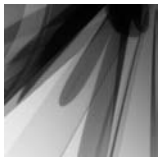

FIRST LANGUAGE

ARTICLE

www.sagepublications.com Vol 25(3): xxx-xxx (200510)
DOI: 10.1177/0142723705059114

A comparative study between mean length of utterance in morphemes (MLUm) and mean length of utterance in words (MLUw)

MATTHEW D. PARKER, *University of Minnesota Duluth*
KENT BRORSON, *University of Minnesota Duluth*

ABSTRACT
Prior to Brown's (1973) introduction of mean length of utterance in morphemes (MLUm), child language researchers and speech-language pathologists used mean length of utterance in words (MLUw) as a measurement of a child's gross language development. After Brown (1973) and others documented MLUm to be a measure which was correlated with the development of morphological and syntactic skills in young children, the practice of counting MLUm became more widely used and accepted. In the present study, MLUw and MLUm scores of 40 language transcripts from typically-developing, English-speaking children between the ages of 3;0 and 3;10 were compared. Results indicated that MLUm and MLUw are almost perfectly correlated. This finding suggests that MLUw can be used as effectively as MLUm as a measurement of a child's gross language development.

KEYWORDS
Assessment; early childhood; language sample; morphology

INTRODUCTION

Speech-language pathologists (SLPs) frequently use a non-standardized form of testing known as Language Sample Analysis (LSA) as an alternative to standardized testing, or to supplement the standardized testing, in evaluation of language skills. Survey results from Hux, Morris-Friehe & Sanger (1993) and Kemp & Klee (1997) illustrated that 80–85% of SLPs in the USA use LSA to supplement standardized testing. LSA is a descriptive approach to language assessment in that it allows the examiner to 'evaluate linguistic achievements within the basic processes of comprehension, production, and observe a child's use of language for communication' (Miller, 1985: 2). Language samples are usually based on observations and a collected conversation sample.

The analysis of language samples has been a recommended clinical practice in speech-language pathology for at least a quarter of a century (Owens, 1999), and now language sample analyses exist that assess virtually every aspect of language structure (Klee, 1992). Through language sampling, 'specific data can be obtained by probing the child's conversational behavior' (Owens, 1999: 120). With these specific data, SLPs are better able to describe language disorders and delays, to select goals and to plan treatment accordingly (Hughes, Fey & Long, 1992). LSA also tends to be more useful than standardized tests in determining specific treatment goals (Paul, 2000). Language sampling is appealing as it allows for a more natural discourse, while standardized tests generally require artificial language usage in which structured responses are expected (Butler, 1992).

Language sample analysis developments in the twentieth century

One procedure commonly used in language analysis is a measurement of a child's mean length of utterance. The measurement of utterance length in children's speech has been the backbone of LSA since the 1920s when Nice (1925) used 'mean length of response' (MLR) to demonstrate the predictable patterns of child language development. MLR measured utterance length by dividing the total number of words by the number of utterances. The way in which utterance length is calculated and described has evolved since Nice first demonstrated its usefulness; one of these changes involved the renaming of MLR to 'mean length of utterance in words' (MLUw). Although MLR changed to MLUw, it still measured utterance length by dividing the total number of words by the number of utterances.

After mean utterance length measurements were shown to be a useful measure of a child's gross language development, researchers began changing the way in which utterance length was computed. Another change involved measuring utterance length by counting the number of words while only counting those utterances that consisted of two words or more, whereas MLUw included utterances of all lengths to complete the count. This measurement became known as 'mean syntactic length' (MSL). MLU in syllables (MLUs) measures utterance length by dividing the total number of syllables by the number of utterances. However, syllable counts are more difficult to conduct due to children's tendencies to duplicate syllables as well as their usage of diminutives (Hickey, 1991). With increased use of duplication of syllables and diminutives, inflation of the MLUs scores is likely to occur (Hickey, 1991).

In 1973 Roger Brown proposed a new method for computing utterance length which is known as 'mean length of utterance in morphemes' (MLUm). MLUm is computed by dividing the total number of morphemes in an utterance by the total number of utterances. Brown (1973: 54) regarded MLUm as 'an excellent simple index of grammatical development'. The use of MLUm as a measure of language growth is based on Brown's work showing that most advances in language development result in increases in length, that is, the addition of words or other linguistic elements to utterances. Brown (1973) and Devilliers & Devilliers (1973) documented that MLUm in the English language was correlated with the development of morphological and syntactic skills in young children. Based on this finding, Brown constructed 5 stages of linguistic development based on MLU values of 1.75, 2.25, 2.75, 3.5 and 4.0, with an MLUm of 1.75 corresponding to Stage I and an MLUm of 4.0 corresponding to Stage V. Brown suggested that up to an average of 4.0 MLUm is considered a good measurement of language complexity. However, once MLUm is over this level, it is no longer considered an accurate measure because henceforth, 'much of the growth in complexity is the result of internal reorganization of utterance form, rather than addition of new structure' (Owens, 1999: 190). Each of Brown's stages is associated with distinct developmental achievements including use of: grammatical morphemes; negation; yes/no questions; Wh- questions; noun phrase elaboration; verb phrase elaboration; complex sentences (Retherford, 2000). Therefore, MLUm is used to predict where a child is, or should be, in the sequence of structural acquisition (Chapman, 1985).

Since Brown introduced MLUm it has been a widely accepted and used measurement of gross language development. A survey of SLP by Loeb, Kinsler and Bookbinder (cited in Eisenberg, Fersko & Lundgren, 2001) indicated that the procedure used most frequently by SLPs in LSA is MLUm. Several studies have documented that MLUm is highly correlated with age for normal children, particularly at very early stages of language development, with MLUm growth decelerating in the later preschool years (Conant, 1987; Miller & Chapman, 1981; Rondal, Ghiotto, Bredart & Bachelet, 1987). MLUm has also been shown to correlate with age for children with specific language impairment (Klee, Schaffer, May, Membrino & Mougey, 1989), and with age for the mental retardation populations (Rondal *et al*., 1987). Another motivation for the use of MLUm has been the assumption that it is a better predictor of grammatical development than is age (Brown, 1973; Devilliers & Devilliers, 1973). Some suggested uses for MLUm include: to determine stage of overall level of language development (Bernstein & Tiegerman-Farber, 1997; Miller & Chapman, 1981); to identify children in need of further language evaluation (Miller & Chapman, 1981); to diagnose or identify a language impairment (Bernstein & Tiegerman-Farber, 1997; Lahey, 1988; Miller & Chapman, 1981; Owens, 1999); to guide further language assessment (Paul, 2000); and to measure change in language production (Fey, 1986; Paul, 2000). However, Miller & Chapman (1981) caution that utterance length measures should be used to describe a child's developing gross language skills and not as an end-all diagnostic tool.

Brown (1973) provided guidelines to follow when attempting to complete MLUm measurements. One guideline states that MLUm counts should include only those morphemes for which there is evidence of productive use or the application of a systematic rule by the child. Therefore, Brown assigns only one morpheme to irregular past tense forms due to the lack of evidence that a child relates the irregular past

tense to the present tense. Words in the catenative form such as 'gonna' and 'wanna' and expressions such as 'lookit' and 'all done' are assigned one morpheme on the grounds that they function as one morpheme in the child's speech (Brown, 1973). A child would be credited with two morphemes for using the negative contraction 'don't' only if there is evidence within the transcript of the use of 'do' and 'not' (Retherford, 2000). When SLPs are following rules provided by Brown (1973), it is their responsibility to interpret what is a productive morpheme and what is not. Many difficulties arise for researchers as well as for SLPs who attempt to use MLUm measure while examining a child's language other than English. Thordardottir & Weismer (1998: 3) concluded that, 'the productivity requirement is especially difficult to meet for languages in which grammatical development has not been documented extensively'. Brown (1973) also noted that studies of highly inflected languages, such as Finnish, Swedish and Spanish all presented difficulties in adapting his rules for MLUm calculations.

Recent research

Although MLUm is a generally accepted procedure for child language analysis, its use is debatable. Dromi & Berman (1982) introduced the measure for utterance length called 'morpheme-per-utterance' (MPU); this measure was used instead of the MLUm index because, in the Hebrew language, increased complexity often does not imply an increase in length, as inflectional morphemes are not added to word roots in a linear fashion. Johnston (2001) suggested calculating MLUm after removing elliptical questions, imitative utterances and single-word responses to help to control discourse variables that may result from pragmatic factors within the language sample collection process. Eisenberg *et al.* (2001) suggested that median or modal utterance length may be a more appropriate way of measuring utterance length as the median and mode scores are also less affected than the mean by the presence of nonrepresentative segments of the sample. Another possibility for analysing utterance length would to look at the length of the longer utterances above the mode (Eisenberg *et al.*, 2001).

Researchers in Dutch, Irish, and Icelandic languages have found correlations of 0.98–0.99 between MLUm and MLUw (Arlman-Rupp *et al.*, 1976; Hickey, 1991; Thordardottir & Weismer, 1998). These strong correlations have led some researchers to believe that MLUw may be a better and more reliable measure for calculating utterance length and a more sensitive measure of a child's language development and complexity (Arlman-Rupp, Van Niekirk-de Hahn & Van de Sandt-Koenderman, 1976; Hickey, 1991; Malakoff, Mayes, Schottenfeld & Howell, 1999; Thordardottir & Weismer, 1998). Hickey (1991: 568) states, 'MLU counted in words was found to be a measure which best balanced effectiveness and ease of application'. Similarly, Arlam-Rupp *et al.* (1976: 233) suggested that 'counting words is easier, faster, more reliable, and theoretically more sound because no ad hoc decisions need to be made. The high correlation between MLUm and MLUw suggests that it is unnecessary to use MLUm as a means of calculating MLU, especially given the uncertain nature of morpheme development (Hickey, 1991: 565).

To reach consensus on whether MLUw is the optimal measure, we need further data on its relationship to the more commonly used, but potentially more problematic, MLUm. The aim of the present investigation is to examine the relationship between

MLUw and MLUm scores in forty language transcripts from typically-developing, English-speaking children aged 3;0–3;10.

METHOD

Language sample source

In most studies of language development it is necessary to recruit subjects and collect language samples from the participants. Instead, this study used the *Child Language Data Exchange System (CHILDES)* (MacWhinney, 2000), an existing internet database, as the source for the collection of 40 language transcripts.

The CHILDES consists of three components: (1) the database of language transcripts, (2) Codes for Human Analysis of Transcripts (CHAT), and (3) Computerized Language Analysis (CLAN). The database contains a large collection of language transcripts. The aim of the second component, CHAT is to 'provide a standardized way of preparing digitized transcripts of spoken or signed language that can then be subjected to computer analysis' (MacWhinney, 2000: 392). All contributing researchers are asked to transcribe their language samples following the specific protocol provided. Component 3, CLAN, enables researchers to analyse any language transcript in the CHILDES database that are in CHAT format.

The nature of the 40 transcripts

The present researchers collected 40 language transcripts from the CHILDES database (MacWhinney, 2000). The selected transcripts exist in the 'English Corpa' file. The transcripts are from normal/typically developing children with an age range of 3;0–3;10. Two groups were established based on age of the children: 3;0–3;5 in the first group and 3;6–3;10 in the second. Each group contained 20 transcripts from the children (50% female, 50% male). The number of contributing researchers totalled 11. The number of transcripts contributed from each researcher, and number of different children are given in Table 1. The target child was interacting with the mother in 12.5% of the transcripts, with the father in 12.5%, with an investigator in 12.5%, with both the mother and father in 25%, with two children in 20%, and with an investigator and parent in 17.5%. The task in which the sample was gathered occurred as follows: 69% while in free-play, 13% while eating, 8% while engaged with a book, and in 10% the task was unknown. The setting in which the language sample was gathered was 40% while in a clinic type environment, 57% while at home, and in 3% of the transcripts the setting was unknown. Language transcripts ranged in size from 652 utterances (Suppes file; #55) to 108 utterances (Kuczaj file; Abe100).

Controlling external validity

To obtain a more representative reflection of a child's MLUm and MLUw measures, it was necessary to control variables that may influence resulting MLU scores. Retherford

Table 1 Researchers' transcripts used in the present study

Researcher	*No. transcripts used*	*No. different children included*
Bliss	1	1
Marty Demastras	4	1
Catherine Garvey	8	7
Jean Berko Gleason	5	5
Stan Kuczaj	3	1
Jacqueline Sachs	3	1
Catherine Snow	3	1
Patrick Suppes	3	1
Lori Van Houten	5	5
Anne Van Kleeck	4	4
Amye Warren-Leubecker	1	1
Total	40	28

(2000), Owens (1999) and Miller (1981) concur that there are varying factors which may affect language sample representativeness including: setting, conversational partner, tasks and topics. Bain, Olswang & Johnson (1992) indicated that children produced more different two-word combinations and with greater frequency in the low-structured situations like free-play. Bornstein, Haynes, Painter & Genevro (2000) and Olswang & Carpenter (1978) found no significant difference in MLU for a group of children between samples that were elicited by either the mother or an unfamiliar adult, or for samples gathered either at home or in the clinic. A more equal status exists between peers interacting, and the MLU may therefore be more representative of a child's language abilities (Youniss, 1980).

The length of the language sample is another important variable. In general, the literature recommends using 50–100 utterances in order to gain a representative sample (Lahey, 1988; Miller & Chapman, 1981, Miller & Chapman, 2000; Retherford, 2000). However, Brorson & Dewey (2005) demonstrated that there was no significant difference between MLUw scores regardless of the length of the transcripts. These findings suggest that the number of utterances needed to calculate MLUw is somewhat arbitrary.

In the present study, in order to increase its external validity, the above factors (setting, conversational partner, tasks and topics) were taken into account when selecting transcripts. An attempt was made to select samples where the target child was in free-play or in other low-structured situations and where the target child was interacting with a variety of different communication partners. It is interesting to note that some researchers, after taking all these given variables into account, have

concluded that it would be extremely difficult to standardize the collection procedures for language sampling (Eisenberg *et al*., 2001). The greater control of these variables strengthens the external validity of the present study.

Rationale for age range selection

Conant (1987), Miller & Chapman (1981) and Rondal *et al*. (1987) demonstrated a strong correlation between age and MLUm. This confirmation guided the selection and grouping of transcripts by age in the present study. The selected age groups could be expected to have MLUms within an early stage IV through V+ (Retherford, 2000). According to Brown's Stages (1973) children are demonstrating only occasional use of grammatical morphemes up to the age of 2;7 and have mastered the use of only five of the 14 grammatical morphemes. Within the selected age range for our study, it can be assumed that the transcripts will contain more grammatically complex structures because – according to Brown's Stages – 10 of the 14 grammatical morphemes should be mastered. The predicted MLUm ranges for the transcripts used in the study is from 3.16 to 4.40 (Miller & Chapman, 1981).

Language sample analysis

Decisions had to be made regarding utterance boundaries and determination of the middle 100 utterances, then computations of MLUw and MLUm scores were completed..

First, utterance boundaries were based on punctuation marks given within the transcripts. Guidelines exist for determining utterance segmentation (Leadholm & Miller, 1992; Miller & Chapman, 1981; Owens, 1999) such as using intonation contours, pauses greater than 2 seconds, and inhalation as utterance boundaries. However, all transcripts were in text so it was not possible to listen for the given cues for utterance boundaries.

Language transcripts that were 100 utterances or longer were collected, following the practice of using 50–100 utterances when analysing a language sample, and determined the middle utterances to be analysed (Lahey, 1988; Miller and Chapman, 2000; Retherford, 2000). Using CLAN, the *chip* command [chip + cCHI @] was executed to determine the total number of utterances in the selected samples (MacWhinney, 2000). Transcripts below 100 utterances were eliminated. On the remaining transcripts, we obtained the total number of utterances in each transcript, divided this by 2, subtracted 50, and then added 100 to find the 100 middle utterances. An option of the *columns* command [columns +h +d +t*CHI +nCHI @] was then run to allow the retrieval of the desired middle 100 utterances (MacWhinney, 2000). Accuracy checks were completed on the initial three transcripts and periodically throughout. The following were eliminated: totally or partially unintelligible utterances; utterances that consisted only of 'fillers'; obvious sentence repetitions (Retherford, 2000). Additional utterances replaced these eliminated utterances as they were available (Retherford, 2000); replacement utterances were collected from the end of the middle 100 utterances in order to complete a total of 100 utterances (Retherford, 2000). The middle 100 utterances of the selected transcripts were printed out.

MLUw and MLUm computations were completed on each of the 40 language transcripts following procedures provided by Retherford (2000: 98, 99), adapted from Brown (1973); see Appendix item 1 for list of procedures. We tallied the total number of words and morphemes after each utterance. Page totals were then tallied to find the total number of words and morphemes in the transcripts.

Inter-rater reliability of MLUw and MLUm counts was evaluated. A second rater scored a randomly selected 5 transcripts from each age group (i.e., 25% of the samples used in this study). A paired means *t*-test was used to compare the means of MLUm and MLUw of the same transcripts analysed by different raters. Discrepancies were minimal and the difference was non-significant for both MLUm and MLUw ($ps > 0.05$).

RESULTS

Since Brown (1973) introduced MLUm it has become the standard measure of child language development. Therefore, criterion validity was examined through correlation coefficients between MLUm and MLUw. The higher the correlation between MLUm and MLUw, the greater the validity of the survey measure is assumed to be (Aday, 1996). Finding of high correlations between MLUw and MLUm measures would suggest that MLUw could be used as effectively as MLUm as measurement of gross language development.

Bivariate correlation was calculated using Pearson's *r*. Variables were defined as MLUw (the total number of words divided by number utterances analysed in the language transcript) and MLUm (the total number of morphemes divided by number utterances analysed in the language transcript). Age groups were represented as 0 (3;0–3;5) and 1 (3;6–3;11). The correlation between MLUm and MLUw was almost perfect, $r = 0.998$, $p < 0.001$. Age was also strongly correlated to MLUm ($r = 0.69$) and MLUw ($r = 0.69$).

DISCUSSION

Some researchers have found difficulties implementing MLUm counts for reasons including: inflection differences that exist across languages; arbitrary decisions regarding productivity of morphemes; and dialectal differences. Studies have attempted to alleviate these difficulties by using MLUw counts instead of MLUm (Arlman-Rupp, 1976; Hickey, 1991; Malakoff *et al.*, 1999; Thordardottir & Weismer, 1998). MLUw has been shown to be strongly correlated with MLUm in several languages other than English, including Dutch (Arlman-Rupp, 1976), Irish (Hickey, 1991), and Icelandic (Thordardottir & Weismer, 1998). All the above have concurred that MLUw measures are simpler to implement, faster, more reliable, and are less arbitrary in nature than MLUm measures.

The question posed in the present study was whether MLUw is correlated with MLUm measures for normally-developing, English-speaking children between the ages

of 3;0 and 3;10. Results clearly demonstrated that MLUw and MLUm measures are, in fact, strongly correlated across the age range studied. This correlation suggests that MLUw could be used as effectively as MLUm in the measurement of a children's gross language development. In the present data, a coefficient of 1.108 indicates the basis for conversion of MLUw to MLUm. For example, a MLUw score of 3.78 could be converted to a MLUm score of 4.19.

Of Brown's (1973) 14 grammatical morphemes, only seven, when used productively, will stand as additional morphemes when completing MLUm measures. The strong correlation that exists between MLUm and MLUw may be attributed to the relatively few grammatical morphemes used by young children that impact MLU scores. Furthermore, some researchers have suggested that the actual opportunities to use grammatical morphemes during the collection of a spontaneous speech sample occur infrequently (Balason & Dollagham, 2002; Lahey, 1994). By demonstrating the strong correlation between MLUw and MLUm in this investigation, further evidence is provided that, although a child may possess the grammatical skills to use the grammatical morphemes productively, the obligatory context did not occur frequently enough during the collection of the transcripts to increase MLUm measures to a level that was significantly different from MLUw.

Our results indicate that MLUw is as a reliable measure of a child's structural development as MLUm, and it can therefore be used as reliably as MLUm. The use of MLUw in place of MLUm has many potential benefits for child language researchers and therapists. First, rules (see Appendix) that are necessary for computing MLUm can be eliminated, because all words will be counted as such. When using MLUw the investigator would not have to make arbitrary decisions about whether morphemes are used productively, such as determining whether a child's usage of a negative contraction 'don't' should be assigned two morphemes (according to Brown (1973), it is necessary to search through the transcript to see if the child used 'do' and 'not' separately). However, if an investigator is counting only the number of words, this step is unnecessary. The elimination of these arbitrary decisions will increase reliability and agreement in issues related to the scoring of MLU, because *ad hoc* decisions will be minimized. Furthermore, MLUw can be used more readily across languages and dialects. Using MLUw will diminish concerns regarding the inflation of MLU scores in highly inflected languages (Arlman-Rupp, 1976; Dromi & Berman, 1982; Hickey, 1991; Thordardottir & Weismer, 1998), and will also alleviate concerns about morphemic differences between Ebonics and standard English (Malakoff *et al*., 1999). Also, with the increased case-load sizes in the schools, MLUw indexes are more appealing to clinicians because they can be calculated faster. MLUw also provides a faster way s to document structural development.

In conclusion, the results of this investigation suggest that MLUw can be used as effectively as MLUm in the measurement of gross language development. MLUw is a more reliable measure of gross language development than MLUm due to the fact that arbitrary decisions regarding morpheme assignment are eliminated. Furthermore, MLUw is a more effective measurement as it can be used more readily and reliably across various languages.

APPENDIX

Rules for counting morphemes (Retherford, 2000)

1. Fillers are not assigned morphemes (*um*, *well*, *oh*).
2. Compound words and closely related are assigned 1 morpheme (*all gone*, *bye-bye*).
3. Indefinite and reflexive compound pronouns are assigned 1 morpheme (*herself*, *anything*).
4. Proper nouns and ritualized reduplications are assigned 1 morpheme (*Mr. Smith*, *night-night*).
5. Diminutive forms of words received 1 morpheme (*funny*, *doggie*).
6. Auxiliary verbs are assigned 1 morpheme.
7. Catenative forms are assigned only 1 morpheme (*gonna*, *wanna*, *hafta*).
8. Inflectional affixes (i.e. plural *–s*, singular and plural possessive *–s*, present third person singular *–s*, regular past tense *–ed*, and *–en*, present participle *–ing*, comparative *–er*, and superlative *–est*) are assigned morphemes.
9. Incorrect uses of inflection are not counted as separate morphemes.
10. Irregular past tense and past participle forms are assigned only 1 morpheme.
11. Negative contractions are assigned two morphemes if there was evidence within the transcript that the child used each part of the contraction separately. If the child does not use each part of the contraction separately 1 morpheme is assigned.
12. All nonnegative contractions are assigned 2 morphemes.
13. Common derivational affixes are assigned there own morphemes.

REFERENCES

Aday, L. A. (1996). *Designing and conducting health surveys: a comprehensive guide* (2nd ed.). San Francisco, CA: Jossey-Bass Publisher.

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

Bain, B., Olswang, L. B. & Johnson, G. A. (1992). Language sampling for repeated measures with language impaired preschoolers: Comparison of two procedures. *Topics in Language Disorder*, *12*, 13–27.

Balason, D. V. & Dollaghan, C. A. (2002). Grammatical morphemes production in four-year-old children. *Journal of Speech, Language, and Hearing Research*, *45*, 961–969.

Bernstein, D. K. & Tiegerman-Farber, E. (1997). *Language and communication disorders in children* (4th ed.). Boston, MA: Allyn & Bacon.

Bornstein, M. H., Haynes, O. M., Painter, K. M. & Genevro, J. L. (2000). Child language with mother and with stranger at the home and in the laboratory: A methodological study. *Journal of Child Language*, *27*, 407–420.

Brorson, K. & Dewey, C. (2005). Effect of language sample size on MLUw. *Hearsay*, *17*, 46–56.

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

Butler, K. (1992) From the Editor. *Topics in Language Disorder*, *12*, iv–v.

Chapman, R. S. (1985). Computing mean length of utterance in morphemes. In J. F. Miller, *Assessing language production in children: Experimental procedures* (pp. 22–35). Baltimore, MD: University Park 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.

DeVellis, R. F. (2003). *Scale development: Theory and applications*: *Applied social research methods* (2nd ed.). Thousand Oaks, CA: Sage Publications.

DeVilliers, J. G. & DeVilliers, P. A. (1973). Development of the use of word order in comprehension. *Journal of Psycholinguistic Research*, *2*, 331–341.

Dromi, E. & Berman, R. A. (1982). A morphemic measure of early language development: Data from modern Hebrew. *Journal of Child Language*, *9*, 403–424.

Eisenberg, S. L., 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.

Fey, M. E. (1986). *Language intervention with young children.* Needham Heights, MA: Allyn & Bacon.

Hickey, T. (1991). Mean length of utterance and the acquisition of Irish. *Journal of Child Language*, *3*, 553–569.

Hughes, D. L., Fey, M. E. & Long, S. H. (1992). Developmental sentence scoring: Still useful after all these years. *Topics in Language Disorder*, *12*, 1–12.

Hux, K., Morris-Friehe, M. & Sanger, D. (1993). Language sampling practices: A survey of nine states. *Language, Speech and Hearing Services in Schools*, *13*, 42–52.

Johnston, J. R. (2001). An alternate MLU calculation: Magnitude and variability of effects. *Journal of Speech, Language, and Hearing Research*, *44* (1), 156–145.

Kahn, J. (2004). Reporting statistics in APA Style. Retrieved 15 April 2004, from http://www.ilstu.edu/~jhkahn/apastats.html.

Kemp, K. & Klee, T. (1997). Clinical language sampling practices: Results of a survey of speech-language pathologist in the United States. *Child Language Teaching and Therapy*, *13*, 161–176.

Klee, T. (1992). Developmental and diagnostic characteristics of quantitative measures of children's language production. *Topics in Language Disorder*, *12*, 28–41.

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.

Lahey, M. (1988). *Language disorders and language development.* New York: MacMillan.

Lahey, M. (1994). Grammatical morpheme acquisition: Do norms exist? *Journal of Speech and Hearing Disorders*, *37*, 1192–1194.

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

Long, S. H. & Channel, R. W. (2001). Accuracy of four language analysis procedures performed automatically. *American Journal of Speech-Language Pathology*, *10*, 180–188.

MacWhinney, B. (2000). *The CHILDES project: tools for analyzing talk.Vol 2: The database* (3rd ed.). Mahwah, NJ: Erlbaum.

MacWhinney, B. (2002). Out of the baby book and into the computer: Child language research comes of age. *APA Review of Books, 47*, 391–394; see also *Journal of Speech and Hearing Disorders*, *45*, 559–563.

Malakoff, M. E., Mayes, L. C., Schottenfeld, R. & Howell, S. (1999). Language production in 24-month-old inner-city children of cocaine-and-other-drug-using mothers. *Journal of Applied Developmental Psychology*, *20*, 159–180.

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

Miller, J. & Chapman, R. (1981). The relationship between age and mean length of utterance in morphemes. *Journal of Speech and Hearing Research*, *24*, 154–161.

Miller, J. & Chapman, R. (2000). *Systematic Analysis of Language Transcripts (SALT)*. Madison, WI: University of Wisconsin-Madison Waisman Center, Language Analysis Laboratory.

Neuman, W. L. (2003). *Social research methods: Qualitative and quantitative approaches* (5th ed.). Boston, MA: Pearson Education, Inc.

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

Olswang, L. B. & Carpenter, R. L. (1978). Elicitor effects on the language obtained from young language-impaired children. *Journal of Speech and Hearing Disorders*, *43*, 76–88.

Owens, R. E. (1999). *Language disorders: A functional approach to assessment and intervention* (3rd ed.). Needham Heights, MA: Allyn & Bacon.

Paul, R. (2000). *Language disorders from infancy through adolescents* (2nd ed.). Saint Louis, MO: Mosby-Year Book.

Retherford, K. S. (2000). *Guide to analysis of language transcripts* (2nd ed.). Eau Claire, WI: Thinking Publications.

Rondal, J. A., Ghiotto, M., Bredart, S. & Bachelet, J. F. (1987). Mean length of utterance of children with Downs Syndrome. *American Journal on Mental Retardation*, *93*, 64–66.

SPSS, Inc. (2002). *SPSS Computer Software, SPSS for Windows v 11.0*. Chicago, IL: SPSS, Inc.

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

Winks Statistics Software (2002). Retrieved 18 April 2004, from http://www.texasoft.com/winkpair.html.

Winstead, R. L. (2002). Null hypothesis and 'p value'. Retrieved 19 April 2004, from http://www.nsm1.nsm.iup.edu/rwinstea/null-p.shtm.

Youniss, J. (1980). *Parent and peers in social development*. Chicago: University of Chicago Press.

ADDRESS FOR CORRESPONDENCE

Dr Kent Brorson
Department of Communication Sciences and Disorders,
University of Minnesota Duluth, 1207 Ordean Court, Duluth, MN 55812, USA
E: kbrorson@d.umn.edu