

Examining the use of standard language production measures in the language samples of African-American toddlers

RAMONDA HORTON-IKARD¹, SUSAN ELLIS WEISMER², &
CLAIRE EDWARDS²

¹*Department of Audiology & Speech Language Pathology, University of Tennessee, Knoxville, USA and*
²*Department of Communicative Disorders & The Waisman Center, University of Wisconsin, Madison, USA*

Abstract

This study examined the relationship between several measures of language production, age, and non-standard speech use in toddlers from African-American English (AAE) backgrounds. Free play language samples were collected from two age groups of typically developing African-American toddlers. Group one consisted of 11 toddlers who were 2½ years old and group two consisted of 11 toddlers who were 3½ years old. Measures of mean length of utterance in morphemes (MLU), mean syntactic length, Index of Productive Syntax scores, and a dialect density rate were extracted from the language samples. Results indicated that MLU, MSL, and IPSYN scores were sensitive to differences in chronological age for the toddlers. These findings suggest that measures other than MLU, such as the IPSYN, may prove to be useful and valid tools for monitoring the language growth of young speakers from AAE backgrounds.

Keywords: *African-American English, language development, toddlers, MLU, MSL, IPSYN*

Currently, there are a number of researchers and clinicians who use language sample analysis (LSA) to evaluate the language performance of children who are from linguistically diverse backgrounds. This practice has been promoted as a viable alternative or supplement to standardized testing procedures that may prove to be biased against individuals who are not learning Standard American English (SAE—Leonard & Weiss, 1994; Craig & Washington, 1998). LSA offers several distinct advantages over standardized testing when evaluating the performance of children from linguistically diverse backgrounds. Stockman (1996) argues that LSA is an ideal tool for evaluating the language of children from linguistically diverse backgrounds because it is culturally sensitive, flexible, and accessible. However, several researchers, including Stockman, have acknowledged that there are some disadvantages to using LSA to evaluate young children's language abilities. These disadvantages include management of context variation, the subjective nature of

Correspondence: RaMonda Horton-Ikard, Ph.D, Department of Audiology and Speech Pathology, 578 South Stadium Hall, University of Tennessee-Knoxville, Knoxville, TN 37996-0740, USA. Tel: 1 865 974 3739. Fax: 1 865 974 1539.
E-mail: rhortoni@utk.edu.

observation, the necessity of detailed and elaborate procedures for transcription, and time constraints. Beyond the concerns of sampling procedure, the production measures extracted from LSA may also be problematic for children who are learning a dialect other than SAE.

Language production measures such as mean length of utterance (MLU), number of different words (NDW), and mean length of response (MLR) have been useful in evaluating several components of language (i.e. morphological, syntactic, lexical, and discourse development) in children for clinical and research purposes. However, many of these production measures are typically used to describe the performance of children who are learning SAE, rather than a cultural or regional dialect of English such as African-American English (AAE).

Currently, AAE is one of the most frequently studied 'non-standard' varieties of the English language (Baugh, 1997; Crystal, 1999). AAE is a linguistic system in which optional variability, unique lexical and grammatical forms, and discourse styles characterize the dialect. In contrast, SAE is defined as the standard language variety used by educated, mainstream Americans. Within the SAE dialect there is minimal use of socially marked forms (Wolfram, 1991; Wolfram, Adger & Christian, 1999). This distinction between standard and non-standard varieties of a language is primarily based on sociological judgments, rather than linguistic ones. For that reason, many of the features associated with AAE have been labeled as socially stigmatizing forms and it is these non-standard forms that are of interest for the current study.

The learning and use of a cultural dialect such as African-American English (AAE) is bound to have an impact on the outcome of standard measures of grammatical development. For example, Kovac and Adamson (1981) posited that 'sociodialectal' processes influenced African-American children's production of variable copula absence. They found that by age 5, children who came from non-SAE backgrounds were more likely to be producing greater instances of variable copula absence in a manner that was not consistent with the simple developmental processes that were demonstrated by white preschool children from SAE backgrounds. These findings suggest that measures of grammatical development, such as MLU, might result in different values than those we typically see for children from SAE backgrounds since MLU is based on the number of morphemes (words and inflections) that appear in a sample.

Empirical data which describes the grammatical development of typically developing African-American children from AAE backgrounds, as measured by MLU, will be especially important for clinical and research purposes. MLU has often been used in clinical settings to screen for possible delays in language. Within research investigations, MLU has also been used to match typically developing participants to a group of participants with atypical language behaviors. However, what remains questionable is the use of this measurement for describing the linguistic growth of typically developing and atypically developing African-American children who may be learning and using AAE. Even though researchers agree that applying the norms of formal tests to groups of children who were not adequately represented in the original standardization sample produces biased results, scant attention has been paid to the biases that might occur when generalizing the norms for spontaneous language production measures to children from diverse backgrounds (Craig & Washington, 1998; Oetting, Cantrell & Horohov, 1999).

The original normative data concerning the use of MLU as a metric of grammatical development was comprised of a sample of children from Madison, Wisconsin, a primarily White, middle to upper class community (Miller & Chapman, 1981). It has been well established that a number of socio-cultural factors impact developmental outcomes, so there is a strong likelihood that these norms are not applicable to children from different

backgrounds. Dollaghan et al. (1999) found that there was a direct relationship between maternal level of education and language production measures of their group of 3-year-old children, such that those children whose parents had the lowest level of education had the lowest MLUs, and those children whose parents had the highest level of education had the highest MLUs. The Dollaghan investigation provided much needed information concerning the relationship between SES and language production measures. Missing from the report, however, was a specific examination of other socio-linguistic variables such as AAE use for the group of African-American children who participated in the study.

Relatively little information is available concerning the performance of young typically developing AAE speakers on standard language productions measures such as MLU. Stockman (1996) described the language performance of a small group of children ($n = 7$) who were between 33 and 36 months of age. Her findings indicated that toddlers from AAE backgrounds produced a minimum MLU of 2.7 and a maximum MLU of 3.74. These MLU values were comparable to the frequently used norms for White toddlers who are learning SAE (Miller & Chapman, 1981). The small data set and the limited age range of the participants included in Stockman's investigation make it difficult to determine if MLU is age sensitive for young AAE speakers, in that growth in utterance length corresponds with growth in chronological age. Due to the limitations of this study the findings can only be viewed as tentative and preliminary. Additionally, there are no other published data available which might support this as a reasonable MLU expectation for toddlers who use AAE.

Beyond MLU there are additional measures of language production that have been used to describe the language performance of young, typically developing and language-impaired children. For example, Klee and Fitzgerald (1985) adopted the use of the metric, Mean Syntactic Length (MSL) to examine the relationship between utterance length and age. The calculation of MSL controlled for possible pragmatic and discourse influences on mean utterance length by eliminating single morpheme responses. MSL values for the group of Klee and Fitzgerald toddlers indicated that there was a better correlation between MSL and age, rather than the traditional MLU metric and age. Johnston (2001) provided greater detail concerning the calculation of another alternative measure-MLU2, quite similar to Klee and Fitzgerald's MSL with a few exceptions. Her calculation of the MLU2 specified the types of utterances that should be eliminated when attempting to minimize the impact of discourse biases. Johnston's participants spanned a wider age range (2:0 – 6:0) and incorporated children diagnosed with specific language impairment. Johnston's findings also indicated significant correlations between age, MLU2, and traditional MLU. The use of this type of alternative measurement with young child speakers has also been limited in that the participants were primarily SAE speakers, or that information related to dialect use was not provided.

The Index of Productive Syntax (IPSYN) has also been used in several studies concerned with early language development and delay (Rescorla, Kahlgaard & Roberts, 2000; Rispoli & Hadley, 2001; Thal, Reilly, Seibert, Jefferies & Fenson, 2004). It is one of the few type-based tools used to evaluate syntactic and morphological complexity in children below age 4. The IPSYN procedure was designed as a tool for primarily evaluating the emergence of grammatical forms, rather than the mastery of forms. Oetting (in press) notes that because the IPSYN searches a child's entire sample for only two instances of 56 morphosyntactic patterns that are found in a number of English dialects, the potential for unfairly penalizing a child for non-standard form usage (i.e. morphological omissions) is minimized. Oetting, Cantrell & Horohov (1999) suggests that the unique scoring system, which focuses on types

rather than tokens, makes the IPSYN an ideal tool for evaluating the language of children who are likely to use dialects other than SAE.

Only a paucity of research exists that examines the use of standard measures of language production in children who are learning a dialect other than SAE. This existing research focuses primarily on children who are at or above 4 years of age. In a previous study concerned with dialect use and language production measures, Oetting et al. (1999) examined the performance of typically developing and language impaired children from a rural area of southeastern Louisiana. Several measures were extracted from the spontaneous language samples of children aged 4–6 years old to determine if dialect use had a significant impact on the calculation of the following production measures: MLU, IPSYN, and Developmental Sentence Scoring (DSS). The findings from the Oetting study indicated that children's use of non-mainstream dialect did not affect the calculations of the IPSYN, DSS and MLU values. This suggests that the use of popular production measures like the IPSYN may still be useful in spite of children's non-standard dialect use.

The findings characterized by the above studies indicate that continued descriptions of normative language behaviors in children that use a dialect other than SAE are warranted in order to understand the nature of impairment in such populations. In particular, those investigations that examine the language behaviors of younger groups of typically developing children than those included in the Oetting et al. (1999) study will be especially relevant to scientific inquiry concerned with the prevention, assessment, and treatment of early language delay in children from culturally and linguistically diverse backgrounds. The purpose of this study was to extend the findings of previous research on the language behaviors of young African-American speakers and the tools used to describe their language behaviors. The current study examined the relationship between three standard measures of language production and AAE dialect use in a group of typically developing African-American toddlers from AAE backgrounds. The three measures of expressive language production that were examined for this study included mean length of utterance (MLU), mean syntactic length (MSL), and the Index of Productive Syntax (IPSYN). The specific questions guiding this study were:

1. Are there age differences in the grammatical development of African-American toddlers on three different measures of expressive language development (MLU, MSL, and IPSYN scores)?
2. Is there a relationship between MLU, MSL, IPSYN scores and AAE dialect use?

Method

Participants

Twenty-two typically developing African-American toddlers from Madison, Wisconsin, participated in this investigation. Children were assigned to two different groups based on chronological age. Group 1 consisted of eleven 2½-year-old African-American toddlers from AAE backgrounds ($M = 31.09$ months, $SD = 2.21$). Group 2 consisted of eleven 3½-year-old toddlers from AAE backgrounds ($M = 42.27$ months, $SD = 1.19$). The two groups were matched as closely as possible based on the distribution of maternal level of education, as measured by the number of years of schooling; this is an index of SES that has been shown to be related to language development (Hoff-Ginsberg, 1991; Chapman, Schwartz & Kay-Raining Bird, 1995). To ensure that differences in SES were not responsible for differences

in the amount and appearance of AAE features, group comparisons using independent sample *t*-tests were carried out on these two variables between the two groups. Differences in maternal level of education between the 2½- and 3½-year-old toddlers were not significant, $t(20) = -0.837$, $p = 0.413$.

Children with a suspected or confirmed hearing loss, cognitive and linguistic delays, chronic otitis media, or developmental disabilities were excluded from this study. An audiologist assessed hearing abilities using distortion product otoacoustic emissions (DPOAEs) guidelines set forth by the Joint Committee on Infant Hearing (JCIH, 2000). All children who participated in this investigation passed the required hearing screening.

Cognitive functioning was assessed based on children's performance on nonverbal items from the *Bayley Scales of Infant Development*, 2nd edition (Bayley, 1993). The *Mental Scale* of the *Bayley* assesses a variety of skills and includes both nonverbal and verbal items. To obtain a non-biased assessment of cognitive abilities in children with suspected language delays, only the nonverbal items from this measure have been used in various studies of typically developing and late-talking toddlers (Rescorla, 1993; Spitz, Tallal, Flax & Benasich, 1997; Ellis Weismer & Evans, 2002) similarly, it was assumed that the nonverbal items would be a more appropriate index of cognition in toddlers from diverse linguistic backgrounds.

Language ability was assessed using the *Preschool Language Scales-3 (PLS-3)* by Zimmerman, Steiner & Pond (1992). All children participating in this study scored no lower than 1.25 standard deviations below the mean for their age on both portions of this measure. Given the potential for cultural and linguistic bias, a score no lower than 1.25 SD below the mean was considered to be an adequate determination of normal range language performance. See Table I for a complete description of participant characteristics and performance on standardized measures.

Dialect status, as confirmed by the first author, was based on exposure to AAE and determined by the following factors: parental dialect use; peer language use; and daycare provider language use. Parents who demonstrated use of at least four AAE features (along morphological, syntactic, or phonological dimensions) during a 15-minute parent-child language sample were considered to be AAE users. Likewise, their toddlers were considered to be from AAE backgrounds (meaning that their primary language exposure and dialect being acquired was indeed AAE). The children who participated in this study did attend daycare and so a brief description of the daycare where these participants were recruited is in order. The African-American toddlers who were recruited to participate in this study from the first daycare facility spent at least 8 hours a day at the center, 4-5 days per week. The

Table I. Group means for the participant characteristics and language abilities for the toddlers, broken down by age group; standard deviations are reported in parentheses.

	2½-year-old toddlers	3½-year-old toddlers
Chronological age ^a	31.09 (2.21)	42.27 (1.19)
Maternal education ^b	13.63 (1.74)	14.36 (2.29)
Total Language-PLS-3 ^c	100.02 (17.02)	108.09 (15.51)
Bayley Nonverbal Cognitive Scores ^d	0.72 (0.17)	0.74 (0.13)

^aAge is reported in months.

^bMaternal education is reported in terms of number of years of schooling.

^cTotal Language Standard scores for the *Preschool Language Scale-3*.

^dProportion correct on items from the *Bayley Scales of Infant Development*

staff was comprised of two co-directors, six teachers, four aides, and anywhere between three and five teenage volunteers a week. All of the staff members were heavy AAE users with only the directors demonstrating code-switching abilities once they were interacting with outsiders/visitors. Peer interactions among the children reflected use of AAE. Interactions between the parents, children, and investigator routinely involved use of AAE. For all intents and purposes, the children at this center were at a point in their development where they received relatively little or no exposure to SAE and a great deal of exposure to AAE. While these are crude measures of dialect use, similar procedures have been effectively used to characterize the dialect usage of participants within other investigations (Cole, 1980; Lucas & Borders, 1994; Washington & Craig, 1994; Jackson, 1998; Smith, Lee & McDade, 2001).

Data collection and procedure

A 15–20-minute language sample was collected from each child during a semi-structured free-play context. These samples involved interaction between each child and the examiner. The examiner who elicited the language samples was African-American and used AAE. Materials for the free play context consisted of a standard set of toys, including a dollhouse with accessories, and family member play figures. Language samples were recorded using a Marantz audio-recorder, a SONY video recorder, lapel microphone worn by the examiner, and a PZM microphone placed at floor level (next to the participant) to enhance video sound recording.

Language sample analysis

Language samples were orthographically transcribed from audiotapes using *Systematic Analysis of Language Transcripts* (SALT, Miller & Chapman, 2000). Several graduate students in communicative disorders, with previous SALT training, transcribed all of the language samples. Prior to this investigation, these individuals had transcribed a substantial number of language samples involving toddlers and preschoolers (ages 2–4 years). The transcribers were not privy to the research questions for this study; nor were they were familiar with the speakers.

It is also important to note that because the primary purpose of this study is to specifically examine the effects of AAE use on measures of grammatical development there were no scoring adjustments for the use of non-standard forms. This means that utterances containing AAE forms were included in the analysis set for all three measurements.

MLU and MSL

Utterance segmentation was based on standard SALT procedures, which define an utterance as a verbal production typically separated by pauses and/or intonation cues and allowing no more than two independent clauses. SALT calculates MLU in morphemes (MLU-M) based on Miller and Chapman's modification of Brown's (1973) original procedures for counting morphemes (1981). Miller and Chapman's procedures include only complete and intelligible utterances in the analysis set and instances of mazes (i.e. false starts, repetitions, and reformulations) are excluded from the analysis.

The second measure of utterance length, MSL has been used in previous studies concerned with grammatical development (Klee & Fitzgerald, 1985; Johnson, 2001). The calculation of MSL is similar to the procedures that SALT uses to calculate MLU, however,

it involves the use of an analysis set that eliminates one-word utterances or 'single morpheme responses'. Using similar procedures as Johnston (2001), the following types of utterances were eliminated from the SALT analyses; repetitions, yes/no acknowledgements or yes/no answers to questions, responses to intonation prompts, imitations, and all other one-word utterances.

IPSYN

The scoring procedures employed within the current investigation follow the guidelines developed by Scarborough in which the occurrences of 56 forms are identified and given a maximum credit of two points when two exemplars of the targeted form have been produced. If only one exemplar is identified then only one point is credited for that specific form. Once two exemplars have been identified no more points can be awarded. Finally, points are tallied and a set of sub-scores for noun phrases, verb phrases, questions/negations, and sentence structure, in addition to total language score, are available to describe syntactic and morphological development. The number of utterances produced during the 15-minute sampling period ranged from 80 to 173 complete and intelligible utterances (excluding mazes, repetitions, yes/no acknowledgements, and imitations). For this reason, scoring for the IPSYN occurred only for the first 75 utterances that met the criteria.

The 75-utterance analysis set was deemed adequate for interpreting the results of the current investigation for three primary reasons. First, Scarborough (1990) provided data on the performance of similarly aged children within her sample of toddlers and preschoolers for corpus lengths. Secondly, the correlation values for IPSYN scores based on as little as a 50-utterance analysis set and a 100-utterance analysis set during the original Scarborough investigation were very high ($r = 0.963$). Finally, because Scarborough provides a conversion table for corpus lengths between 50 and 95 utterances we can make some preliminary predictions about how IPSYN values for this group of children might look if corpus lengths were greater than those obtained for the present study.

Coding for features of AAE

Coders examined each of the completed samples and assigned a feature type to each of the vernacular forms based on a compiled list of morpho-syntactic features typically associated with use of AAE which have been previously observed in children 4 years of age or younger (Horton-Ikard & Ellis Weismer, in press). Coders were trained in assigning AAE form codes by the first author during earlier projects examining AAE use in children (Horton-Ikard, Ellis Weismer & Evans, 2001; Horton-Ikard & Miller, 2004). It is estimated that these individuals had coded between 69 and 134 language transcripts of preschool and school-aged African-American children who used AAE, prior to the current investigation, and so they did have experience in transcribing the language samples of children from non-SAE backgrounds. Frequencies of the various types of AAE forms were obtained to aid in the calculation of total number of features and dialect density for each child's sample and transferred into a database for later statistical analyses.

Dialect density measure

Several researchers have currently adopted the use of a dialect density measure to describe the amount of dialect used within a language sample (Oetting & McDonald, 2002; Washington & Craig, 2002). A dialect density measure was calculated for each of the

participants within this investigation to describe non-standard speech use on all of the complete and intelligible utterances. Dialect density was calculated by dividing the number of AAE tokens by the number of utterances appearing in the child's sample. It should be noted that the dialect density measure does not imply that the children are solely using AAE dialect. This measure simply indicates how much of their talk contains those non-standard forms that are typically associated with the AAE dialect. The use of the dialect density measure allows the results from the current investigation to be adequately compared to other studies that have examined non-standard dialect use in young children (Oetting et al., 1999; Washington & Craig, 2002).

Transcription and coding reliability

The reliability procedures for the orthographic transcription of language samples using the SALT program were as follows. One person initially transcribed and entered each sample into the program. A second person then listened to and reviewed the same sample in order to note and correct errors or disagreements about coding decisions and actual transcription of utterances with the first transcriber. Four out of 22 of the participant samples (22%) were selected to examine agreement for utterance segmentation and morpheme-by-morpheme transcription. Initial agreement for utterance segmentation was 95% (343 agreements/384 total judgments); morpheme-by-morpheme transcription agreement was 91% (1822 agreements/2003 total judgments). Point-by-point agreement on the identification of AAE morpho-syntactic features was 86% (158 agreements/184 total judgments). The reliability of the IPSYN was obtained by examining the same four samples. IPSYN reliability was at 89% (34 disagreements for points/the 382 total number of points assigned).

Results

Preliminary analyses were conducted to establish that the language samples obtained across the groups of toddlers were roughly equivalent with respect to standard language production measures. Table II provides descriptive information concerning total utterance production for all measures extracted from the language samples, as well as the number of multiword utterances used for the calculation of MSL. There were no significant differences in the number of total utterances produced, $t(20) = 0.332$, $p = 0.744$. In addition, there were no significant differences in the total number of multi-word utterances produced, $t(20) = 0.267$, $p = 0.792$.

Standard language production measures

Multivariate analysis of variance (MANOVA) was performed to determine if there were significant differences between the $2\frac{1}{2}$ - and $3\frac{1}{2}$ -year-old toddlers on the three dependent

Table II. Mean summary information for language samples, broken down by age-groups; standard deviations are reported in parentheses.

	$2\frac{1}{2}$ -year-old toddlers	$3\frac{1}{2}$ -year-old toddlers
Number of utterances	105 (27)	106 (34)
Number of multi-word utterances	78 (24)	82 (32)

variables (MLU, MSL, and IPSYN scores). Partial eta squared (η_p^2) was used as a measure of effect size. This measure reflects the proportion of the effect plus error variance that is attributable to that effect; consequently, it provides an index of the amount of variance accounted for in the sample. The results of the MANOVA indicated that age had a significant effect on: MLU [$F(1,20) = 4.813$, $p = 0.040$, $\eta_p^2 = 0.194$]; MSL [$F(1, 20) = 11.904$, $p = 0.003$, $\eta_p^2 = 0.373$]; and IPSYN scores [$F(1,20) = 35.89$, $p = 0.000$, $\eta_p^2 = 0.642$]. See Table III for group means on MLU, MSL, and IPSYN scores. On all three metrics of grammatical development the 2½-year-old toddlers demonstrated lower performances than their 3½-year-old peers. The largest effect size for age was apparent with the IPSYN scores, indicating that 64% of the variance could be accounted for by age. Table IV provides descriptive information for comparisons with other studies that have examined MLU and IPSYN scores for toddlers learning SAE. As can be seen from Table IV, both age-groups of toddlers performed within the predicted range for MLU and IPSYN scores.

Use of AAE forms

T-tests indicated that the group of 2½-year-olds demonstrated a significantly lower mean dialect density rate of 0.26, while the 3½-year-olds demonstrated a dialect density rate of 0.35 [$t(20) = 3.211$, $p = 0.004$], suggesting that 3½-year-olds are producing more AAE than their younger peers.

A two-part descriptive analysis of the forms produced was conducted to determine if the trends represented by the dialect density measure were maintained at the level of the individual forms. The tokens were separated into two types of AAE forms. Type one included those forms with an explicit SAE obligatory context. Type two was comprised of

Table III. Mean performances of toddlers on the three production measures under investigation; standard deviations are reported in parentheses.

	2½-year-old toddlers	3½-year-old toddlers
MLU-M ^a	2.70 (0.34)	3.08 (0.48)
MSL ^b	3.24 (0.25)	3.90 (0.58)
IPSYN ^c	51.27 (5.16)	70.19 (8.52)
IPSYN Score Range	46–61 points	60–91 points

^aMean length of utterance in morphemes.

^bMean syntactic length.

^cIndex of Productive Syntax Scores for a 75-utterance corpora.

Table IV. MLU and IPSYN scores for toddlers in other studies; standard deviations reported in parentheses.

	2½-year-old toddlers	3½-year-old toddlers
Predicted MLU ^a	2.54 (0.571)	3.78 (0.817)
Predicted MLU Range ^a	1.97–3.11	2.96–4.60
IPSYN Scores ^b	52.73 (10.13)	72.20 (7.23)
Predicted IPSYN Range for 100 utterances	50–62 points	68–85 points

^aMiller & Chapman (1981).

^bScarborough (1990).

Table V. Average number of target forms, obligatory contexts, and percent of use for individual forms with an SAE contrast: standard deviations are reported in parentheses.

	2½-year-old toddlers			3½-year-old toddlers		
	Average number of target forms	Obligatory contexts	Percent of use	Average number of target forms	Obligatory contexts	Percent of use
Variable copula absence	12 (7)	25 (7)	46 (27)	11 (7)	23 (8)	45 (28)
Absent past tense	0.36 (0.8)	1 (1)	13 (31)	3 (4)	5 (3)	55 (41)
Subject/verb agreement	1 (2)	40 (3)	2 (1)	1 (1)	64 (5)	2 (2)
Absent 3rd person	3 (3)	4 (4)	45 (39)	4 (2)	7 (3)	55 (33)
Absent plural	0.55 (1)	3 (2)	22 (41)	0.63 (1)	5 (3)	16 (32)
Absent possessive	0.73 (1)	2 (2)	29 (39)	0.55 (1)	2 (2)	16 (31)

those forms without an explicit obligatory context. Table V provides a summary of descriptive data for each groups' mean production of the 'type one' forms and opportunities for use. Table VI provides frequency data for the 'Type two' AAE forms.

Part one of the descriptive analysis examined the percentage of use in possible contexts for the category one forms listed in Table V. This type of examination revealed two primary findings. First, there were very few differences across groups in the average number of obligatory contexts for each of the individual forms across the groups. Secondly, the average percentage of feature use in contexts differed by age group with the exception of variable copula absence. The use of variable copula absence is virtually the same regardless of age.

Part two of the descriptive analysis examined the frequency data for those features without an explicit SAE obligatory context (see Table VI). Examination of these data indicated that there was very little difference in the production of the *go* copula for both age groups of toddlers. However, the 3½-year-old toddlers demonstrated less use of undifferentiated pronoun case. The 3½-year-old toddlers also demonstrated greater use of overgeneralization than their 2½-year-old peers.

Finally, Pearson correlation coefficients were obtained to determine if there was a relationship between AAE use (as measured by density rate) and the three production measures. As can be seen from Table VII, there was a significant correlation between each of the three measures. However, there were no significant correlations between dialect density and any of the three measures for the toddlers in this study ($p > 0.05$).

Discussion

This study examined two primary issues concerned with the use of language production measures for African-American toddlers who are learning AAE. The first issue was whether there were differences between two age groups of African-American toddlers on three standard measures of language production (MLU, MSL, and IPSYN scores). The second issue concerned characterizing the relationship between MLU, MSL, the IPSYN, and non-standard speech use. Results are discussed in accordance with the primary research questions that guided this study.

The finding of significant differences for measures of language production due to age suggests that the three standard measures are sensitive to age differences in African-American toddlers. The above results concerning MLU is consistent with research comprised of young SAE speakers in which increases in MLU is associated with growth in

Table VI. Frequency counts for 'Type two forms' for toddlers

	2½-year-old toddlers	3½-year-old toddlers
Overgeneralization	13	21
Undifferentiated pronoun case	31	18
<i>Gonna/Ona Gone/On</i> semi-auxiliary	9	6
<i>Go</i> copula	19	18
<i>Sposeta/Bouta</i>	17	12
<i>Finna/Fixin to</i>	4	6
Pronoun extension	4	6
Existential <i>it/they</i>	1	2
Multiple negation	2	2
Indefinite article	0	3
Habitual <i>be</i>	1	0
Total count	101	94

Table VII. Correlation coefficient values for the relationship between dialect density rates (DDR), MLU, MSL, and IPSYN scores.

	MSL	IPSYN	DDR
MLU	0.617 ^a	0.709 ^b	0.196
MSL		0.693 ^c	0.212
IPSYN			0.366

^a*p* = 0.002.^b*p* = 0.000.^c*p* = 0.000.

chronological age until Stage V (Brown, 1973; Miller & Chapman, 1981). Children from AAE backgrounds demonstrate a developmental trend similar to toddlers from SAE backgrounds. However, comparisons between the MLU data for the current sample and the original Miller and Chapman norms revealed that even while the children from the current sample scored within the norms, they still performed below the participants in the original Miller and Chapman sample. In addition, further in-depth examination of the MLU values for each of the 3½-year-old toddlers from the current sample revealed that this older group of toddlers scored at the lower end of the predicted MLU ranges for the Miller and Chapman data (Retherford, 2000). Recall that Dollaghan et al. (1999) found that there was a relationship between low maternal education and low MLU values. The participants in the current study came from a more heterogeneous socio-economic background than the ones in the Miller and Chapman study. SES diversity may account for some of the differences in performance on MLU by the toddlers who participated in this investigation.

Unlike MLU, MSL is a metric that has only been examined in a few studies concerned with language development and its use has not been applied to studies concerned with linguistic development in non-SAE child speakers. The mean MSL value for the 2½-year-old toddlers resulted in what appears to be a 20% increase in utterance length when compared to the mean MLU value. The mean MSL value for the 3½-year-old toddlers was also increased by 25%. These increases are similar to those reported in the previous studies which used this type of metric. Johnston (2001) reported that the calculation of her alternate MLU measure resulted in increases in utterance length per subject that ranged from 3 to

49%. The larger MSL values for the current study indicate that discourse variables (i.e. single word utterances and yes–no acknowledgements) are having an impact on the calculation of traditional utterance length measures (i.e. MLU) for these toddlers from AAE backgrounds.

The findings related to IPSYN scores for each of the age groups indicate that the toddlers from AAE backgrounds are performing quite similarly to SAE child speakers on this task. The mean IPSYN scores for the African-American toddlers in the current study are comparable to the mean IPSYN scores for a 75-utterance corpora in Scarborough's original sample of 15 children at ages 30 and 42 months old (1990). Examination of the range of IPSYN scores obtained from the conversion chart of Scarborough's original report suggest that if a full 100-utterance sample had been used for the analysis set, these African-American toddlers would still have performed similarly to their peers from SAE backgrounds. The IPSYN's sensitivity to differences in age and similar performance scores by these African-American toddlers suggest that the IPSYN can be a useful research tool.

Finally, there was no relationship between dialect density and the three measures of grammatical development. But the evidence provided by the descriptive analysis highlights two significant points. The first point is that the examination of only those features with the SAE obligatory context (listed in Table III) may mask some very important differences. And secondly, toddlers from AAE backgrounds are more likely to maintain or increase their production of certain non-standard speech forms. Kovac and Adamson's (1981) early investigation concerning non-standard dialect use in preschoolers indicated that African-American children from working class backgrounds demonstrate a sharp increase in the variable copula absence between the ages of 3 and 5 years old. Perhaps our findings provide additional evidence that even at the earlier stages of linguistic development toddlers are beginning to use a linguistic system that is consistent with their adult and community models of language. The greater use of AAE speech forms by the older toddlers in this investigation is rarely seen in the language production of toddlers who are learning SAE. A recent investigation concerned with the use of these features in school-age children suggest that there will be a decrease in the use of such forms as children develop their language (Isaacs, 1996). Isaacs failed to find differences in the use of non-standard dialect forms between her sample of school-aged African-American and White children. It is possible that by the time these children reach their school-age years their use of non-standard speech will demonstrate the same trends that appeared in the Isaacs investigation.

Generalization of these findings should be carefully applied, given the small sample size for each of the age groups and the corpus size used for IPSYN scores. While correlation values between smaller and larger corpus sizes were relatively high in the original Scarborough investigation, it is possible that longer samples might have resulted in different results. In addition, while these children represented a range of socio-economic backgrounds, they still came from a community that is remarkably different from the large urban and smaller rural communities where a large number of African-American children in this country reside (Logan, 2001; United States Census Bureau, 2002). Future research should begin to examine the use of MLU, MSL, and IPSYN scores for distinguishing between typically developing and language delayed toddlers who are learning AAE.

Acknowledgments

This project was funded by a Minority Supplement to Grant R01 DC03731 from the National Institutes of Health (NIH), National Institutes on Deafness and Other Communication Disorders and by The University of Wisconsin's Department of

Communicative Disorders Fredric Gruber Award. The authors extend their thanks to the parents, children and daycare staff of the Early Childhood Learning Centers and the Kids First Daycare Center who participated in this project, to Jim Kessenich, Jennifer Mulder, Rhonda Parker, Melanie Combs, and Kristen Cameron for the transcription, analysis, and coding of the language samples, and to Chris Hollar for her assistance throughout the data collection process.

References

- Bayley, N. (1993). *Bayley scales of infant development* (2nd ed.) New York: Cambridge University Press.
- Brown, R. (1973). *A first language: The early stages*. Cambridge, MA: Harvard Press.
- Chapman, R., Schwartz, S., & Kay-Raining Bird, E., 1991, Language skills of children and adolescents with Down Syndrome: comprehension. *Journal of Speech and Hearing Research*, 34, 1106–1120.
- Craig, H., & Washington, J. (1998). Average c-unit lengths in the discourse of African-American children from low-income, urban homes. *Journal of Speech, Language, and Hearing Research*, 41, 433–444.
- Dollaghan, C., Campbell, T., Paradise, J., Feldman, H., Janosky, J., Pitcairn, D., & Kurs-Lasky, M., 1999, Maternal education and measures of early speech and language. *Journal of Speech, Language, and Hearing Research*, 42, 1432–1445.
- Ellis Weismer, S., & Evans, J. (2002). The role of processing limitations in early identification of specific language impairment. *Topics in Language Disorders*, 22, 15–29.
- Hoff-Ginsberg, E. (1991). Mother-child conversation in different social classes and communicative settings. *Child Development*, 62, 782–796.
- Horton-Ikard, R., & Miller, J.F. (2004). It's not just the poor kids, age and the use of AAE forms by African-American children from middle SES communities. *Journal of Communication Disorders*, 37, 467–489.
- Horton-Ikard, R., Ellis Weismer, S., & Evans, J. (2001, February). *AAE usage in mono-racial and biracial parent-toddler dyads*. Poster presented at the 3rd Biennial Texas Research Symposium on Linguistic Diversity, Austin, TX.
- Horton-Ikard, R., & Ellis Weismer, S. (in press). Distinguishing African-American English from developmental errors in the language production of toddlers. *Applied Psycholinguistics*.
- Johnston, J. (2001). An alternate MLU calculation: Magnitude and variability effects. *Journal of Speech, Language, and Hearing Research*, 44, 156–164.
- Joint Committee on Infant Hearing (2000). JCIH year 2000 position statement: Principles and guidelines for early hearing detection and intervention programs. *American Journal of Audiology*, 9, 9–29.
- Klee, T., & Fitzgerald, M. (1985). The relation between grammatical development and mean length of utterance in morphemes. *Journal of Child Language*, 12, 251–269.
- Kovac, C. & Adamson, H.D. (1981). Variation theory and first language acquisition. In D. Sankoff & H. Cedergren (Eds), *Variation omnibus*. Edmonton, Alberta, Linguistic Research, Inc.
- Leonard, L.B., & Weiss, A. (1994). Application of non-standardized assessment procedures to diverse linguistic populations. In K. Butler (Ed.), *Cross-Cultural perspectives in language assessment and intervention*. Gaithersburg, MD: Aspen Publishers, Inc.
- Logan, J. (2001). *Ethnic diversity grows, neighborhood integration is at a standstill*. Lewis Mumford Center for Comparative Urban and Regional Research, State University of New York at Albany, NY.
- Miller, J.F. (1981). *Assessing language production in children*. Austin, TX: Pro Ed.
- Miller, J.F., & Chapman, R. (1981). The relation between age and mean length of utterance in morphemes. *Journal of Speech and Hearing Research*, 24, 154–161.
- Miller, J.F., & Chapman, R. (2000). *Systematic analysis of language transcripts* [computer software]. Madison, WI: Language Analysis Laboratory, Waisman Center.
- Oetting, J. (in press). Promising Language Assessment tools for children who speak a non-mainstream dialect of English. In Ball, M. (Editor). *Clinical Sociolinguistics*. Blackwell Publishing.
- Oetting, J., Cantrell, J., & Horohov, J. (1999). A study of specific language impairment (SLI) in the context of non-standard dialect. *Clinical Linguistics and Phonetics*, 13, 25–44.
- Oetting, J., & McDonald, J. (2001). Nonmainstream dialect use and specific language impairment. *Journal of Speech, Language, and Hearing Research*, 44, 207–223.
- Oetting, J., & McDonald, J. (2002). Methods for characterizing participants' nonmainstream dialect use in studies of specific language impairment. *Journal of Speech, Language, and Hearing Research*, 45, 505–518.
- Rescorla, L. (1993, April). *Outcome of toddlers with specific expressive delay (SELD) at ages 3, 4, 5, 7, & 8*. Poster presented at the Biennial Meeting of the Society for Research in Child Development, New Orleans, LA.

- Rescorla, L., Dahlsgaard, K., & Roberts, J. (2000). Late-talking toddlers: MLU and IPSYN outcomes at 3;0 and 4;0. *Journal of Child Language*, 27, 643–664.
- Retherford, K. (2000). *Guide to analysis of language transcripts third edition*. Eau Claire, WI: Thinking Publications.
- Rickford, J. (1999). *African American vernacular English*. Malden, MA: Blackwell Publishing.
- Rispoli, M., & Hadley, P. (2001). The leading edge: the significance of sentence disruptions in the development of grammar. *Journal of Speech, Language, and Hearing Research*, 44, 1131–1144.
- Scarborough, H.S. (1990). Index of Productive Syntax. *Applied Psycholinguistics*, 11, 1–12.
- Smitherman, G. (1977). *Talkin' and testifyin': The language of Black America*. Boston: Houghton Mifflin (reprinted in 1986 by Wayne State University Press).
- Spitz, R., Tallal, P., Flax, J., & Benasich, A. (1997). Look who's talking: A prospective study of familial transmission of language impairments. *Journal of Speech, Language, and Hearing Research*, 40, 990–1001.
- Stockman, I. (1996). Promises and pitfalls of language sample analysis. *Language Speech and Hearing Services in Schools*, 27, 355–366.
- Thal, D., Reilly, J., Seibert, L., Jefferies, R., & Fenson, J. (2004). Language development in children at risk for language impairment: Cross population comparisons. *Brain & Language*, 88, 167–171.
- United State Census Bureau, 2002, *Current Population and Income Reports*. Washington, DC: US Government Printing Office.
- Washington, J., & Craig, H. (1994). Dialectal forms during discourse of urban, African-American preschoolers living in poverty. *Journal of Speech and Hearing Research*, 37, 816–823.
- Washington, J., & Craig, H. (2002). Morpho-syntactic forms of African-American English used by young children and their caregivers. *Applied Psycholinguistics*, 23, 209–231.
- Wolfram, W., Adger, C., & Christian, D. (1999). *Dialects in schools and communities*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Wolfram, W., & Schilling-Estes, N. (1998). *American English: Dialects and variation*. Oxford, UK: Basil Blackwell.
- Zimmerman, I., Steiner, V., & Pond, R. (1992). *The Preschool Language Scale-3*. Chicago, IL: The Psychological Corporation.