Mean Length of Utterance in Children With Specific Language Impairment and in Younger Control Children Shows Concurrent Validity and Stable and Parallel Growth Trajectories

Mabel L. Rice
University of Kansas, Lawrence

Sean M. Redmond
University of Utah, Salt Lake City

Lesa Hoffman
Pennsylvania State University, State College

Purpose: Although mean length of utterance (MLU) is a useful benchmark in studies of children with specific language impairment (SLI), some empirical and interpretive issues are unresolved. The authors report on 2 studies examining, respectively, the concurrent validity and temporal stability of MLU equivalency between children with SLI and typically developing children.

Method: Study 1 used 124 archival conversational samples consisting of 39 children with SLI (age 5;0 [years;months]), 40 MLU-equivalent typically developing children (age 3;0), and 45 age-equivalent controls. Concurrent validity of MLU matches was examined by considering the correspondence between MLU and developmental sentence scoring (DSS), index of productive syntax (IPSyn), and MLU in words. Study 2 used 205 archival conversational samples, representing 5 years of longitudinal data collected on 20 children with SLI (from age 5;0) and 18 MLU matches (from age 3;0). Evaluation of growth dimensions within and across groups was carried out via growth-curve modeling.

Results: In Study 1, high levels of correlation among the MLU, DSS, and IPSyn measures were observed. Differences between groups were not significant. In Study 2, temporal stability of MLU matches was robust over a 5 year period.

Conclusions: MLU appears to be a reliable and valid index of general language development and an appropriate grouping variable from age 3 to 10. The developmental stability of MLU matches is indicative of shared underlying growth mechanisms.

KEY WORDS: specific language impairment, mean length of utterance, vocabulary development, growth curves

Since Margaret Morse Nice (1925) suggested that "average sentence length may well prove to be the most important single criterion for judging a child's progress in the attainment of adult language" (p. 378), measures of children's utterance length have been widely used as a benchmark of linguistic maturation (e.g., Barrie-Blackley, Musselwhite, & Rogister, 1978; Bennett-Kastor, 1988; Brown, 1973; de Villiers & de Villiers, 1973; Leadholm & Miller, 1992; McCarthy, 1930; Miller, 1981; Miller & Chapman, 1981; Shriner, 1969; Templin, 1957; Tyack & Gottsleben, 1974). Mean length of utterance in morphemes (MLU) played a prominent role in Roger Brown's influential stage model of grammatical development and in subsequent clinical adaptations of this framework. Over the last three decades, the properties of MLU have been explored in its capacity as both an independent and dependent variable.

A common design in studies of children with specific language impairment (SLI) is to compare affected children's performance to that of a control group of typically developing children of similar chronological age (age controls) and to that of another control group of typically developing children who display similar levels of MLU (MLU controls). The logic behind the use of dual control groups is that observed linguistic deficiencies in the SLI group's performance relative to age expectations can then be compared to immature, but typically developing, linguistic systems. Any similarities between the SLI group and their MLU controls in terms of overall proficiency or error typology on a particular language task are then considered to be evidence of delays within the SLI group. On the other hand, observed differences between the SLI group and the MLU controls would represent an unexpected developmental disruption or a qualitative deviation from the course of typical linguistic maturation.

MLU-based comparisons appear to be particularly important for studies of morphosyntactic and syntactic development. Opportunities for children to produce key grammatical forms such as inflectional affixes, relative clauses, and question forms would logically be influenced by the extent to which children produce utterances that are sufficiently long enough to support such structures. Thus, utterance length represents an important confound that should be addressed experimentally—and MLU equivalency is one way to accomplish this.

The dual comparison design has been highly successful in producing a consistent body of evidence that suggests both delays and disruptions are present within the linguistic symptoms associated with SLI. In particular, a profile of selective morphosyntactic disruptions within the context of coexisting general delays in other areas of language represents the desiderata of current theories (Chiat, 2000; Leonard, 2004; Rice, 2003; van der Lely, 1998). The extent to which children with SLI follow a delay or disruption pattern is highly relevant to the evaluation of etiological models and the possible role of genetic or environmental factors in language impairment.

Although the use of MLU levels as a reference for the developmental status of affected versus unaffected children has great potential interpretive value, the literature yields mixed perspectives regarding measurement and interpretation, as well as important empirical gaps. The two studies reported here addressed these issues in complementary research designs: The first study focused on measurement robustness within a dual comparison, cross-sectional design, comparing MLU to concurrent language measures. The second study is the first report of longitudinal measurements of affected and unaffected children in the age range of 3 to 10 years, comparing the growth of MLU across groups and across linguistic dimensions.

Study 1 was motivated by important challenges to the use of MLU as a control variable (Eisenberg et al., 2001; Johnston, 2001; Johnston & Kamhi, 1984; Lahey et al., 1992; Plante et al., 1993; Rollins, Snow, & Willett, 1996). Implementation issues have been noted. For example, cautions have been expressed regarding the integrity of available normative samples, the effect of different conversational parameters on representativeness, and the reliability of MLU calculated on samples of insufficient lengths. Other concerns have focused on interpretative issues associated with MLU. Does MLU measure what it was designed to measure? For example, if MLU benchmarks children's progression through "general levels of language development," then it should be highly correlated with other indices of growth, such as chronological age and overall vocabulary level. Furthermore, as a purported measure of grammatical development, it should also be consistent with other general grammatical indices. Two indices have been evaluated: developmental sentence scoring (DSS; Lee, 1974), which is a technique to score eight categories of grammatical forms in a sample of 50 sentences, and the index of productive syntax (IPSyn; Scarborough, 1990), which is an estimate of emerging grammatical complexity in a sample of 100 spontaneous utterances in the categories of noun phrases, verb phrases, questions and negatives, and sentence structures. Alternative measures of average utterance length, such as utterance length in words, should also show strong associations. Some studies have failed to find such associations (Chan, McAllister, & Wilson, 1998; Klee & Fitzgerald, 1985; Scarborough, Wyckoff, & Davidson, 1986; Scarborough, Rescorla, Tager-Flusberg, Fowler, & Sudhalter, 1991).

Furthermore, the logic of using dual control groups based on any index of "general language level" has been challenged. Plante, Swisher, Kiernan, and Restrepo (1993) pointed out that the notion of a "language match" is inherently misleading because by its nature the equivalency between children with SLI and typically developing children will introduce extraneous age effects. Children with SLI will inevitably be older than their MLU controls and will thus bring different levels of experience and world knowledge, attention, motivation, and social development to language tasks. Plante et al. argued further that since language represents a multidimensional construct, it is possible that superficially similar MLU values from different children could reflect very different levels of actual language development (see Johnston & Kamhi, 1984; Lahey et al., 1992; and Rollins, Snow, & Willett, 1996, for similar arguments). Plante et al. (1993) suggested that longitudinal measures of MLU and other language variables would be helpful in sorting out these issues.

Study 2 addressed the need for longitudinal data. There have been surprisingly few studies with a developmental view of MLU in children with SLI, that is, that have directly examined the relationship of age with MLU in affected and unaffected children. Klee, Schaffer, May, Membrino, and Mougey (1989) documented that children with SLI have lower MLU levels than their age peers during the 24–50 month age range. Scarborough, Wyckoff, and Davidson (1986) obtained longitudinal samples from 12 typically developing children from ages 24 to 60 months to evaluate MLU and age relationships. Scarborough, Rescorla, Tager-Flusberg, Fowler, and Sudhalter (1991) followed up with longitudinal analysis of 20 preschoolers with language delay. Their samples included children with fragile X syndrome, Down syndrome, or autism, but none with SLI.

Brief Review of Implementation Issues

The manner in which conversational samples are collected from children with SLI and typically developing children could compromise the integrity of using MLU as a control measure. Johnston et al. (1993) examined the effects of an interview collection format on 24 children's productions (age range: 2;6–7;8 [years;months]). These investigators found a significant inverse relationship between the proportion of questions used by the adult examiner and children's MLU values. More importantly, the dampening effect of examiner questions on MLU was particularly apparent in the samples collected from children with SLI. Thirty-six percent of the utterances produced by these children represented brief, elliptical responses to adult questions compared to 24% of the typically developing children's utterances. Johnston et al. suggested that because of an awareness of their language limitations, children with SLI may respond differently to adult questioning than typically developing children. In particular, children with SLI may perceive adult questioning as challenging or threatening and may be more reluctant to engage during interview interactions. Thus, MLU measures based on samples with high levels of adult questioning might consistently underestimate the grammatical competence of children with SLI. Under these conditions, MLU equivalency between groups of affected and unaffected children would become problematic.

The integrity of any language level control made between children with SLI and typically developing children also depends on the reliability of the measure used. Concerns have been raised regarding the appropriate sample size needed for MLU controls. Gavin and Giles (1996) examined the temporal stability of MLU at different sample sizes. These investigators collected two 20-min language samples from a group of 20 typically developing preschool children (age range: 2;7–3;10). Samples were edited to different sizes, based on the total number of utterances (25–175, in increments of 25). Acceptable levels of temporal reliability (i.e., $r > .90$) were only observed for samples containing 175 utterances. In contrast, experimental studies matching children with SLI to younger typically developing children have often been based on samples containing a significantly smaller number of utterances (50–100 utterances), for which Gavin and Giles reported much lower levels of reliability (rs = .61–.82). Thus, equating groups of children on the basis of sample sizes smaller than 175 utterances probably compromises the integrity of the match.

Brief Review of Interpretative Issues and Developmental Evidence

Significant correlations between MLU and other indices of growth (e.g., age, vocabulary level, DSS, IPSyn, utterance length measured in words) are needed to support the use of MLU as a general index of grammatical development. Brown (1973) described MLU as

> an excellent simple index of grammatical development because almost every new kind of knowledge increases length: the number of semantic roles expressed in a sentence, the addition of obligatory morphemes, coding modulations of meaning, the addition of negative forms and auxiliaries used in interrogative and negative modalities, and, of course, embedding and coordinating (pp. 53–54).

However, later in the same discussion he suggested that as a measure of grammatical development MLU probably loses its integrity beyond 4.0 morphemes. He suggested that at this point children would have control over such diverse sentence structures that MLU would probably depend more on the nature of the interaction rather than on the limits of children's grammatical knowledge. Empirical reports on the linearity of the age–MLU relationship have been somewhat mixed on this issue. Miller and Chapman (1981) pooled data from a sample of 123 typically developing children from age 1;5 to 4;11 and found a significant correlation (r = .88) between age and MLU. A linear relationship was also reported. MLU values were based on "a minimum of 50 intelligible utterances" (Miller & Chapman, 1981, p. 24). Other studies have reported similar levels of correlation and linearity (rs = .70–.76) for children within this age range (Blake, Quoataro, & Onorati, 1993; de Villiers & de Villiers, 1973; Rondal, Ghiotti, Bredart, & Bachelet, 1987).

However, Klee and Fitzgerald (1985) presented conflicting evidence based on a sample of 18 two- to four-year-olds. These investigators found a considerably lower correlation between age and MLU (r = .26). MLU values were based on 100 consecutive, intelligible utterances. Chan, McAllister, and Wilson (1998) found similar results for 75 two- to three-year-olds based on 50 consecutive, intelligible utterances (rs = .22–.37). Conant (1987) reanalyzed the Klee and Fitzgerald data and found that when the data for the 3-year-olds was analyzed separately there was a moderate correlation with age (r = .75). Scarborough, Wyckoff, and Davidson (1986) observed in both cross-sectional and longitudinal samples of 2–5-year-olds that a linear relationship existed only for children up to 4 years, beyond which a nonlinear relationship was apparent, supporting Brown's (1973) initial concerns. In this case, MLUs were based on samples of 100 intelligible utterances.

Klee et al. (1989) compared the relationship between MLU and age in language samples collected on preschool children with SLI to samples collected on age-matched typically developing children (age range: 2;0–4;2). MLU values were based on an average of 156 complete and intelligible utterances (range: 65–242 utterances). The results showed that MLU values were significantly and moderately correlated with age in both groups (rs = .75 and .77, respectively). Furthermore, Klee et al. were the first to demonstrate that the affected children's MLU levels were lower than those of their age peers, a generalization that holds for Cantonese-speaking children as well (Klee, Stokes, Wong, Fletcher, & Gavin, 2004).

Limited information exists on the MLU–age relationship for school-age children. Chabon, Kent-Udolf, and Egolf (1982) used a picture description task to investigate the temporal stability of MLU values collected on three groups of typically developing children (age range: 3;6–9;6) by having children repeat the task over 3 consecutive days. These authors found significantly lower levels of stability within the oldest group (3;6–4;6, r = .68; 5;6–6;6, r = .70; 8;6–9;6, r = .40). It is difficult to integrate these findings based on children's picture descriptions with the rest of the literature because most studies on MLU have been based on children's conversational productions.

In addition to age, a measure of general grammatical development should correspond well with other indices of grammatical growth. There is some evidence that relative to other measures, MLU may not be as sensitive to developmental differences between children with and without SLI. Scarborough, Rescorla, Tager-Flusberg, Fowler, and Sudhalter (1991) examined the relationship between MLU and IPSyn measures within samples collected on preschool children with and without language delays (although their samples did not include children with SLI). Results indicated that regression curves were similar for the two groups of children, especially when MLU values were below 3.0. However, for some older children with language delays, the authors suggested that MLU significantly overestimated their actual syntactic production.

Johnston and Kamhi (1984) reached similar conclusions when they compared 10 children with SLI to 10 MLU-matched typically developing controls (MLU range: 4.17–5.48). These investigators found that although the language samples were matched on the basis of MLU, there were significant differences between groups on several syntactic and semantic measures. Measures from the DSS procedure were used. Some of these differences favored the MLU matches (number of propositions per utterance, number of correctly inflected verbs), whereas others favored the SLI group (number of progressive sentences and two-argument predicates). Johnston and Kamhi characterized the observed group differences and apparent trade-off effects as an example of how the "same can be less" within language matched comparisons between children with and without language impairments. In this case, MLU calculations based on language samples of children with SLI appeared to be superficially similar to language samples of typically developing children but were in fact qualitatively different.

In a recent re-evaluation of the "same can be less" issue, Leonard and Finneran (2003) examined the effect that reduced rates of particular grammatical morphemes—in this case, finite verbal forms—could have on the types of sentences children produce. These investigators provided a mathematical and data based demonstration that affected and unaffected children who are equivalent on MLU but who also show significantly different rates of grammatical morpheme use would not necessarily have to have qualitatively different sentence structures or overall language levels. In other words, because of the modest impact that finite verbal forms appeared to have on MLU calculations, Leonard and Finneran argued it was not necessary to invoke differential trade-off effects such that children with SLI would need to compensate in other areas of language in order to achieve comparable MLU levels. Similarly, in their study of child speakers of Southern American English, with and without SLI, Oetting, Cantrell, and Horohov (1999) showed that MLU calculations were not significantly altered by the presence of a nonstandard dialect that allows omissions of grammatical morphemes. In this study, obtained MLU values for the child speakers of Southern American English without SLI were close to the normative values available for child speakers of Standard American English (i.e., Leadholm & Miller, 1992).

In summary, investigators have expressed various concerns regarding the appropriateness of equating children with SLI to younger typically developing children on the basis of MLU. Some of these concerns can be addressed by maintaining high levels of experimental control of the sampling contexts and collection procedures. Expanding available normative samples would also improve the diagnostic integrity of MLU. These issues are important and have likely had an impact on our understanding of MLU and its value as an index of general language development. For example, some of the inconsistencies reported across studies could be attributed to different sample sizes or differences in the amount of adult questioning that took place.

However, more pointed concerns have focused on the possibility that MLU values for children with SLI are based on different dimensions of development than MLU values for typically developing children. In contrast, Rice (2004a, 2004b) argued that the morphosyntactic development of children with SLI and younger unaffected children shows striking parallels, which in turn suggest adherence to the same underlying growth mechanisms. The argument hinges on similarities in grammatical acquisition and avoidance of grammatical errors. An important exception to this generalization is the acquisition of grammatical tense marking, which is out of synchrony with the growth of general clausal mechanisms (as indexed by MLU) in affected children. Yet grammatical tense marking follows similar growth trajectories in affected children as in younger children, but the growth is offset in children with SLI, such that they start late and are moving through the patterns of growth at older ages than unaffected children.

A focus on growth brings an interest in possible predictors. Two variables have been of interest in previous studies: children's nonverbal intelligence and their mother's education. Although each of these variables has been identified as predictive of children's language acquisition in the broader literature (cf. Dollaghan et al., 1999; Tomblin, 1996), a consistent finding has been that growth in grammatical tense marking is not predicted by children's nonverbal intelligence or their mother's education (Rice, Tomblin, Hoffman, Richman, & Marquis, 2004; Rice, Wexler, & Hershberger, 1998; Rice, Wexler, Marquis, & Hershberger, 2000).

In the context of general growth mechanisms, it is important to clarify the stability of growth in MLU during the period of basic morphosyntactic acquisition, and to evaluate whether mother's education or children's nonverbal intelligence are predictors. A valuable way to investigate similarity of growth mechanisms across groups is to determine if growth follows similar trajectories for multiple dimensions of language. Here we evaluate the extent to which MLU and vocabulary acquisition follow similar trajectories in development for MLU equivalent groups of affected and unaffected children. In effect, this constitutes a tough test of the validity of MLU as a grouping variable, in that it evaluates the extent to which MLU in affected as well as unaffected children shows stable and parallel growth, and, furthermore, grows in tandem with other dimensions of language in the expected way.

Questions Directing the Current Analyses

The concurrent validity of MLU matching warrants further examination. A longitudinal investigation of MLU growth in children with SLI and typically developing children is also needed that extends beyond the preschool/kindergarten time frame and employs conversationally felicitous interactions with language samples of sufficient size. Under these conditions, the integrity of MLU as a control variable over the course of grammatical development could be evaluated. If MLU estimates for children with and without SLI are qualitatively different and based on different dimensions of development, then we should expect different patterns of associations between MLU and other general language measures. We also need to check on the association of MLU in morphemes with MLU in words. Although a high level of association is expected, it is not always found. In the case of children with SLI, who show morphological differences from unaffected children, it is especially relevant to examine if a morphological index of growth shows a different pattern of association than a word-level index of growth in utterance length. If MLU estimates for children with and without SLI are qualitatively different, then different growth curves for MLU should be observed between the two groups of children because there would be no reason to predict that spurious initial MLU matches would remain stable over time. If MLU, however, taps into the same underlying growth mechanisms in children with and without SLI, then initial matches should be stable. Growth curves and observed MLU associations for the two groups of children should likewise be similar, as well as predictor relationships.

Study 1 examined the concurrent validity of MLU equivalency in a cross-sectional study of 5-year-old children with SLI and 3-year-old children with typical development. The specific questions for Study 1 were as follows:

1. Are there significant group differences between age-equivalent and MLU-equivalent groups on measures derived from the DSS analysis?
2. Are there significant group differences between age-equivalent and MLU-equivalent groups on measures derived from the IPSyn?
3. Are there significant group differences between age-equivalent and MLU-equivalent groups on average utterance length in morphemes?
4. Is MLU associated with age, vocabulary level, DSS, and IPSyn scores?

Study 2 is a longitudinal study that examined the temporal stability of initial MLU equivalency over a 5-year period and compared the growth of MLU to growth in vocabulary levels. The specific questions for Study 2 were as follows:

1. Are there significant differences between initial MLU equivalent groups in the observed rates of growth in MLU?
2. Are there significant differences between groups in the observed rates of growth in vocabulary levels?

3. Does mother's education or children's nonverbal intelligence predict growth in MLU, and, if so, does the prediction vary by group?

Study 1: Cross-Sectional at Ages 5 and 3 Years

Method

Participants. The 124 conversational samples used to examine the concurrent validity of MLU matches were taken from Rice and Wexler's (1996) investigation of children's command of grammatical tense marking. Participant characteristics are summarized in Table 1. These data include three groups of children, combining the sample of Rice, Cleave, and Oetting (2000) with an additional sample recruited for the longitudinal study of Rice, Wexler, and Hershberger (1998): 39 children with SLI, with a mean chronological age of 58 months (range = 52–68 months), a control group of 40 younger typically developing children (M = 36 months; range = 30–44 months) who were equivalent to the SLI group on the basis of MLU (each child in the MLU group had an MLU within 0.10 morphemes of at least 1 child in the SLI group), and a control group of 45 same-age typically developing children (M = 60 months; range = 52–67 months). DSS and IPSyn data are available for 10 of the age control children primarily to confirm that the other two groups did differ from age expectations on these measures.

Table 1. Participant profiles for the cross-sectional sample in Study 1 (N = 124)[a]: Group means and (standard deviations).

Variable	SLI	MLU equivalent	Age equivalent
Sample size (males)	39 (24)	40 (19)	45 (24)
Age in months	58.23 (4.22)	35.55 (3.49)	59.91 (4.12)
CMMS age deviation score	96.11 (10.14)	109.74 (9.20)	115.10 (14.18)
MLU in morphemes	3.49 (0.56)	3.58 (0.49)	4.51 (0.83)
MLU in words	3.23 (0.48)	3.25 (0.41)	4.06 (0.79)
TOLD–2:P spoken language quotient	76.05 (6.67)		112.39 (9.43)
PPVT–R raw score	32.10 (7.19)	26.20 (8.65)	64.60 (10.62)
DSS utterance score	4.94 (1.55)	5.23 (1.54)	7.89 (1.90)
IPSyn	69.95 (8.38)	70.93 (8.43)	84.30 (9.39)

Note. CMMS = Columbia Mental Maturity Scale; MLU = mean length of utterance; TOLD–2:P = Test of Language Development—Second Edition: Primary; PPVT–R = Peabody Picture Vocabulary Test—Revised; DSS = developmental sentence scoring; IPSyn = index of productive syntax.

[a] The number of participants per dependent variable varied somewhat: Complete (124) for age, MLUm, and MLUw; because of age floors, the CMMS was available for only 35 SLI, 19 MLU-equivalent, and 21 age-equivalent children; because of protocol differences across studies, the TOLD–2:P was available for only 21 SLI and 23 age-equivalent children; the DSS and IPSyn data were available for 37 SLI, 40 MLU-equivalent, and 10 age-equivalent children.

All of the participants with SLI had receptive and expressive language impairment, without severe speech impairment or "limited intelligibility" and met the following criteria: (a) previously diagnosed as having a language impairment by a certified speech-language pathologist; (b) receptive language performance on the Peabody Picture Vocabulary Test—Revised (PPVT–R; Dunn & Dunn, 1981) 1 *SD* or more below the mean; (c) expressive language performance 1 *SD* or more below age expectations as measured by MLU from a sample of at least 150 utterances (Leadholm & Miller, 1992); (d) normal or above normal intellectual functioning (above 85 standard score) as measured by the Columbia Mental Maturity Scale (CMMS; Burgemeister, Blum, & Lorge, 1972); (e) a passing score on a probe screening for articulation competency with consistent use of final *-t, -d, -s,* and *-z,* and only minor mispronunciation on the Goldman Fristoe Test of Articulation (GFTA; Goldman & Fristoe, 1986); and (f) normal hearing levels as measured by a hearing screening at 25 dB HL at 1000, 2000, and 4000 Hz. In addition, the Test of Language Development—Second Edition: Primary (TOLD–2:P; Newcomer & Hammill, 1988) was administered to the participants in Rice, Wexler, and Hershberger (1998). The group mean of the SLI group (N = 21) on the spoken language quotient (standardization sample: M = 100, SD = 15) was 76; standard deviation was 7; for the age control group (N = 23), M = 112, SD = 9. TOLD–2:P scores are not available for the participants recruited for Rice, Cleave, and Oetting (2000) because the focus of the study was lexical acquisition (hence the PPVT–R as a common variable) and the children in the SLI group were recruited from clinicians' caseloads, where they shared the criterion of performance 1 *SD* or below on a standardized omnibus language test but the tests varied across clinical settings. Rice and Wexler (1996) reported that the SLI samples for the two studies were equivalent in levels of morpheme use, and for grammatical tense marking these levels were significantly lower than those of the MLU group.

Children in the control groups were recruited from the same school and preschool attendance centers as the affected children. Control children had to meet the following criteria: (a) identified as typically developing by teachers and parents; (b) receptive language skills within normal limits or above (i.e., standard score of 85 or above) as measured by the PPVT–R; (c) expressive language skills within normal limits or above (i.e., 1 *SD* below the age level mean as minimal level of performance) as measured by MLU (Leadholm & Miller, 1992); (d) normal intellectual functioning (i.e., standard score of 85 or above) as measured by the CMMS, normal articulation (percentile of 15 or above) as measured by the GFTA, and normal hearing as measured by a hearing screening.

Conversational sampling procedures. The language samples were collected using a standard set of age-appropriate toys selected to elicit a variety of grammatical forms and sentence types, consisting of toy people, house/garage and furniture, and toy animals. The samples were audio-recorded adult–child conversational interactions using a dual-microphone set-up, to maximize audio clarity. The aim was for a minimum of 200 child utterances (*M*s = 253.54 for the SLI group, 211.36 for the age control group, and 221.08 for MLU control group), which usually required about 30 min of interaction, although more time could be required for affected children. The samples were subsequently transcribed and coded for grammatical morphemes following the conventions of the Kansas Language Transcript Database (KLTD; Howe, 1996). Utterance segmentation followed Miller (1981, p. 14), that is, terminal intonation contour, pauses of 2–3 s. In addition, although these utterances are rare in the samples, utterances comprising more than two independent clauses conjoined by *and* were broken preceding the second conjunction, in order to avoid spurious lengthening due to clausal chaining.[1]

Graduate research assistants with clinical experience collecting language samples from young children were trained to follow "best practice guidelines" regarding sample collection. This included following the children's conversational lead, engaging in parallel talk and parallel play focusing on everyday event schemas such as household activities, sharing personal anecdotes and experiences, and introducing topics related to past and ongoing events during their conversational interactions. Research assistants were also trained to keep the use of "yes"–"no" and *Wh-* questions to a minimum. A monitoring system was in place such that each sample was checked by a project supervisor for sufficient number of utterances, conversational content, and adherence to collection procedures. Samples that did not meet these criteria were rejected and an additional sample was collected from the same child–adult dyad within a couple of weeks of the initial sample.

Dependent measures. The Systematic Analysis of Language Transcripts program (SALT; Miller & Chapman, 1991) was used to generate values for MLU and MLU in words. DSS and IPSyn measures were coded and calculated by hand by research assistants.

Reliability. Conversational sample transcription and analysis followed a written protocol (cf. Howe, 1996) representing different steps and checking passes to which individual transcribers were trained to 85% agreement or better with trained transcribers prior to carrying out transcription assignments. All transcripts were checked by second and third transcribers for possible errors; any detected disagreements were resolved through consensus agreement. Rice and Wexler's (1996) original estimations of transcription reliability associated with the data set were based on pairwise comparisons across two coders over 11 transcripts and yielded an overall level of agreement of 96% (range = 89%–98%).

Levels of agreement for DSS scoring were based on pairwise comparisons between two coders over 10 transcripts. The overall correlation between coders was .86. The average proportion of utterances with identical scores was .78 (range = .56–.88). A percentage of agreement value was calculated by adding up the total DSS for both transcripts and then dividing by the larger number. This calculation yielded an overall level of agreement of 92% (range = 77%–100%).

Agreement levels for IPSyn scoring were based on pairwise comparisons between two coders over 10 transcripts. A point-by-point comparison yielded an overall level of agreement of 89% (range = 83%–96%).

Results and Discussion

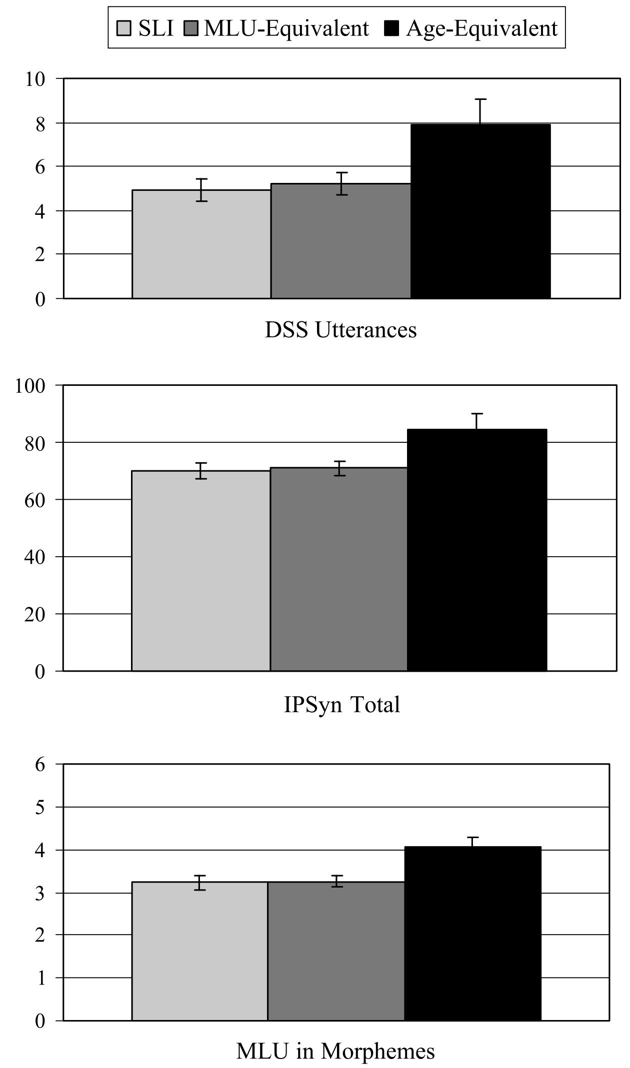
Potential group differences in MLU in morphemes, DSS, and IPSyn between the SLI group and the typically developing control groups (MLU equivalent and age equivalent) were examined with independent analyses of variance (ANOVAs). Homogeneity of variance was obtained unless otherwise noted, and post hoc comparisons were conducted with Tukey's honestly significant difference. Means and error bars are provided for each outcome and each group in Figure 1; means and standard deviations are provided in Table 1. Although the group comparisons of most interest are the SLI and MLU-equivalent groups, the age-equivalent group is also included to substantiate that the other two groups performed at lower levels on the target variables. We predicted that there would be overall group differences but that these differences would be age equivalent > SLI = MLU equivalent. Each of the three research questions is addressed in turn.

Question 1: Are There Significant Group Differences Between Age-Equivalent and MLU-Equivalent Groups on Measures Derived From the DSS Analysis?

As expected, there were significant differences across the three groups in DSS utterances, $F(2, 84) = 14.13$, $MSE = 2.51$, $p < .001$, $\eta^2 = .25$, and DSS total scores, $F(2, 84) = 14.10$, $MSE = 6307.63$, $p < .001$, $\eta^2 = .25$. For

[1]The criteria for utterance segmentation included the following: terminal intonation contour, pauses of 2–3 s, a limit of two independent clauses conjoined by *and* in one utterance, and clauses joined by subordinating conjunctions (such as *after, before, but, if, when*) included in a single utterance. Between-examiner reliability estimates for utterance boundaries following these conventions at each round of data collection ranged from 87% to 93%, with only one time below 90%.

Figure 1. Group means and two standard errors for developmental sentence scoring (DSS), index of productive syntax (IPSyn), and mean length of utterance (MLU) in morphemes.



both outcomes, the age-equivalent group had higher scores than the SLI or MLU-equivalent groups ($ps < .001$), which did not differ from each other.

Question 2: Are There Significant Group Differences Between Age-Equivalent and MLU-Equivalent Groups on Measures Derived From the IPSyn?

As expected, there were significant differences across the three groups in IPSyn scores, $F(2, 84) = 11.83$, $MSE = 72.51$, $p < .001$, $\eta^2 = .22$, such that the age-equivalent group had higher scores than the SLI or MLU-equivalent groups ($ps < .001$), which did not differ from each other.

Question 3: Are There Significant Group Differences Between Age-Equivalent and MLU-Equivalent Groups on Average Utterance Length in Morphemes?

Because homogeneity of variance did not hold for the age-equivalent group, Browne-Forsythe tests were used to assess the significance of mean differences. There were significant differences across the three groups in average utterance length in morphemes, $F^*(2, 105.78) = 33.67$, $p < .001$, such that the age-equivalent group had higher scores than the SLI or MLU-equivalent groups ($ps < .001$), which did not differ from each other. An identical pattern of group differences was found for average utterance length in words as well.

Question 4: Is MLU Associated With Age, Vocabulary Level, DSS, and IPSyn Scores?

Correlations and their standard errors within each of the SLI and MLU-equivalent groups are given in Table 2. It is clear that the expected associations among the variables were uniformly strong for the younger control group but not in the SLI group. For example, the age-expected correlation with MLU was not significant in the SLI group ($r = .11$ for SLI; $r = .51$ for MLU equivalent). It could be thought that the lack of correlation for the SLI group was attributable to a relatively narrow age range, but the age range was no more restricted for the SLI group than for the MLU group, where a correlation was evident. The difference in the magnitude of the age–MLU correlation was examined with a Group × Age interaction term in a linear regression, and the difference was found to be significant at the .06 level (estimated power of .44 to detect an increase in $R^2 = .042$ for $N = 79$). Similarly, the expected association of MLU with PPVT–R raw scores was not significant for the SLI group ($r = .08$ for SLI; $r = .43$ for MLU equivalent), although a Group × PPVT–R raw score interaction term in a linear regression was not significant (estimated power of .21 to detect an increase in $R^2 = .018$ for $N = 79$).

Overall, the results of Study 1 suggest that children in the SLI and the MLU-equivalent groups were producing utterances that were very similar with regard to the number of words produced per utterance and grammatical complexity, as indexed by DSS and IPSyn. Further, the associations among utterance length and complexity were strong in the affected group as well as for the controls. The normative association of age and lexical growth with utterance length was not evident in the affected group, suggesting that the mechanisms for growth in

Table 2. Observed pearson correlations within SLI and MLU-equivalent groups in Study 1.

SLI correlations[a]	1	2	3	4	5	6
1. MLU in morphemes	1.00					
2. MLU in words	.98*	1.00				
3. Age in months	.11	.19	1.00			
4. DSS utterance score	.56*	.61*	.38	1.00		
5. IPSyn total score	.70*	.71*	.27	.81*	1.00	
6. PPVT raw score	.08	.11	.60*	.39	.28	1.00
MLU-equivalent correlations[a]	**1**	**2**	**3**	**4**	**5**	**6**
1. MLU in morphemes	1.00					
2. MLU in words	.98*	1.00				
3. Age in months	.51*	.49*	1.00			
4. DSS utterance score	.70*	.67*	.45*	1.00		
5. IPSyn total score	.80*	.76*	.62*	.75*	1.00	
6. PPVT raw score	.43*	.48*	.41*	.41*	.36	1.00

[a] The standard error of the correlations was calculated as $1/\sqrt{n-3}$; for SLI = .17, and for MLU equivalent = .16. Spearman nonparametric correlations were largely similar.

*$p < .01$.

utterance length are dissociated with age and with vocabulary growth for affected children. This conclusion differs from that of Klee et al. (1989), who reported a strong association of MLU and age in a sample of affected children ages 2;0–4;2, with an MLU range of 1.5–3.69. It may be that the age–MLU relationship is more robust for the earlier stages of MLU growth in affected children than in the later stages studied here. For the indices of clausal complexity, MLU was positively associated with growth in complexity, suggesting that for the affected children as for the younger children MLU is a valid index of clausal complexity, even when it becomes disconnected from age and growth in receptive vocabulary. These outcomes are highly supportive of the assumption that the concurrent validity is robust for matching children with SLI to younger typically developing children on the basis of MLU, given careful attention to best practice for sample elicitation and coding.

Study 2: Longitudinal Study

Method

Participants. The participants were the SLI group and the MLU-equivalent group of Rice, Wexler, and Hershberger's (1998) longitudinal study of morphosyntactic development: 20 five-year-olds with SLI and 18 children in the MLU-equivalent group. Note that this is a subset of the participants in Study 1. Ethnicity/race data are available for the longitudinal study participants only, of whom 1 control child was African American, 2 affected children were Hispanic, and the remaining participants were Caucasian. Every 6 months thereafter for the next 4 years (and 1 whole year later), additional language samples were collected from both groups of children, for a total of 5 years of data.

Dependent measures. A total of 205 conversational samples constituted the material for calculation of MLU over time. The same collection, transcription, and monitoring procedures outlined for the cross-sectional study were used throughout the duration of the longitudinal study. The SALT program was used to generate values for MLU. A total of up to 9 data points were available for each participant. Raw scores from the PPVT–R were used to estimate children's growth in receptive vocabulary. A total of up to 4 annual data points were available for each participant. The levels of mother's education at the outset of the study were measured by a scale where 1 = *some high school* and 5 = *some graduate work*. The mean for the SLI group was 2.5. They were distributed across the entire range; 16 were high school graduates or above. The mean for the MLU-equivalent group was 4.5. All were high school graduates or above.

Reliability. The same safeguards and procedures of the cross-sectional study were followed in the longitudinal study. Interrater agreement, assessed over 10 rounds of data collection, was calculated at 90% or better for utterance boundaries, morpheme transcription, morpheme coding, and morpheme counting. Pairwise monitoring across six different transcribers was also found to reach consistently high levels of agreement of 85% or better.

Results and Discussion

The main questions of interest were as follows:

1. Are there significant differences between initial MLU-equivalent groups in the observed rates of growth in MLU?
2. Are there significant differences between groups in the observed rates of growth in vocabulary levels?
3. Does mother's education or children's nonverbal intelligence predict growth in MLU, and, if so, does the prediction vary by group?

Preliminary analyses examined mean differences per time of measurement per group for MLU in morphemes and PPVT–R raw scores using the LSMEANS procedure in SAS PROC MIXED. As shown in Table 3, the two groups remained at equivalent levels of MLU at each time of measurement; there were no significant differences across the nine times of measurement. This

Table 3. Differences for mean length of utterance (morphemes) and PPVT–R raw scores: Means (and standard errors) for the longitudinal sample by group and time of measurement.

Years in study	MLU in morphemes				PPVT–R raw scores			
	SLI	MLU equivalent	*p*	Cohen's *d*[a]	SLI	MLU equivalent	*p*	Cohen's *d*[a]
0.0	3.67 (0.15)	3.75 (0.15)	.71	−.12	32.40 (1.90)	24.89 (1.84)	.01	.92
0.5	3.97 (0.15)	4.21 (0.16)	.28	−.36				
1.0	4.16 (0.14)	4.17 (0.14)	.97	−.02	53.65 (2.51)	50.06 (2.64)	.33	.32
1.5	4.46 (0.13)	4.52 (0.14)	.76	−.09				
2.0	4.76 (0.15)	4.60 (0.16)	.46	.24	69.05 (2.37)	70.44 (2.50)	.69	−.13
2.5	4.74 (0.14)	4.88 (0.15)	.52	−.22				
3.0	4.85 (0.16)	4.86 (0.16)	.95	−.01	79.75 (1.99)	88.61 (2.09)	.01	−1.00
4.0	4.92 (0.17)	4.84 (0.18)	.74	.11				
5.0	5.32 (0.17)	5.04 (0.18)	.26	.37				

[a]Cohen's *d* effect size was calculated as the SLI minus MLU mean/pooled standard deviation.

suggests remarkable consistency in the group equivalency on this measure over time, even beyond the 4.0 benchmark of Brown (1973). For PPVT–R raw scores, the SLI group had a significantly higher level of performance at the first time of measurement (M = 32.40, SD = 8.50 for the SLI group; M = 25, SD = 7.81 for the MLU group) and a significantly lower level of performance at the last time of measurement (M = 79.75, SD = 8.90 for SLI and M = 88.61, SD = 8.87 for MLU). Thus, it appears that the older SLI group had somewhat better receptive vocabulary for MLU level as compared to the younger children, perhaps as a function of accumulated word-learning experiences. The two groups were at parallel levels of receptive vocabulary for 2 intervening years before the younger children showed an advantage at the final time of measurement.

Evaluation of growth dimensions within and across groups for MLU and PPVT–R was carried out via growth-curve modeling. Linear mixed models were estimated for each outcome variable using SAS PROC MIXED. Restricted maximum likelihood was used to estimate model parameters and to assess the significance of random effects; degrees of freedom were estimated using the Satterthwaite method. Excellent overviews of growth curve modeling are available elsewhere (e.g., Singer & Willett, 2003), and thus the models are briefly presented here.

The analysis assumes a multilevel model in which the first level comprises the multiple measurements for the individual at different times on a specific outcome variable, and the second level comprises measurements for the individual of predictors that do not vary across time, and which may be related to the individual's level of the outcome variable or the change in the outcome variable across time. For these analyses, time was centered (i.e., set equal to 0) at the first occasion. Equation 1 shows the Level 1 model for the individual growth curves:

$$\text{Level 1}: Y_{it} = \pi_{0i} + \pi_{1i}(\text{time}_{it}) + \pi_{2i}(\text{time}_{it}^{\;2}) + e_{it}, \quad (1)$$

where Y_{it} is the outcome variable for person i at time t, and time$_{it}$ indicates the measurement occasion for individual i. The intercept (π_{i0}) represents the expected level of the outcome variable for individual i at the first occasion (age 3 for the MLU-equivalent group and age 5 for the SLI group). The linear slope for time (π_{i1}) represents the expected change in each outcome for a 1-year interval, and the quadratic slope (π_{i2}) represents the expected rate of change in the linear slope for a 1-year interval. The Level 1 residual is represented by e_{it}, or the difference between the observed and model-predicted value for each individual at each time point.

Characteristics of the individual (i.e., time-invariant predictors) that may affect the outcome variable may be represented by using the parameters at Level 1 as dependent variables. In the current study, these include group (SLI vs. MLU equivalent), mother's education, and age-standardized scores for the CMMS. Mother's education was centered such that mothers with a college education served as the reference group, whereas CMMS was centered at 100, such that children with average levels of CMMS performance served as the reference group. Equation 2 provides the Level 2 model for the prediction of the individual growth parameters:

$$\begin{aligned} \text{Level 2}: \pi_{0i} &= \beta_{00} + \beta_{01}(\text{Group}_i) + \beta_{02}(\text{MomEd}_i) \\ &\quad + \beta_{03}(\text{CMMS}_i) + U_{0i} \\ \pi_{1i} &= \beta_{10} + \beta_{11}(\text{Group}_i) + \beta_{12}(\text{MomEd}_i) \\ &\quad + \beta_{13}(\text{CMMS}_i) + U_{1i} \\ \pi_{2i} &= \beta_{20} + \beta_{21}(\text{Group}_i) + \beta_{22}(\text{MomEd}_i) \\ &\quad + \beta_{23}(\text{CMMS}_i) + U_{2i,} \end{aligned} \quad (2)$$

where π_{i0}, π_{i1}, and π_{i2} are the individual intercepts, linear rates of change, and quadratic rates of change, respectively, that serve as outcomes in the Level 2 model. The fixed effects (i.e., regression slopes) are represented by β, such that β_{00}, β_{10}, and β_{20} indicate the expected intercept, linear rate of change, and quadratic rate of change, respectively, for a child in the SLI group with a mother with a college education and a CMMS score of 100. β_{01}, β_{11}, and β_{21} indicate the expected differences between the SLI and MLU-equivalent groups on the intercept, linear rate of change, and quadratic rate of change, respectively. β_{02}, β_{12}, and β_{22} indicate the expected difference in the intercept, linear rate of change, and quadratic rate of change, respectively, for a one-unit difference in mother's education. β_{03}, β_{13}, and β_{23} indicate the expected difference in the intercept, linear rate of change, and quadratic rate of change, respectively, for a one-unit difference in CMMS. Although not shown in Equation 2, the interaction between group and CMMS on the individual intercepts, linear rates of change, and quadratic rates of change was also estimated. Finally, each individual is expected to have a random deviation from the model-predicted intercept, linear rate of change, and quadratic rate of change, as indicated by the individual random effects of U_{0i}, U_{1i}, and U_{2i}, respectively. The variances of these individual random effects are also known as the Level 2 residual variances.

Means for each time point were estimated with a saturated model (i.e., an ANOVA-type model in which each mean is estimated separately), and are plotted with the model-predicted trajectories for each outcome in Figures 2 and 3, as described below. The model outcomes are described first and are then summarized in terms of the individual questions addressed by this study. Model parameters for each outcome are given in Table 4.

For MLU, significant fixed effects were observed for the linear and quadratic growth parameters. As seen in Figure 2, a negatively accelerating growth function was observed over the 5 years of measurement. Significant individual differences (i.e., random effects) were observed for both initial status and the linear effect of time. Although there were no differences in any of the growth parameters by group, CMMS was significantly related to the intercept, such that higher CMMS scores corresponded to a higher initial level of performance. Thus, the answer to the question of whether there are significant differences between initial matched MLU groups in the observed rates of growth in MLU is "no." Affected children and younger children follow the same growth trajectories. Further, children's nonverbal cognitive levels predicted initial status in MLU to the same extent in both groups (i.e., no Group × CMMS interaction). It is noteworthy that mother's education did not predict growth or intercept.

For PPVT–R raw scores, significant fixed effects were observed for the linear and quadratic growth parameters. As seen in Figure 3, a negatively accelerating growth function was observed across the 3 years of measurement. Significant individual differences (i.e., random effects) were observed for initial status only. Group was significantly related to the intercept and to the linear rate of growth, such that the MLU-equivalent group had a lower average initial status, but a greater degree of linear change relative to the SLI group. CMMS was significantly

Figure 2. Predicted and actual growth in MLU morphemes for SLI and MLU-equivalent groups by year in study (SLI: 5–9 years; MLU matches: 3–7 years).

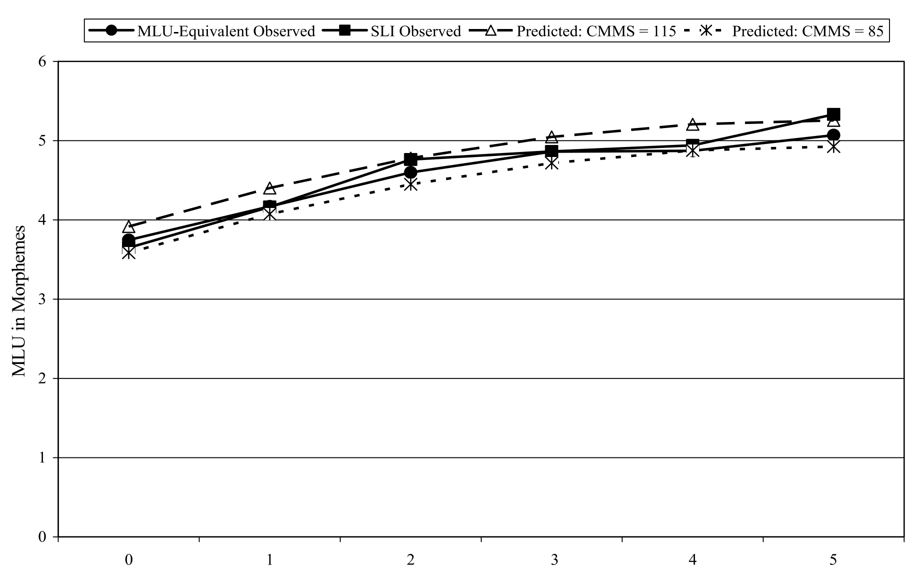
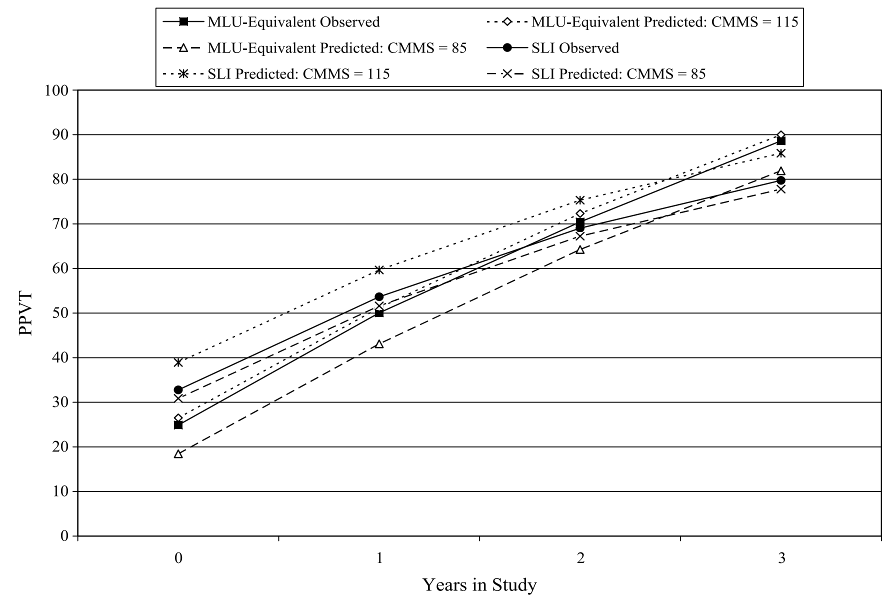


Figure 3. Predicted and actual growth in PPVT–R raw scores for SLI and MLU-equivalent groups by year in study (SLI: 5-8 years; MLU matches: 3–6 years).



related to initial status, such that higher CMMS scores corresponded to a higher initial level of performance, but the effect of CMMS again did not differ by group. Predicted growth curves for CMMS = 85 and CMMS = 115 are also shown in Figure 3. Thus, the answer to the question of whether there are significant differences between groups in the observed rates of growth in vocabulary levels is "yes," of a complex sort: The groups differed at onset, and then the younger children grew to equivalent levels throughout the middle years, leading to higher levels than the affected children. Further, children's nonverbal cognitive levels again predicted initial status in MLU to the same extent in both groups (i.e., no Group × CMMS interaction). As with the MLU outcomes, mother's education did not predict growth in PPVT–R raw scores.

General Discussion

The empirical issue of the integrity of using MLU matches to test hypotheses regarding the nature of language development in children with SLI is addressed in

Table 4. Parameter estimates (Est.) for growth curve models of Study 2.

Parameter	MLU Est.	MLU *SE*	PPVT–R Est.	PPVT–R *SE*
Fixed effects				
Intercept	3.75*	0.09	35.37*	2.39
Linear growth	0.54*	0.06	22.08*	1.80
Quadratic growth	−0.05*	0.01	−2.16*	0.54
SLI vs. MLU Equivalent × Intercept			−13.30*	3.72
SLI vs. MLU Equivalent × Linear Growth			5.55*	0.97
CMMS × Intercept	.01*	.01	0.27*	0.13
Covariance parameters				
Residual variance	0.22*	0.02	42.24*	5.79
Intercept variance	0.16*	0.06	44.94*	14.74
Intercept–slope covariance	−0.02	0.12		
Slope variance	0.02*	0.01		

*$p < .05$.

the outcomes of this study, with findings that bear on concerns about issues of implementation and interpretation. We explored concurrent validity as well as growth trajectories, examining MLU as one of three measures of clausal development (DSS and IPSyn as other indicators) and tracking the growth in MLU as well as the growth in receptive vocabulary.

Overall, the outcomes show robust reliability and validity for MLU as an index of clausal development that tracks young children's language acquisition, for affected as well as unaffected children. With regard to concurrent validity, at school entry age the MLU levels of children with SLI are equivalent to those of children about 2 years younger. MLU matching at the level of morphemes is also equivalent at the level of words. Further, MLU matching captures equivalent levels of performance on two different and independent indicators of clausal complexity, the DSS utterance and total scores, and the IPSyn scores, and shows high levels of correlation with DSS and IPSyn for both groups of children.

It is important to note that although DSS and IPSyn scores include elements of morphological development, they are not sensitive to observed differences in affected children versus MLU controls in the development of grammatical tense marking in obligatory contexts. As noted earlier, it is now well documented that grammatical tense marking is out of synch with MLU development in affected children, at levels of performance lower than expected for MLU levels (cf. Rice, 2004a, 2004b). This lack of sensitivity to deficits in grammatical tense marking may be attributable to the broader scope of clausal measurement, as in the DSS, or by an emphasis on emergence of forms, as in the IPSyn, where up to two instances of a particular form are scored without regard to probability of use in obligatory contexts. Either limitation must be considered when evaluating growth estimates based on DSS or IPSyn measures, and when evaluating characterizations of affected children based on these measures.

The longitudinal data provide the strongest evidence to date of robust reliability and validity of MLU as an index of language acquisition in affected and younger unaffected children, and how it holds up over time. Simply put, unreliable measurement would wreak havoc with the detection of meaningful growth trajectories and stable group similarities over time. The obtained growth trajectories are stable and meaningful, and show consistent benchmarking of MLU equivalency across groups across years of measurement. To the extent that growth curves reflect underlying mechanisms of change, the similarity across the two groups of children suggests that both groups are tapping into similar mechanisms of change. Although the SLI group requires more time to get to the same level of MLU as the unaffected children, the growth in MLU in the range observed here (roughly 3–5 morphemes) follows the same pattern in both groups. Explanations of the nature of language impairments must take these parallels into account. Finally, the predictive relationship with nonverbal cognitive ability is also the same across groups, adding further support to developmental similarities. Mother's education was not a predictor for either group.

In effect, the longitudinal outcomes inform the issue of how age is related to MLU growth in affected and unaffected control children (cf. Klee et al., 1989). Study 1 found no association of age with MLU in the affected group, although Klee et al. (1989) reported such an association for an earlier acquisition period for affected children. The growth trajectories of Study 2 track a strong role for time-modulated effects on MLU in the affected children that parallels that of unaffected younger children. Even though the chronological age levels of the affected children at the outset did not predict their levels of MLU, they nevertheless followed the same growth path as the younger children. The longitudinal outcomes are highly supportive of time-graduated effects but the exact relationship to age levels requires consideration of onset (i.e., when word combinations begin), the shape of the acquisition curves (i.e., linear change of even acceleration over time, as well as points of acceleration change for nonlinear effects), and the portion of the growth curves studied. Eventually, it is likely that the apparent differences in estimates of age–MLU associations can be reconciled by longitudinal data from the earlier stages of MLU acquisition of affected children.

Lexical acquisition during the same period of MLU growth in the studies reported here indicates differences as well as similarities between the groups. Relative to MLU equivalency at the outset, the affected children benefit from their 2-year age advantage with a somewhat greater receptive vocabulary score, but by the final time of measurement the younger children's faster rate of lexical growth leads to performance above the affected group. A preferred method of examining the relationship between MLU and PPVT–R scores over time would be multivariate analyses of growth, allowing for evaluation of within-person correlations. However, this was not possible because of a lack of variance in the linear growth for PPVT–R scores. It is probably too strong to suggest that the affected children are "deviant" in lexical growth relative to the MLU-equivalent children. The predictive relationship with nonverbal cognitive ability, as in the case of MLU, is the same across groups. Further, the nonpredictive outcome for mother's education on children's vocabulary growth also held across groups.

Overall, the outcomes of this investigation point toward both similarities and distinctions in the ways in which MLU and receptive vocabulary grow over time in affected and unaffected children. A caveat is that the affected children in the studies reported here were selected on the basis of receptive vocabulary limitations. The broader clinical diagnosis of SLI can include children with or without such vocabulary deficits, and the literature includes studies in which the affected group is mixed in this regard. It remains to be determined if the generalizations here hold for samples of children ascertained with a broader diagnostic criterion.

To return to the concerns that have been raised by other investigators about the robustness of MLU as a way of equating groups or measuring language growth, the caveat of this study is that considerable care must be taken with MLU measurement, at the level of sampling procedures, data coding, and analyses. In these studies, efforts in this direction included an adequate number of participants and careful definition of "affectedness," explicit examiner training for how to elicit conversationally felicitous speech from children, control of the stimulus materials for comparability across examiners, use of procedures that focus on descriptive comments about ongoing play activities (avoiding narrative contexts or inconsistent intermingling of story narratives with descriptive commentary), collecting conversational samples of sufficient length, and monitoring interexaminer reliability to ensure suitable levels. Finally, a relatively narrow window of MLU equivalency was followed for forming equivalent groups (i.e., within 0.10 morphemes), a control against Type II errors that had been applied inconsistently in previous studies (cf. discussion in Mervis & Robinson, 2003). Under these conditions, it is clear that MLU is a very useful way of benchmarking general language acquisition and furthering our understanding of the ways in which children with SLI do and do not follow the growth patterns of unaffected children. Accurate developmental description of this sort is essential for evaluation of competing models of the nature of language impairment and underlying etiological sources.

Acknowledgments

This study was supported by National Institute on Deafness and Other Communication Disorders Award R01 DC01803. Further support was provided by grants from the National Institutes of Health to the University of Kansas through the Mental Retardation and Developmental Disabilities Research Center (P30HD002528), the Center for Biobehavioral Neurosciences in Communication Disorders (P30DC005803), and R01DC005226.

We express special appreciation to Karla Barnhill and Shannon Wang for their assistance in data collection. Our special appreciation is expressed to the children, parents, and teachers who participated in this study, and to the following schools who supported this research: Edgerton Elementary in Edgerton, KS; Allen Elementary, Avenue A Elementary, Morgan Elementary School, Stafford Elementary, South Hutchinson Elementary, and Union Valley Elementary in Hutchinson KS; Centennial Elementary, Deerfield Elementary, Hillcrest Elementary, Hilltop Child Development Center, Kennedy Elementary, New York Elementary, Quail Run Elementary, St. John's Elementary, Sunflower Elementary, Sunset Hills Elementary, and Wakarusa Valley Elementary in Lawrence, KS; Meadow Lane Elementary in Lee's Summit, MO; Scarborough Elementary in Olathe, KS; Garfield Elementary, Hawthorne Elementary, and Lincoln Elementary in Ottawa, KS; Holy Trinity Elementary, Nativity Parish in Overland Park, KS; and Wellsville Elementary in Wellsville, KS.

References

Barrie-Blackley, S., Musselwhite, C. R., & Rogister, S. H. (1978). *Clinical oral language sampling*. Danville, IL: Interstate Printers and Publishers.

Bennett-Kastor, T. (1988). *Analyzing children's language: Methods and theories*. New York: Basil Blackwell.

Blake, J., Quartaro, G., & Onorati, S. (1993). Evaluating quantitative measures of grammatical complexity in spontaneous speech samples. *Journal of Child Language, 20,* 139–152.

Brown, R. (1973). *A first language: The early stages*. Cambridge, MA: Harvard University Press.

Burgemeister, B. B., Blum, L. H., & Lorge, I. (1972). *The Columbia Mental Maturity Scale*. San Antonio, TX: The Psychological Corporation.

Chabon, S. S., Kent-Udolf, L., & Egolf, D. B. (1982). The temporal reliability of Brown's mean length of utterance (MLU-M) measure with post Stage V children. *Journal of Speech and Hearing Research, 25,* 117–124.

Chan, A., McAllister, L., & Wilson, L. (1998). An investigation of the MLU-age relationship and predictors of MLU in 2- and 3-year old Australian children. *Asia Pacific Journal of Speech, Language, and Hearing, 3,* 97–108.

Chiat, S. (2000). *Understanding children with language problems*. New York: Cambridge University Press.

Conant, S. (1987). The relationship between age and MLU in young children: A second look at Klee and Fitzgerald's data. *Journal of Child Language, 14,* 169–173.

de Villiers, J., & de Villiers, P. (1973). A cross-sectional study of the acquisition of grammatical morphemes in child speech. *Journal of Psycholinguistic Research, 2,* 267–278.

Dollaghan, C. A., Campbell, T. F., Paradise, J. L., Feldman, H. M., Janosky, J. E., & Pitcairn, D. N. (1999). Maternal education and measures of early speech and language. *Journal of Speech, Language, and Hearing Research, 42,* 1432–1443.

Dunn, A., & Dunn, A. (1981). *Peabody Picture Vocabulary Test—Revised*. Circle Pines, MN: American Guidance Service.

Eisenberg, S. L., McGovern Fersko, T., & Lindgren, C. (2001). The use of MLU for identifying language impairment in preschool children: A review. *American Journal of Speech-Language Pathology, 10,* 323–342.

Gavin, W. J., & Giles, L. (1996). Sample size effects on temporal reliability of language sample measures of preschool children. *Journal of Speech and Hearing Research, 39,* 1258–1262.

Goldman, R., & Fristoe, M. (1986). *Goldman Fristoe Test of Articulation*. Circle Pines, MN: American Guidance Service.

Howe, M. (1996). *Transcription and coding manual for analysis of language samples: Kansas language transcript datatbase*. Lawrence: University of Kansas.

Johnston, J. R. (2001). An alternate MLU calculation: Magnitude and variability effects. *Journal of Speech, Language, and Hearing Research, 44,* 156–164.

Johnston, J., & Kamhi, A. (1984). Syntactic and semantic aspects of the utterances of children and language-impaired children: The same can be less. *Merrill-Palmer Quarterly, 30,* 973–978.

Johnston, J., Miller, J. F., Curtiss, S., & Tallal, P. (1993). Conversations with children who are language impaired: Asking questions. *Journal of Speech and Hearing Research, 36,* 973–978.

Klee, T., & Fitzgerald, M. D. (1985). The relation between grammatical development and mean length of utterance in morphemes. *Journal of Child Language, 12,* 251–269.

Klee, T., Schaffer, M., May, S., Membrino, I., & Mougey, K. (1989). A comparison of the age-MLU relation in normal and specifically language impaired preschool children. *Journal of Speech and Hearing Disorders, 54,* 226–233.

Klee, T., Stokes, S. R., Wong, A. M.-Y., Fletcher, P., & Gavin, W. J. (2004). Utterance length and lexical diversity in Cantonese-speaking children with and without specific language impairment. *Journal of Speech, Language, and Hearing Research, 47,* 1396–1410.

Lahey, M., Liebergott, J., Chesnik, M., Menyuk, P., & Adams, J. (1992). Variability in children's use of grammatical morphemes. *Applied Psycholinguistics, 13,* 373–398.

Leadholm, B. J., & Miller, J. F. (1992). *Language sample analysis: The Wisconsin guide*. Milwaukee: Wisconsin Department of Public Instruction.

Lee, L. (1974). *Developmental sentence analysis*. Evanston, IL: Northwestern University Press.

Leonard, L. B. (2004). Specific language impairment in children. In R. D. Kent (Ed.), *The MIT encyclopedia of communication disorders* (pp. 402–405). Cambridge, MA: MIT press.

Leonard, L. B., & Finneran, D. (2003). Grammatical morpheme effects on MLU: "The same can be less" revisited. *Journal of Speech, Language, and Hearing Research, 46,* 878–888.

McCarthy, D. A. (1930). The language development of the pre-school child. *University of Minnesota Institute of Child Welfare Monograph Series IV.* Minneapolis: University of Minnesota Press.

Mervis, C. B., & Robinson, B. F. (2003). Methodological issues in cross-group comparisons of language and cognitive development. In Y. Levy & J. Schaeffer (Eds.), *Language competence across populations: Toward a definition of specific language impairment* (pp. 233–258). Mahwah, NJ: Erlbaum.

Miller, J. (1981). *Assessing language production in children*. Baltimore: University Park Press.

Miller, J., & 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., & Chapman, R. (1991). *Systematic Analysis of Language Transcripts*. Madison: University of Wisconsin Language Analysis Lab.

Newcomer, P. L., & Hammill, D. D. (1988). *Test of Language Development—Second Edition: Primary*. Austin, TX: Pro-Ed.

Nice, M. M. (1925). Length of sentences as a criterion of a child's progress in speech. *Journal of Educational Psychology, 16,* 370–379.

Oetting, J. B., Cantrell, J. P., & Horohov, J. E. (1999). A study of specific language impairment (SLI) in the context of non-standard dialect. *Clinical Linguistics and Phonetics, 13,* 25–44.

Plante, E., Swisher, L., Kiernan, B., & Restrepo, M. A. (1993). Language matches: Illuminating or confounding? *Journal of Speech and Hearing Research, 36,* 772–776.

Rice, M. L. (2003). A unified model of specific and general delay: Grammatical tense as a marker of unexpected variation. In Y. Levy & J. Schaeffer (Eds.), *Language competence across populations: Toward a definition of specific language impairment* (pp. 63–95). Mahwah, NJ: Erlbaum.

Rice, M. L. (2004a). Growth models of developmental language disorders. In M. L. Rice & S. F. Warren (Eds.), *Developmental language disorders: From phenotypes to etiologies* (pp. 207–240). Mahwah, NJ: Erlbaum.

Rice, M. L. (2004b). Language growth of children with SLI and unaffected children: Timing mechanisms and linguistic distinctions. Plenary talk at the Boston University Conference on Language Development. In A. Brugos, L. Micciulla, & C. Smith (Eds.), *Proceedings of the 28th Annual Boston University Conference on Language Development* (pp. 28–49). Somerville, MA: Cascadilla Press.

Rice, M. L., Cleave, P. L., & Oetting, J. B. (2000). The use of syntactic cues in lexical acquisition by children with SLI. *Journal of Speech, Language and Hearing Research, 43,* 582–594.

Rice, M. L., Tomblin, J. B., Hoffman, L., Richman, W. A., & Marquis, J. (2004). Grammatical tense deficits in children with SLI and nonspecific language impairment: Relationships with nonverbal IQ over time. *Journal of Speech, Language, and Hearing Research, 47,* 816–834.

Rice, M. L., & Wexler, K. (1996). Toward tense as a clinical marker of specific language impairment in English-speaking children. *Journal of Speech and Hearing Research, 39,* 1239–1257.

Rice, M. L., Wexler, K., & Hershberger, S. (1998). Tense over time: The longitudinal course of tense acquisition in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 41,* 1412–1431.

Rice, M. L., Wexler, K., Marquis, J., & Hershberger, S. (2000). Acquisition of irregular past tense by children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 43,* 1126–1145.

Rollins, P. R., Snow, C. E., & Willett, J. B. (1996). Predictors of MLU: Semantic and morphological developments. *First Language, 16,* 243–259.

Rondal, J. A., Ghiotto, M., Bredart, S., & Bachelet, J. F. (1987). Age-relation, reliability, and grammatical validity of measures of utterance length. *Journal of Child Language, 14,* 433–446.

Scarborough, H. (1990). Index of productive syntax. *Applied Psycholinguistics, 11,* 1–22.

Scarborough, H. S., Wyckoff, J., & Davidson, R. (1986). A reconsideration of the relation between age and mean length of utterance. *Journal of Speech and Hearing Research 29,* 394–399.

Scarborough, H., Rescorla, L., Tager-Flusberg, H., Fowler, A. E., & Sudhalter, V. (1991). The relation of utterance length to grammatical complexity in normal and language disordered groups. *Applied Psycholinguistics, 12,* 23–45.

Shriner, T. H. (1969). A review of mean length of response as a measure of expressive language development in children. *Journal of Speech and Hearing Disorders, 34,* 61–68.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York: Oxford University Press.

Templin, M. (1957). *Certain language skills in children. University of Minnesota Institute of Child Welfare Monograph Series 26*. Mineapolis: University of Minnesota Press.

Tomblin, J. B. (1996). Genetic and environmental contributions to the risk for specific language impairment. In M. L. Rice (Ed.), *Towards a genetics of language* (pp. 191–210). Mahwah, NJ: Erlbaum.

Tyack, D., & Gottsleben, R. (1974). *Language sampling, analysis and training*. Palo Alto, CA: Consulting Psychologists Press.

van der Lely (1998). SLI in children: Movement, economy and deficits in the computational-syntactic system. *Language Acquisition, 7,* 161–192.

Received February 28, 2005

Revision received September 4, 2005

Accepted December 6, 2005

DOI: 10.1044/1092-4388(2006/056)

Contact author: Mabel L. Rice, 1000 Sunnyside Avenue, 3031 Dole Center, Child Language Doctoral Program, University of Kansas, Lawrence, KS 66045. E-mail: mabel@ku.edu

Lesa Hoffman is now at University of Nebraska—Lincoln.