller, J., Chapman, R., 2000. Systemic Analysis of Language Transcripts (SALT) [Computer Software]. University of Wisconsin, Language Analysis Lab, Madison, WI.

Mufwene, S., Rickford, J. R., Bailey, G., Baugh, J., 1998. African American English: Structure, history and use. Routledge, London.

Rickford, J., Thrberge-Rafal, C., 1996. Preterite had + v-ed in the narrative of african american adolescents. American Speech 71.

Roberts, J. E., Burchinal, M., Medley, L. P., Zeisel, S. A., Mundy, M., Rousch, J., Hooper, S., Bryant, D., Henderson, F., 1995. Otis media, hearing sensitivity, and maternal responsiveness in relation to language during infancy. Journal of Pediatrics 126, 481–489.

Terry, J. M., Evangelou, E., Smith, R. L., 2009. Dialect switching and african american english: Implications for early childhood acheivement, manuscript.

Wake, W. K., Gardner, H., Kornhaber, M. L., 1996. Intelligence: multiple perspectives. Harcourt Brace College Publishers, Fort Worth, TX.

Waters, G. S., Caplan, D., 2001. Age, working memory and on-line syntactic processing in sentence comprehension. Psychology and Aging 16, 128–144.

Wolfram, W., 1974. The relationship of white southern speech to vernacular black english. Language 50, 498–527.

Wolfram, W., Fasold, R., 1974. The study of social dialects in American English. Prentice Hall, Englewood Cliffs, NJ.

Young, G. A., Smith, R. L., 2005. Essentials of statistical inference. Cambridge University Press, Cambridge, UK.

Morphemes coded	Example in MAE
Past tense -ed	Jill walked to school.
Participle -en	Jill has written a letter.
Past tense copula was/were	Jack and Jill were late for school.
Auxiliary have	Jill has climbed that hill.
3rd person singular -s	Jill eats a lot of ice-cream.
Counterfactual Conditional if+ed	If Jill walked to school, she would get there earlier than she does.

Table 1: The Six Mophosyntactic Features Analyzed

	Correlation	p-value	Lower bound	Upper bound
Past tense -ed	-0.1394290	0.23290	-0.090394	0.3551520
Participle -en	0.1242560	0.28820	-0.105686	0.3415850
Past tense copula was/were	0.1899870	0.10260	-0.038640	0.3997140
Have	0.1556990	0.18220	-0.073873	0.3695100
Counterfactual conditional if+ed	0.4315980	0.00011	0.2268588	0.5998054
Present 3rd singular -s	0.5606250	0.00000	0.3823090	0.6986860

Table 2: The correlation between students' overall scores and $\alpha_{i,k}^{M}$, the effect of a feature on a student.

	Mean of $\alpha_{i,k}^{M}$	Standard deviation of $\alpha_{i,k}^{M}$	Coefficient of variation
Past tense -ed	-0.028219	0.046712	-1.655342
Participle -en	-0.072917	0.286117	-0.392388
Past tense copula was/were	-0.075168	0.033766	-0.449214
Have	-0.103199	0.034436	-0.033368
Counterfactual conditional if+ed	0.121299	0.050015	0.412331
Present 3rd singular -s	0.502120	0.681340	-1.357046

Table 3: The correlation of students' scores and averaged over students, standard deviation, and coefficient of variation

	Correlation	p-value	Lower bound	Upper bound
Past tense -ed	0.1383601	0.2365	-0.09147423	0.355419925
Participle -en	0.1102468	0.3464	-0.1197104	0.3289770
Past tense copula was/were	0.1784003	0.1257	-0.05061082	0.38958768
Have	0.1396023	0.2323	-0.0902181	0.3553064
Counterfactual conditional if+ed	0.419099	0.0001823	0.2123343	0.5899455
Present 3rd singular -s	0.7213507	0.00000	0.5911756	0.814892

Table 4: The correlation between students' overall scores, $\alpha_{i,k}^{M}$ (the effect of a feature on a student), and students' AAE production.

	Correlation	p-value	Lower bound	Upper bound
Past tense -ed	0.1326645	0.2565	-0.09722475	0.34911459
Participle -en	0.1539759	0.1872	-0.07562888	0.36807480
Past tense copula was/were?	0.1497014	0.1999	-0.0799777	0.36428611
Have	0.2104058	0.07	-0.091738633	0.41743347
Counterfactual conditional if+ed?	0.4345985	0.0	0.2303586	0.6021645
Present 3rd singular -s	0.727889	0.00000	0.6000587	0.8194717

Table 5: The correlation between students' overall scores, $\alpha_{i,k}^{M}$ (the effect of a feature on a student), and the proportion of students' production of third person singular -s.

	Correlation	p-value	Lower bound	Upper bound
Past tense -ed	0.05078606	6696	-0.1814007	0.2776097
Participle -en	0.4432727	0.00	0.2374179	0.6110260
Past tense copula was/were	0.1302823	0.2719	-0.1028683	0.3498629
Have	-0.07661842	0.5194	-0.1562023	0.3013733
Counterfactual conditional if+ed	0.4432727	0.000	0.2374179	0.6110260
Present 3rd singular -s	0.5800434	0.00000	0.4038725	0.7147306

Table 6: The correlation $\alpha_{i,k}^{M}$ and test performance in a model with a parameter of students' spatial reasoning.

	Mean of $\alpha_{i,k}^{M}$	Standard deviation of $\alpha_{i,k}^{M}$	Coefficient of variation
Past tense -ed	-0.0371303428065367	0.0447035897231647	-1.20396382969281
Participle -en	-0.222999536775056	0.0426470790500726	-0.191242904208772
Past tense copula was/were	-0.332554668659929	0.0390566696744618	-0.117444358342181
Have	-0.226751692754703	0.0416611386569064	-0.183730221154181
Counterfactual conditional if+ed	0.101406525950873	0.0506289227621783	0.49926690898282 7
Present 3rd singular -s	-0.0682975000919536	0.067916055275314	-0.994414951994935

Table 7: The variation of $\alpha_{i,k}^{M}$ in a model of test performance including a parameter of working memory span

	Correlation	p-value	Lower bound	Upper bound
Past tense -ed	0.1168306	0.3249	-0.1163644	0.3378177
Participle -en	0.06432672	0.5887	-0.1682303	0.2901007
Past tense copula was/were	0.1874357	0.1123	-0.04455287	0.40024332
Have	0.1294739	0.2749	-0.1036818	0.3491411
Counterfactual conditional if+ed	0.1874357	0.1123	-0.04455287	0.40024332
Present 3rd singular -s	0.5799727	0.00000	0.4037834	0.7146785

Table 8: The correlation between students' overall scores, $\alpha_{i,k}^{M}$ and students' working memory.

	Correlation	p-value	Lower bound	Upper bound
Possessive pronoun	0.1560828	0.181100	-0.8348210	0.36993950
Possessive -s	0.3543088	0.001816	0.13848310	0.53800710
Contractible be	0.1304694	0.264600	0.09943697	0.34715500
Contractible have	0.1420290	0.224200	-0.08776185	0.35746783

Table 9: The correlation between students' overall scores and $\alpha_{i,k}^{M}$.

	Mean	Standard deviation	Coef. of variation
Possessive pronoun	0.087639	0.044155	-0.50383000
Possessive -s	0.030060	0.071893	2.46077500
Contractible be	0.068930	0.031360	-0.45777600
Contractible have	0.094694	0.325494	-0.34426100

Table 10: The mean, standard deviation, and coefficient of variation of $\alpha_{i,k}^{M}$