Methodological Issues in the Calculation of Mean Length of Utterance

Helena Oosthuizen and Frenette Southwood

Stellenbosch University

Abstract

Mean length of utterance (MLU) is widely used as a diagnostic, monitoring and group matching measure. This study investigated methodological issues regarding the calculation of MLU. The aim was to establish whether different calculation procedures render different MLUs, and whether there is a high correlation between MLU measured in words (MLU-w) and in morphemes (MLU-m). Language samples from 15 Afrikaans-speaking 6-year-olds with and 15 without specific language impairment were analyzed. MLU was calculated eight times for each participant, varying sample size (50 or 100 utterances), unit counted (words or morphemes) and calculation method (traditional or alternate). Significant differences in resultant MLUs were due to the calculation method used, rather than sample size or unit counted. A high positive correlation (>0.96) between MLU-w and MLU-m was found. The results imply that researchers and clinicians should clearly state their MLU calculation procedures, otherwise reliable comparisons between MLU scores from different sources cannot be made. The results furthermore imply that, in order to generalize research results and make diagnostic decisions based on MLU, consistent procedures should be used, not only with regard to language sampling, but also to MLU calculation.

Keywords: MLU calculation, MLU-w, MLU-m, Afrikaans, SLI

Spontaneous language measures form an important part of the language evaluation protocol (Dunn, Flax, Swilinski & Aram, 1996; Evans & Miller, 1999) due to the limitations of standardized tests and their limited availability in certain languages, such as Afrikaans (Southwood & Russell, 2004). Because language sampling enables one to assess behaviours directly in a naturalistic context, error and other analyses of spontaneous language samples may be more sensitive to language deficits and less vulnerable to cultural and dialectal bias than standardized language testing (Dunn et al., 1996; Hewitt, Hammer, Yont, & Tomblin, 2005). Various measures have been developed for use during language sample analysis; the most widely used of these appear to be mean length of utterance in morphemes (MLU-m). In a recent survey of speech-language pathologists conducted by Loeb, Kinsler and Bookbinder in the USA (reported in Eisenberg, Fersko, & Lundgren, 2001), 93% of the respondents reported using language sample analysis. Mean length of utterance (MLU) was the most widely used measure, employed by 91% of respondents.

Although measures of utterance length had been used in child language studies since the early 20th century (e.g., Nice, 1925), MLU-m was first popularized by Roger Brown in 1973. Brown found MLU to be more accurate than chronological age in predicting grammatical development, at least up to what he set out as Stage V of language development, which correlates with an MLU of 4. He found evidence of comparable linguistic development between children in each of his five proposed stages; this resulted in many researchers and clinicians employing MLU as a measure of morphosyntactic complexity.

A number of uses of MLU have since been suggested. Two of the main ones are diagnosing a language disorder and selecting intervention goals (Loeb et al. reported in Eisenberg et al., 2001; Miller & Chapman, 1981; Shipley & McAfee, 2004). MLU has also been recommended as a screening tool to identify children in need of further language evaluation (Klee & Fitzgerald, 1985) and to determine the overall level of language development (Miller & Chapman, 1981). This last use has led to the widespread employment of MLU as a matching variable in child language research. According to Eisenberg et al. (2001), there is a need for the validity of MLU to be established separately for each of its uses, if MLU is to be used clinically.

Contact:
Dr Frenette Southwood
Department of General Linguistics
Stellenbosch University
Private Bag X1
Matieland 7602 South Africa
email: fs@sun.ac.za

Despite the widespread use of MLU in language sample analysis, there remains disagreement about the validity and reliability of MLU. A main criticism concerns the absence of a manual specifying the purpose, administration and scoring procedures, normative sample, appropriate reference data, as well as evidence of reliability and validity (Eisenberg et al., 2001). Administrators of standardized language tests expect to find this information about a test in its examiner's manual, which will enable them to determine the appropriateness of a test for a particular child, to repeat test procedures, and to assess its effectiveness. Measures of language sample analysis (including MLU) need to be subjected to the same rigorous criteria as those applied to standardized tests if such measures are to be used diagnostically (Gavin & Giles, 1996).

That no standardized procedure currently exists for calculating MLU makes it difficult to generalise across studies and is confusing for clinicians and researchers faced with conflicting criteria for calculating MLU. This leads to inconsistency with regard to MLU calculation. Against this background, some problematic aspects of MLU will be discussed.

Firstly, MLU calculation depends critically on how utterances are segmented (Eisenberg et al., 2001), yet it remains uncertain exactly what constitutes an utterance. The failure to clearly operationalise the notion of 'utterance' was one of the first criticisms against Brown's original MLU measure (Crystal, 1974). In Chomskyan generative grammar, sentences are considered to be units of language competence, i.e., of fluent speakers' unconscious knowledge of the grammar of their language, whereas utterances are considered units of language use, where *language use* refers to what a person actually says or understands from what another person is saying at a given moment. An utterance is potentially influenced by a variety of nonlinguistic factors such as fatigue, memory limitations, and external distractions, and is therefore often an imperfect reflection of language competence. As stated by Botha (1995), "one and the same sentence can be realized by various utterances which differ from one another" (p. 12), e.g., with regard to acoustic properties such as pitch, intensity, and duration. Although competence is not directly reflected in performance, it is presupposed by every instance of performance. The clinician or child language researcher who is primarily interested in the child's level of linguistic competence has to take the indirect route of using the child's utterances in measuring this competence.

It is more difficult to define the notion 'utterance' than that of 'sentence' (Crystal, Fletcher & Garman, 1976). An utterance could be practically anything verbally produced by the child, any "unit of language" (Rondal, Ghiotto, Bredart, & Bachelet, 1987). Therefore, it is understandable that, thus far, most of the research on MLU has used a process of elimination, focusing on what an utterance is not, rather than on what it is. The original rules provided by Brown (1973) still serve as the basis for excluding and including utterances in a sample, but additional criteria have since been added. Miller and Chapman (1981) segmented utterances "primarily by apparent terminal intonation contour" (p. 155), but reported interrater disagreement of 10-15% for the utterances, which renders this rule insufficient (Eisenberg et al., 2001). Garman (1989) considered a single word, a phrase, or a single clause with its own prosodic identification to be an utterance. Leadholm and Miller (1992) took pauses greater than two seconds to indicate utterance boundaries, and also formulated a rule for dealing with multiple conjoining in order to avoid unnecessarily long utterances. Klee and Fitzgerald (1985) determined utterance boundaries based on major clausal syntactic units, intonation contours, pauses, and speaker turns. In Dutch child language research, the definition of an utterance has generally been based on the notion of the T-unit: one main clause plus any subordinate clause or nonclausal structure attached or embedded in it (Bol, 2003). In short, definitions of *utterance* vary from study to study, and no single definition has yet been agreed upon by researchers.

The second problematic aspect of MLU is that it is influenced by discourse variables (Johnston, 2001) and sampling procedures. Certain pragmatic variables, such as a high frequency of single-morpheme responses and elliptical responses to an adult's questions, can underestimate a child's linguistic abilities, especially in language-impaired populations (Johnston, Miller, Curtiss, & Tallal, 1993; Klee & Fitzgerald, 1985). Children with specific language impairment (SLI) may be overly conscious of their linguistic deficits and therefore reluctant to engage in conversations which will reveal these deficits. Whatever the reasons for the ellipsis, high rates of questioning by clinicians or researchers will probably result in a skewed MLU measure. Johnston et al. (1993) investigated the effect of adult questioning on children's conversations, using a standard interview protocol to elicit a language sample from preschoolers diagnosed with SLI and typically developing children matched for language level. They concluded that at least 35% of the children's utterances would have been longer and more complete had the examiner asked no questions. Furthermore, the children with SLI used more ellipsis in their utterances than did typically developing children, and were more likely to do so as questioning increased.

Johnston (2001) further examined the effects of removing elliptical question responses, imitative utterances, and single-word *yes/no* responses before calculating MLU. She found that this alternate calculation (which she termed *MLU2*) can lead to an increase in children's MLU of 3% to 49%, effectively placing them in the next MLU stage. She concluded that the MLU index contains a discourse-related component which varies in size from sample to sample and does not reflect the child's true linguistic abilities. She argues for the removal of this component to improve stability and developmental sensitivity of the MLU measure. Similarly, Klee and Fitzgerald (1985) reasoned that, because MLU is supposedly an index of syntactic ability, the removal of single-word utterances, such as *yes/no* responses, from the count should allow for greater sensitivity of the MLU measure.

Different elicitation methods have also been shown to affect MLU. Southwood and Russell (2004) compared three different methods of language sample elicitation–story generation; freeplay; and conversation–and found that whereas story generation yielded longer and more complex utterances, freeplay elicited more utterances. In a similar study, Wagner, Nettelbladt, Sahlén, and Nilholm (2000) found that, although MLU in words (MLU-w) was higher in narration than in conversation, children used more complex verb forms in conversation than in narration.

A third criticism against MLU is that, despite its frequent use, it remains relatively unclear what MLU actually reflects in terms of a child's linguistic knowledge. In their study on the association between MLU and measures of expressive vocabulary and morphosyntax, DeThorne, Johnson, and Loeb (2005) concluded that MLU is better viewed as a global measure of expressive language ability, despite its original introduction as a measure of morphosyntactic ability. Eisenberg et al. (2001) recommended that MLU should rather not be regarded as a measurement of morphosyntax, but recognized for what it is, namely "one of several possible ways of measuring utterance length" (p. 324), a conclusion supported by Leonard and Finneran (2003).

Brown himself (1973) noted that the nature of MLU is such that one cannot assume that the utterance length of individual speakers is always the result of the same linguistic means. It does seem possible that a larger vocabulary could translate into longer utterances–as Brown (1973) observed, "almost every new kind of knowledge increases length" (p. 53). New content words allow for the expansion of noun and verb phrases, whereas the acquisition of new function words allows speakers to create entirely new phrases and to conjoin or embed multiple phrases (DeThorne et al., 2005). However, longer utterances are not necessarily more sophisticated than shorter utterances (Crystal et al., 1976). For example, although **We did played* is four morphemes long and *We played* only three, the former is ungrammatical, whereas the latter is not. This dissociation between utterance length, on the one hand, and syntactic complexity and sophistication, on the other, can cause one to overestimate a child's grammatical abilities. Whereas a low MLU can be interpreted as supporting a diagnosis of language impairment, a higher than expected MLU cannot be taken as evidence that no impairment exists (Eisenberg et al., 2001). Alternatively, children with language impairment could go undiagnosed or be included in typically developing control groups on the basis of MLUs which are similar in length but differ qualitatively, as illustrated by **We have play* and *We played*, which both contain three morphemes.

A common design in studies on SLI entails comparing the performance of children with language disorder to that of two different control groups of typically developing children, namely younger children matched according to MLU and age-matched children (Leonard & Finneran, 2003; Rice, Redmond, & Hoffman, 2006). Dual control is employed to enable researchers to compare observed linguistic deficiencies in the SLI group's performance (relative to age expectations) to immature, but typically developing, linguistic systems (Rice et al., 2006). MLU-matching is most appropriate if the dependent measures are influenced by utterance length (Leonard & Finneran, 2003). However, the use of MLU as a matching variable has been questioned for the following reasons. Firstly, children with language disorder are usually older than the MLU-matched children with normal language to whom they are compared (Bol, 2003). It is possible that the linguistic abilities of these children with SLI have been influenced by additional cognitive and nonlinguistic experiences, and can therefore not be compared directly to the linguistic abilities of younger children. A second concern, illustrated by the latter set of examples above, is that children with language impairment, who are known to have specific problems with grammatical morphology, must be compensating in their language production with other aspects in order to have the same MLU as the typically developing children. For instance, Johnston and Kamhi (cited in Leonard & Finneran, 2003) found that, whereas children with SLI made more syntactic errors–mainly omissions of grammatical morphemes–than their typically developing, age-matched peers, the SLI group was also more likely to express progressive aspect (i.e., main verb plus *-ing*) in their utterances. It seems, then, that differences favouring one group might be balanced out by differences favouring the other group.

A fourth problematic aspect of MLU concerns the sample size used in the calculation thereof. Brown (1973) recommended samples of 100 utterances, but–partly because of the difficulty in obtaining a spontaneous speech sample from some children, and also because transcribing is time-consuming–50 (or even fewer) utterances are frequently used (see, e.g., Miller & Chapman, 1981). Gavin and Giles (1996) expressed concern about using samples containing fewer than 100 utterances for MLU calculation, because of low test-retest reliability reported for such small sample sizes. These authors found that spontaneous language measures only have sufficiently high test-retest reliability when sample sizes reach 175 complete and intelligible utterances.

A fifth problematic aspect of MLU concerns the unit of measurement. Although traditionally measured in morphemes, some researchers have found it useful to measure MLU in words for highly inflected languages such as Icelandic (Thordardottir & Weismer, 1998), as well as for Dutch (Arlman-Rupp, Van Niekerk de Haan, & Van de Sandt-Koenderman, 1976). According to Arlman-Rupp et al. (1976), "counting words is faster, easier and theoretically more justifiable than counting morphemes, since no *ad hoc* decisions are necessary" (p. 269), a conclusion also reached by Hickey (1991) with regards to use of measures of utterance length for Irish. MLU-w inevitably leads to a measure equal to or smaller than MLU-m, as bound morphemes are not included in the count (Shipley & McAfee, 2004).

Some researchers report a high correlation between the traditional MLU-m and MLU-w (Arlman-Rupp et al., 1976; Hickey, 1991), whereas others report a low correlation (Klee & Fitzgerald, 1985). A high, positive correlation would suggest that MLU-w could possibly be used in the place of MLU-m with little or no loss of information, as is in fact recommended when grammatical morphemes are the independent variables of a study.

In recent years, language analysis programs such as the Systematic Analysis of Language Transcripts (SALT), developed by Jon Miller and Robin Chapman (1981-1998), and the Child Language Data Exchange System (CHILDES), developed by Brian MacWhinney and Catherine Snow in 1986, have provided clinicians and researchers with alternatives for both the transcription and analysis of children's language samples (Evans & Miller, 1999). Because MLU is calculated in the same way for every transcript which is entered for analysis (according to the conventions of the specific program), variability between results is lessened. Researchers and clinicians should, however, be aware of each program's scoring conventions and use caution when comparing their results to a normative population which differs from the population from which their language sample was taken.

From the preceding discussion, it can be seen that there are several unresolved issues surrounding MLU which could potentially influence its reliability and validity. In spite of this, many researchers are still using MLU, for example to compare results of different studies, without explicating their method of calculation. This illustrates that, although MLU is a popular measure, in practice the procedures used to calculate it remain unclear.

In the South African context, where there is a lack of standardized language assessment instruments for expressive morphology and syntax in, amongst others, Afrikaans, MLU is often used diagnostically. It is thus even more important for South African clinicians to specify their MLU calculation procedures. The general aim of this study was to examine whether the MLU of Afrikaans-speaking 6-year-olds is influenced by methodological issues regarding its calculation. In an attempt to achieve this aim, the following two hypotheses were tested:

Hypothesis 1: Using different sets of criteria when calculating MLU for the same sample will result in a significant difference between obtained scores.

Hypothesis 2: Given that Afrikaans has relatively sparse bound morphology, there is a high correlation between MLU-w and MLU-m for Afrikaans.

Furthermore, the diagnostic strength of MLU for a language other than English (the language on which the most MLU research has been done) was to be determined. The aim was to establish whether MLU successfully differentiates typical language development from atypical language development. In an attempt to achieve this aim, a third hypothesis was tested:

Hypothesis 3: The MLU of the typically developing Afrikaans-speaking 6-year-olds will be significantly higher than that of the Afrikaans-speaking 6-year-olds with SLI.

METHOD

Participants

Thirty monolingual, Afrikaans-speaking 6-year-olds participated. Fifteen were children with SLI. To recruit these participants with SLI, speech-language therapists at government-funded institutions and in private practice were asked to identify from their case loads all 6-year-olds from monolingual Afrikaans-speaking homes who presented with language problems in the absence of hearing, intellectual, socio-emotional, and neurological problems. A total of 16 children were identified and the parents of 15 children consented to participation in the study. Apart from having a language problem, the children also had to have normal intellectual functioning (i.e., a nonverbal IQ score of 85 or above; see Stark & Tallal, 1981) and normal hearing sensitivity, as determined by a hearing screening test, performed according to the American Speech, Language, and Hearing Association's guidelines (ASHA, 1997-2006). As there is currently no agreed-upon protocol for the identification of SLI in Afrikaans-speaking children, mainly due to the lack of Afrikaans-medium language assessment instruments, the judgement of the speech-language therapist of each of these children was used to determine whether a potential participant had SLI.

The other 15 participants were deemed to be typically developing by their parents and teachers. They were matched to the SLI group according to age in months. Participants for inclusion in the typically developing, age-matched (TDA) group were recruited from four aftercare centres in the Stellenbosch area of the Western Cape Province. These children had to meet the following criteria: typically developing in all respects according to their teachers; normal intellectual functioning according to their teachers and parents; and normal hearing as determined by a hearing screening test, performed according to the ASHA guidelines mentioned above.

Procedures

Language sampling. A language sample of 30 minutes was elicited from each participant by the second author, using the same procedure and the same set of manipulable toys for all participants (wooden blocks, figurines with accessories, and plastic kitchen furniture). Each sample was transcribed independently by two graduate research assistants and the transcription was checked against the tape recording by the second author.

Language sample analysis. The transcribed samples were divided into utterances. Following Hunt (1970), an utterance was considered to be a T-unit, i.e., "one main clause plus whatever subordinate clause and non-clausal expressions are attached to or embedded within it" (p. 4).

To test the first hypothesis, MLU was calculated eight times for each child, by systematically varying method (traditional vs. alternate) and unit (words vs. morphemes) each for samples of either 50 or 100 utterances in length.

The method termed *traditional* followed the original rules set out by Brown (1973) as well as rules added by Miller and Chapman (1981), and Leadholm and Miller (1992). The term *traditional* here refers to the fact that these rules have been most widely used by clinicians and researchers since the introduction of MLU. Also, the SALT program (Miller & Chapman, 1981-1998) and its accompanying comparison databases are mostly based on this method. (Short) utterances which reflect the nature of the interaction, rather than the child's actual morpho-syntactic abilities, are often included in the count when using the traditional method, but are removed before calculating MLU according to the alternate method. The alternate method has been shown to be more effective in addressing discourse bias than the traditional method (see Johnston, 2001), and was therefore selected as the second method for this study. Where existing sets of rules were in any way unclear or insufficient, additional rules were formulated in keeping with the overall principles present in the original set of rules. The rules for both methods are given in Appendix A; the alternate method follows the same rules as the traditional method, unless otherwise specified.

To evaluate the reliability of the application of the different sets of criteria and the calculation of the MLU scores, four samples were randomly selected and independently analyzed by the two authors. A high interrater reliability for resultant MLU scores (.987) was found.

Statistical Analysis

A repeated-measures analysis of variance (ANOVA) was performed to examine differences in MLU scores for both groups that resulted from the eight different procedures by which MLU was calculated-i.e., there were eight factors, each corresponding to a method/unit/sample size combination. By using differences between sample means, ANOVA allows one to draw inference about the presence or absence of differences between population means. In repeated measures designs, the same participant serves under more than one treatment condition (e.g., under Traditional, 50 utterances, MLU-w but also under Alternate, 50 utterances, MLU-m). This design is frequently used where the same set of participants are measured repeatedly in the same dependent variable, in this case, MLU, making it appropriate for the type of comparisons made in this study.

Where ANOVA indicated that the overall differences were significant, post hoc pairwise comparisons were made using Tukey's Honest Significant Difference (HSD) procedure. The latter is a multiple comparison procedure (it compares each pair of means with appropriate adjustment for multiple testing), designed to hold the error rate at alpha for a set of comparisons, by comparing every mean with every other mean, while taking into consideration the number of pairwise comparisons among groups.

An alpha level of .05 was used for all statistical tests. The transcriptions of two children in the TDA group and one in the SLI group were of insufficient length to calculate an alternate MLU-w or MLU-m for 100 utterances, and these children's scores were therefore not entered into the statistical analysis.

Ethical Considerations

Clearance was obtained from the ethical committee of the research committee of a university training hospital for those participants who were recruited via organisations related to this hospital. Throughout the study, the ethics and safety standards of the National Research Foundation of South Africa were adhered to.

Written informed consent was obtained from parents for participation of their children in the study, and oral informed assent was obtained from the children. Parents and children were informed of their right to discontinue their participation in the study at any time, with no reasons needed for this decision. Children were informed of their right to rest at any stage during language sample collection and to request that language sampling be terminated, without having to provide reasons for such requests. Anonymity was ensured throughout the study.

RESULTS AND DISCUSSION

Effect of Different Sets of Criteria on MLU Calculation

The mean MLU (i.e., the average MLU based on all eight MLUs of all participants) was 4.13 for the SLI group and 5.31 for the TDA group. Statistically significant results ($p \leq .001$) were obtained in the overall analysis of variance with regard to the two groups, the calculation procedure (i.e., the eight method/unit/sample size combinations), as well as the interaction between the group and the method used, as discussed below.

Differences between the experimental and control groups. There was a statistically significant difference between the averaged sample means of the two groups, $F(1, 25) = 22.15$; $p = .001$), indicating that these observations were not sampled from the same population. The SLI group was indeed distinctly different from the group with typically developing language, and vice versa. The mean of the two groups not only differed significantly statistically, but also placed them in two different developmental stages according to Brown (1973): The SLI group would be in Stage V and the TDA group in the Post-V Stage. This lends support to the third hypothesis stated above. See Appendix B for the MLU-w, MLU-m, and associated MLU stage of each participant.

Differences due to the method used. The mean MLU of the two groups combined was 4.39 for the traditional method with 100 utterances, 4.22 for the traditional method with 50 utterances, 5.83 for the alternate method with 100 utterances, and 5.61 for the alternate method with 50 utterances. The results of the repeated-measures ANOVA indicate that, for both groups, a statistically significant difference exists with regard to the calculation procedures which were used, *F*(3, 75) = 86.65; *p* = .001), i.e., the eight method/unit/sample size combinations. This lends support to the first hypothesis, namely that using different procedures when calculating MLU for the same sample will result in a significant difference between obtained scores. However, no significant differences were found between procedures where the same method was used–i.e., there was no significant difference between scores obtained using the traditional method (regardless of unit or sample size) and there was no significant difference between scores obtained using the alternate method (regardless of unit or sample size)–indicating that the method (traditional or alternate) accounted for the significant difference in scores, and not the unit counted or the sample size. This was confirmed by the results of Tukey's HSD (alpha = .05).

This finding corresponds with that of Johnston (2001), who reported considerable variability in MLU scores due to use of the alternate method of MLU calculation. Johnston illustrated these changes by assigning individual participants to one of Brown's (1973) MLU stages, based on their corresponding absolute MLU values. She concluded that "the alternate calculation procedures can 'jump' children over MLU intervals that are equivalent to the extent of an entire stage" (p. 162). However, the original MLU stages, as set out by Brown (1973), only extend to Stage V, which corresponds to an MLU of 4.0. Other researchers (e.g., Miller & Chapman, 1981) have since elaborated on these stages, for example by adding a Post-V stage which corresponds to an MLU of 4.5+ and an age of 56 months and older. All the participants in this study were 6-year olds who had relatively high MLUs, especially in the TDA group. For example, participant TDA11 had a traditional MLU-m of 5.56 and an alternate MLU-m of 7.14, based on 100 utterances. Both these MLUs would place her in the Post-V stage, thereby obscuring the considerable difference in MLU of 1.58. For this reason, a proportional difference variable was used instead of Brown's stages to determine the magnitude of difference in MLU which resulted from using the alternate procedure.

This score was calculated for the four sets of procedures which differed only in terms of the method used, i.e., traditional [100 utterances; MLU-w] vs. alternate [100 utterances; MLU-w]; traditional [100 utterances; MLU-m] vs. alternate [100 utterances; MLU-m]; traditional [50 utterances; MLU-w] vs. alternate [50 utterances; MLU-w]; and traditional [50 utterances; MLU-m] vs. alternate [50 utterances; MLU-m]. Group means for percentage difference (%DIF) are included in Table 1, where %DIF = (alternate MLU – traditional MLU) / traditional MLU. Proportional difference refers to the relative change that occurred between two scores. For example, there is a 100% increase in MLU from 4.0 to 8.0, but only a 20% increase in MLU from 4.0 to 5.0.

Table 1: *Mean Proportional Gain in MLU Scores due to Use of Altnate instead of Traditional Calculation Method, per Group*

Group	MLU-w/ 100 utterances	MLU-m/ 100 utterances	MLU-w/ 50 utterances	MLU-m/ 50 utterances
TDA	0.42	0.42	0.45	0.44
SLI	0.27	0.27	0.26	0.27

The mean proportional gain in MLU scores due to use of the alternate method was 0.27 for the SLI group, and 0.43 for the TDA group. The MLU score of no participant in the SLI group increased by more than 50%, whereas MLU scores increased by 50-110% for a third of the typically developing participants. The TDA group showed the greatest proportional gain in scores due to use of the alternate method, indicating that the MLUs for the TDA group were more sensitive to the method of calculation than those of the group with language impairment.

There was also considerable variation within the groups, with increases in MLU of as little as 5% to as much as 110%, indicating that individual children's MLUs were influenced to varying degrees by the alternate method of calculation. Johnston (2001) also reported considerable changes in scores for both groups, due to use of the alternate method. In an attempt to explain these findings, Johnston (2001) explored the effect of child variables (such as IQ and language level) as well as discourse variables (such as percentage of adult utterances that were questions and percentage of child utterances that followed adult questions) on the magnitude of the difference scores. She found that the strongest predictor of the magnitude of the %DIF score for her TDA group was the percentage of child responses to questions. In other words, typically developing children whose samples included many responses to adult questions were most affected by the alternate MLU calculation procedures. However, a mixed picture was found for her SLI group. In addition to proportion of question responses, a child's expressive language level–as measured by MLU, amongst others–was found to have an influence on %DIF scores. In light of these findings, Johnston (2001) recommended that an alternate method of MLU calculation should be used whenever 30-40% or more of a child's utterances are replies to questions.

In the present study, the statistically significant differences between groups, as well as within groups, due to the method (traditional vs. alternate) used to calculate MLU can most probably be explained by the differences between the two methods with respect to discourse bias. The alternate method removes most of the (short) discourse-related utterances, whereas the traditional method includes them, leading to lower MLUs. The effect was greatest for the TDA group, where some children's scores increased by more than 80%, solely because the alternate instead of the traditional method was used. This finding contrasts with that of Johnston et al. (1993) viz. that children with SLI made more use of ellipsis in their answers to questions than children with typically developing language, and therefore showed greater changes in their MLU when these utterances were removed. However, Johnston (2001) failed to find this group difference. These different findings can probably be attributed to the different methods of language elicitation used. Johnston et al. (1993) employed an interview protocol containing high levels of questioning, as the focus of their study was on the effect of questions on children's MLU. It is possible that highly structured interactions such as these conveyed implicit messages of expectation rather than friendly interest (Johnston et al., 1993). For the present study, as in the study by Johnston (2001), language samples were elicited in a play context and questioning was kept to a minimum, creating a relaxed atmosphere where there was probably less perceived pressure on participants with language impairment to perform.

It is further possible that the participants with typically developing language were pragmatically superior to their peers with language impairment, and that this competence was reflected in a proportionally higher use of single-word discourse markers. For example, by using single-word responses such as *uh-huh, mm,* and *yes/no* to acknowledge the adult's previous utterance, children are demonstrating that they are aware of the needs of their conversational partner. Therefore, using the traditional method to calculate MLU will most likely penalize children for what is, in fact, a developing conversational skill.

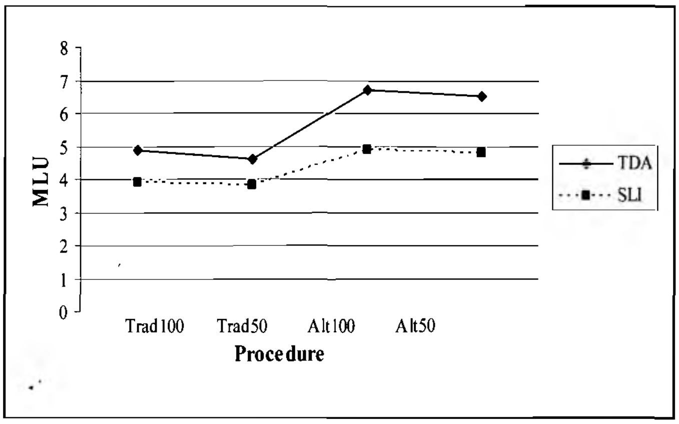

Figure 1. Interaction between Procedure and Group for MLU-w

Whether MLU is used diagnostically or as a matching variable, the underlying assumption is that it should be able to differentiate between a child with language impairment and one with typically developing language. In this study, the alternate method was better able to discriminate between the two groups than the traditional method (see Figure 1), making an alternate MLU the preferred measure for researchers as well as clinicians. The results in Figure 1 specifically pertain to MLU-m. However, similar results were found for MLU-w and therefore also for a combination of MLU-m and MLU-w.

Relationship between MLU-w and MLU-m

Results of the intraclass correlation (ICC) procedure show an ICC agreement correlation above 0.96 for all four methods, as shown in Table 2. For this procedure, MLU scores from both groups were considered simultaneously. The ICC agreement correlation indicates that there is a high, positive correlation between MLU-w and MLU-m for Afrikaans, regardless of the method or sample size used. This is consistent with findings for Irish (Hickey, 1991), Dutch (Arlman-Rupp et al., 1976), and Icelandic (Thordadottir & Weismer, 1998). These findings lend support to the second hypothesis, which predicted a high positive correlation between MLU-w and MLU-m for Afrikaans.

Table 2: *Intraclass Correlation between MLU-w and MLU-m*

Method	ICC agreement	ICC consistency
Traditional (100)	0.968	0.997
Traditional (50)	0.969	0.995
Alternate (100)	0.966	0.998
Alternate (50)	0.960	0.995

As expected, MLU-w was lower than MLU-m. However, an ICC consistency above 0.99 was found for all procedures, indicating that this difference was not significant. This suggests that MLU for Afrikaans could be calculated either in words or in morphemes, without much loss of information. This would make MLU-w the preferred measure of the two, as it is simpler and can be better motivated on theoretical grounds than MLU-m (Arlman-Rupp et al., 1976). However, if MLU is to be used diagnostically, by comparing an individual child's score to some normative group to determine whether the child has a language delay or disorder, MLU-m would be the preferred measure. Given that MLU-w will always be equal to or lower than MLU-m, there is, theoretically, the risk of underestimating a child's language ability if MLU-w is used diagnostically. Also, as stated before, using MLU-m could theoretically overestimate the language abilities of children with SLI–for example, English-speaking children with SLI who overuse the present progressive tense, may artificially increase their MLU by doing so. However, in practice, this risk of underestimating or overestimating an Afrikaans-speaking child's language abilities due to choice of MLU-w over MLU-m, or vice versa, is small, given that the scores are highly correlated. For languages which are typologically different to Afrikaans, this might not necessarily be the case. On the other hand, if the focus of research is on morphological development in children with SLI, MLU-w appears to be the more appropriate matching variable. As stated by Miller and Deevy (2003), care has to be taken not to create a confound: If morphemes are being examined, then employing MLU measured in morphemes as a matching criterion between experimental and control groups seems inappropriate. MLU-m is more suitable as the dependent variable in these cases, also because it is sensitive to changes in morphological development.

CONCLUSION

This study demonstrated that variations in calculation procedures have a significant effect on the MLU for Afrikaans-speaking 6-year olds. Using different sets of criteria to calculate MLU led to significant differences in scores, for both the language-impaired and typically developing groups.

These results have several implications for clinical practice. The first concerns the two different methods used to calculate MLU in the present study. For a spontaneous language measure such as MLU to be used diagnostically, it should be able to differentiate between normal language and impaired or delayed language. Of the two methods used in this study, the alternate method was better able to differentiate between the language-impaired and typically developing groups and is therefore recommended for calculating MLU. Furthermore, Johnston (2001) also showed that the alternate method addressed discourse bias more effectively than the traditional method, as discourse-specific utterances, such as elliptical utterances and single-word *yes/no* responses, are removed before MLU calculation. It is thus recommended that clinicians use the alternate method instead of the traditional method, especially when more than 30-40% of a child's utterances are responses to questions (Johnston, 2001).

The second clinical implication concerns the size of the sample from which MLU was calculated. It was found that MLUs calculated for samples of 50 utterances were generally lower than those calculated for 100 utterances. Use of small samples (50 utterances or less) is not recommended when MLU is used diagnostically, as a lower MLU would place children at a lower language development level than they actually are. Small sample sizes have also been shown to have lower levels of test-retest reliability than samples of 100 utterances or more (Gavin & Giles, 1996). Ideally, samples should consist of 175 complete and intelligible utterances, as these samples have been shown to have high levels of test-retest reliability (Gavin & Giles, 1996). However, in practice, it is often difficult to elicit sufficiently long language samples from children–even more so if they are severely language-impaired or shy. It is therefore recommended that clinicians use sample sizes of at least 100 utterances to calculate MLU, and use samples of 50 utterances or less only when monitoring a child's progress during intervention.

A third clinical implication concerns the unit counted when calculating MLU, namely words or morphemes. A high positive correlation has been found between MLU-w and MLU-m for Afrikaans, indicating that either of these measures could be used without much loss of information. Although counting words is easier and faster, and can be better justified on theoretical grounds than counting morphemes (Arlman-Rupp et al., 1976), clinicians need to bear in mind that MLU-w will always be smaller than, or equal to, MLU-m. This means that there is, theoretically, the risk of underestimating a child's language abilities when using MLU-w diagnostically, and MLU-m would then be the more appropriate measure for diagnostic purposes.

The results from this study also have implications for research. Researchers should at all times clearly state their procedures for calculating MLU; failure to do so would mean that results from different studies cannot be compared reliably. Furthermore, many researchers employ MLU either as a matching variable or as a dependent variable. If morphemes are being examined, MLU-w would be a more appropriate matching variable, whereas MLU-m would be more suitable as a dependent variable in an experimental context, as it is sensitive to morphological changes.

The present study has certain limitations. Firstly, a relatively small number of participants were used, which could limit generalisability of results. Secondly, participants in this study were all monolingual 6-year-old Afrikaans-speaking children. Therefore, results must be interpreted with caution when applied to a population other than the one described in the present study. A third limitation concerns the two sets of criteria regarding the method of MLU calculation. For the purposes of the present study, the set of rules guiding the two methods of MLU calculation had to be rather stringent. The majority of clinicians and researchers would probably use broader criteria than those used here, supplementing existing rules with other best-practice principles. The possibility exists that, in practice, different methods of MLU calculation might result in less significant differences between scores, because the methods are defined by broader criteria than those used for the purposes of this study. This implies that the calculation methods used in this study do not necessarily replicate real-life clinical decisions regarding MLU calculation. Further research is needed to explore the effect of different, but less narrowly defined, sets of criteria on the calculation of MLU.

The attractiveness of MLU lies in the fact that it is simple to understand and easy to calculate, and allows for preliminary ordering of the data from child language samples. However, in order to generalise findings or to make diagnostic decisions based on MLU, consistent procedures should be used not only with regard to language sampling, but also to MLU calculation.

AUTHOR NOTE

This material is based on work financially supported by The

National Research Foundation of South Africa. Any opinion, findings, conclusions or recommendations expressed in this material are those of the author and therefore the NRF does not accept any liability in regard thereto.

REFERENCES

Arlman-Rupp, A. J., Van Niekerk de Haan, D., & Van de Sandt-Koenderman, M. (1976). Brown's early stages: Some evidence from Dutch. *Journal of Child Language, 3*, 267-274.

American Speech, Language, and Hearing Association. (1997-2006). Hearing screening. http://www.asha.org/public/hearing/testing/ (accessed 25 October 2007).

Bol, G. W. (2003). MLU-matching and the production of morphosyntax in Dutch children with specific language impairment. In Y. Levy & J. Schaeffer (Eds.), *Language Competence across Populations: Towards a Definition of Specific Language Impairment* (pp. 259-271). Mahwah, NJ: Lawrence Erlbaum Associates.

Botha, R. P. (1995). The world of language: A Carrollinian canvas. *Stellenbosch Papers in Linguistics, 29.*

Brown, R. (1973). *A First Language – The Early Stages.* London: George Allen and Unwin Ltd.

Chomsky, N. (1965). *Aspects of the Theory of Syntax.* Cambridge, MA: MIT Press.

Crystal, D. (1974). Review of R. Brown, A first language. *Journal of Child Language, 1*, 289-307.

Crystal, D., Fletcher, P., & Garman, M. (1976). *The Grammatical Analysis of Language Disability: A Procedure for Assessment and Remediation.* London: Edward Arnold.

DeThorne, L. S., Johnson, B. W., & Loeb, J. W. (2005). A closer look at MLU: What does it really measure? *Clinical Linguistics and Phonetics, 19*, 635-648.

Dunn, M., Flax, J., Sliwinski, M., & Aram, D. (1996). The use of spontaneous language measures as criteria for identifying children with specific language impairment: An attempt to reconcile clinical and research incongruence. *Journal of Speech and Hearing Research, 39*, 643-654.

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

Evans, J. & Miller, J. W. (1999). Language sample analysis in the 21st century. *Seminars in Speech and Language, 20*, 101-116.

Garman, M. (1989). The role of linguistics in speech therapy: Assessment and interpretation. In P. Grunwell, & A. James (Eds.), *The Functional Evaluation of Language Disorder* (pp. 29-57). London: Croom Helm.

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.

Hewitt, L. E., Hammer, C. S., Yont, K. M., & Tomblin, J. B. (2005). Language sampling for kindergarten children with and without SLI: Mean length of utterance, IPSYN, and NDW. *Journal of Communication Disorders, 38*, 197-213.

Hickey, T. (1991). Mean length of utterance and the acquisition of Irish. *Journal of Child Language, 18*, 553-569.

Hunt, K. W. (1970). Syntactic maturity in school children and adults. *Monographs of the Society for Research in Child Language Development* (35)1. Serial no. 134.

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

Johnston, J. R., 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. (1985). The relation between grammatical development and mean length of utterance in morphemes. *Journal of Child Language, 12*, 251-269.

Leadholm, B. J., & Miller, J. F. (1992). *Language Sample Analysis: The Wisconsin Guide.* Madison, WI: Wisconsin Department of Public Instruction.

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.

MacWhinney, B. & Snow, C. (1986). *Child Language Data Exchange System (CHILDES).* http://childes.psy.cmu.edu/ (accessed 25 October 2007).

Miller, J. F. & Chapman, R. S. (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. S. (1981-1998). *Systematic Analysis of Language Transcripts (SALT).* Retrieved on October, 25, 2007 from http:// www. languageanalysislab.com/ salt.

Miller, C. A. & Deevy, P. (2003). A method for examining productivity of grammatical morphology in children with and without specific language impairment. *Journal of Speech, Language, and Hearing Research, 46*, 1154-1166.

Nice, M. M. (1925). Length of sentences as a criterion of a child's progress in speech. *Journal of Educational Psychology, 6*, 370-379.

Rice, M. L., Redmond, S. M. & Hoffman, L. (2006). Mean length of utterance in children with specific language impairment and in younger control children shows concurrent validity and stable and parallel growth trajectories. *Journal of Speech, Language, and Hearing Research, 49*, 793-808.

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

Shipley, K. G. & McAfee, J. G. (2004). *Assessment in Speech-Language Pathology: A Resource Manual* (3rd ed.). San Diego, CA: Singular Publishing Group.

Southwood, F. & Russell, A. F. (2004). Comparison of conversation, freeplay, and story generation as methods of language sample elicitation. *Journal of Speech, Language, and Hearing Research, 47*, 366-376.

Stark, R. E., & Tallal, P. (1981). Selection of children with specific language deficits. *Journal of Speech and Hearing Disorders, 46*, 114-122.

Thordardottir, E. T. & Weismer, S. E. (1998). Mean length of utterance and other language sample measures in early Icelandic. *First Language, 18*, 1-32.

Wagner, C. R., Nettelbladt, U., Sahlén, B., & Nilholm, C. (2000). Conversation versus narration in pre-school children with language impairment. *International Journal of Language and Communication Disorders, 35*, 83-93.

APPENDIX A

Traditional and Alternate Methods of Calculating MLU

Traditional method.

Exclude:

(a) Totally or partially unintelligible utterances.

(b) Dysfluencies–the word is counted once in the most complete form produced.

(c) Fillers such as *mm* or o 'oh' (and their equivalents, such as *a, um, uh).*

(d) Utterances with "a long string of conjoined words or phrases based on, for example, objects in the room" (Miller & Chapman, 1981. p. 156.)

(e) False starts, reformulations, and revisions–the most complete form of the utterance is included.

Include:

(a) Exact utterance repetitions.

(b) Single-word utterances, such as *ja* 'yes' and *nee* 'no' (and their equivalents, such as *huh-uh, uh-huh, jip,* yup, *OK*), including interjections, such as *wow, cool, jislaaik* 'gee wiz'.

(c) Social and formulaic utterances, such as *ek weet nie* 'I don't know'*, daar's hy,* 'there you go (literally: there he is)', *nee dankie* 'no thanks', *hey, foeitog* 'shame', *tannie* 'auntie', *ekskuus* 'sorry / excuse me'.

(d) Utterances where the child completes the adult's utterance.

(e) Incomplete/abandoned utterances not followed by a revision.

(f) Sounds that are incorporated into the meaning of the utterance, such as *dan gaan ek* so sshhoo 'then I go sshhoo'.

(g) Idiosyncratic words or utterances, as in **looka hy kop is aan* ' looka him head is on'.

Count as one morpheme:

(a) Words produced for emphasis (count each occurrence).

(b) Proper names.

(c) Irregular past tense verb forms, such as *was* 'was', *kon* 'could', *wou* 'wanted to', *sou* 'would'.

(d) Diminutives, such as *hondjie* 'little dog', *mannetjie* 'little man'.

(e) Auxiliary verbs, such as *het* 'have', *wil* 'want to', *gaan* 'going to', *kan* 'can', *sal* 'will'.

(f) Catenatives, such as *'t* (*het* 'has'), 's (*is* 'is'), *dis* (*dit+is* 'this+is'). If the words *is* and *het* appear in uncontracted form elsewhere in the child's sample, the contracted versions are counted as two morphemes; otherwise, one (see Miller & Chapman, 1981). If the word *is* appears separately in the child's sample, the contracted version *dis* is counted as one word and two morphemes.

(g) Compound words (i.e., two or more free morphemes), such as *see-speel-goeters* 'sea-play-things', *hierso* 'over here', *graad twee* 'grade two'.

(h) Ritualized reduplications, such as *speel-speel* 'in a playful/easy way', *nou-nou* 'just now'.

Count as two or more morphemes all inflected word forms, including:

(a) Plural nouns.

(b) Regular past tense verb forms, such as *gewerk* 'work-PAST PARTICIPLE'.

(c) Inflected adjectives denoting degrees of comparison, e.g., *kleiner* 'smaller', *kleinste* 'smallest'.

(d) Complex verbs consisting of a verb + preposition, such as *afspring* 'jump off', *opklim* 'climb up'.

Alternate method.

Exclude:

(a) Exact self-repetitions.

(b) Exact repetitions of the adult partner.

(c) Single-word responses *ja* 'yes', *nee* 'no', and their equivalents, whether occurring (i) as an answer to a question, (ii) as an acknowledgement of the adult's previous utterance, or (iii) during self-talk.

(d) Responses to *wh*-questions in which only the queried constituent was provided. If the child answers a constituent query with a full sentence, the utterance is not removed. Also, answers to open-ended questions are treated as exceptions to this rule.

(e) Incomplete or abandoned utterances not followed by a revision.

(f) Utterances where the child completes the adult's utterance.

(g) Single-word utterances used by the child to gain the listener's attention.

(h) Social or formulaic utterances, such as *wat is dit?* 'what is this?', *nee dankie* 'no thanks', *kyk hier* 'look here', so 'like this'.

Include:

(a) Spontaneous single-word utterances that do not reflect discourse bias, e.g., during spontaneous naming of objects or self-talk.

(b) *Yes/no* responses immediately followed by a full-clause elaboration.

APPENDIX B: MLU-w, MLU-m, and Associated MLU Stage for Each Participant

Participant	Measure	Traditional 100		Traditional 50		Alternate 100		Alternate 50	
		MLU	Brown	MLU	Brown	MLU	Brown	MLU	Brown
SLI1	MLU-w	3.47	Early IV	3.24	Early IV	4.39	V	4.16	V
	MLU-m	3.65	Late IV	3.44	Early IV	4.61	Post-V	4.44	V
SLI2	MLU-w	3.4	Early IV	3.38	Early IV	4.34	V	4.4	V
	MLU-m	3.51	Late IV	3.56	Late IV	4.51	Post-V	4.58	Post-V
SLI3	MLU-w	3.44	Early IV	3.24	Early IV	4.61	Post-V	4.64	Post-V
	MLU-m	3.66	Late-IV	3.48	Early IV	4.78	Post-V	4.96	Post-V
SLI4	MLU-w	2.89	Late III	2.72	Early III	4.11	V	3.94	V
	MLU-m	3.22	Early IV	2.98	Late III	4.5	V	4.42	V
SLI5	MLU-w	4.03	V	3.74	Late IV	4.68	Post-V	4.32	V
	MLU-m	4.21	V	3.90	V	4.95	Post-V	4.5	V
SLI6	MLU-w	3.23	Early IV	3.1	Early IV	4.14	V	3.92	V
	MLU-m	3.37	Early IV	3.24	Early IV	4.3	V	4.06	V
SLI7	MLU-w	4.18	V	3.78	V	5.05	Post-V	4.7	Post-V
SLI8	MLU-w	4.38	V	4.54	Post-V	5.91	Post-V	5.52	Post-V
	MLU-m	4.53	Post-V	4.74	Post-V	6.11	Post-V	5.74	Post-V
SLI9	MLU-w	3.68	Late IV	3.74	Late IV	4.54	Post-V	4.56	Post-V
	MLU-m	3.84	V	3.88	V	4.8	Post-V	4.72	Post-V
SLI10	MLU-w	3.47	Early IV	3.26	Early IV	4.69	Post-V	4.02	V
	MLU-m	3.74	Late IV	3.5	Late IV	5.08	Post-V	4.42	V
SLI11	MLU-w	3.82	V	3.56	Late IV	4.99	Post-V	5.04	Post-V
	MLU-m	4.08	V	3.78	V	5.36	Post-V	5.5	Post-V
SLI12	MLU-w	3.62	Late IV	3.18	Early IV			4.12	V
	MLU-m	3.84	V	3.4	Early IV			4.38	V
SLI13	MLU-w	3.79	V	4	V	4.21	V	4.34	V
	MLU-m	4.03	V	4.38	V	4.52	Post-V	4.7	Post-V
SLI14	MLU-w	3.82	V	4.18	V	5.14	Post-V	4.88	Post-V
	MLU-m	3.96	V	4.34	V	5.47	Post-V	5.1	Post-V
SLI15	MLU-w	3.85	V	3.8	V	4.14	V	4.36	V
	MLU-m	4.13	V	4	V	4.39	V	4.66	Post-V

Participant	Measure	Traditional 100		Traditional 50		Alternate 100		Alternate 50	
		MLU	Brown	MLU	Brown	MLU	Brown	MLU	Brown
TDA1	MLU-w	3.84	V	3.32	Early IV	5.26	Post V	4.86	Post V
	MLU-m	4.07	V	3.52	Late IV	5.66	Post V	5.16	Post V
TDA2	MLU-w	3.45	Early IV	3.58	Late IV	5.5	Post V	4.68	Post V
	MLU-m	3.57	Late IV	3.7	Late IV	5.76	Post V	4.82	Post V
TDA3	MLU-w	4.53	Post V	4.12	V	5.41	Post V	5.02	Post V
	MLU-m	4.68	Post V	4.24	V	5.58	Post V	5.12	Post V
TDA4	MLU-w	4.58	Post V	4.86	Post V			6.4	Post V
	MLU-m	4.88	Post V	5.16	Post V			6.86	Post V
TDA5	MLU-w	4.52	Post V	3.6	Late IV	7.48	Post V	6.68	Post V
	MLU-m	4.81	Post V	3.96	V	7.87	Post V	7.12	Post V
TDA6	MLU-w	4.04	V	3.26	Early IV	6.51	Post V	6.56	Post V
	MLU-m	4.24	V	3.34	Early IV	6.88	Post V	6.88	Post V
TDA7	MLU-w	6.96	Post V	7.14	Post V	8.02	Post V	8.1	Post V
	MLU-m	7.35	Post V	7.64	Post V	8.32	Post V	8.48	Post V
TDA8	MLU-w	6.12	Post V	6.24	Post V	6.54	Post V	6.48	Post V
	MLU-m	6.4	Post V	6.54	Post V	6.77	Post V	6.74	Post V
TDA9	MLU-w	5.31	Post V	4.62	Post V	6.39	Post V	6	Post V
	MLU-m	5.62	Post V	4.88	Post V	6.73	Post V	6.38	Post V
TDA10	MLU-w	3.82	V	3.54	Late IV	5.72	Post V	5.38	Post V
	MLU-m	3.98	V	3.78	V	5.98	Post V	5.66	Post V
TDA11	MLU-w	5.25	Post V	5.9	Post V	6.77	Post V	7.06	Post V
	MLU-m	5.56	Post V	6.4	Post V	7.14	Post V	7.52	Post V
TDA12	MLU-w	3.58	Late IV	4.02	V			5.76	Post V
	MLU-m	3.76	V	4.18	V			6.06	Post V
TDA13	MLU-w	3.86	V	3.54	Late IV	7.09	Post V	6.5	Post V
	MLU-m	4.1	V	3.8	V	7.48	Post V	6.94	Post V
TDA14	MLU-w	5.31	Post V	4.54	Post V	6.48	Post V	6.22	Post V
	MLU-m	5.55	Post V	4.82	Post V	6.8	Post V	6.52	Post V
TDA15	MLU-w	3.65	Late IV	3.68	Late IV	6.39	Post V	6.16	Post V
	MLU-m	3.78	V	3.78	V	6.71	Post V	6.44	Post V

Note: *The samples of three of the participants* (*SLI12, TDA4, TDA12) did not contain 100 or more utterances when using the alternate method.*