1984

Comparison of an on-line procedure for MLU calculation with the traditional MLU procedure.

Kari L. Morris
The University of Montana

Let us know how access to this document benefits you.
Follow this and additional works at: https://scholarworks.umt.edu/etd

Recommended Citation
https://scholarworks.umt.edu/etd/7197

This Thesis is brought to you for free and open access by the Graduate School at ScholarWorks at University of Montana. It has been accepted for inclusion in Graduate Student Theses, Dissertations, & Professional Papers by an authorized administrator of ScholarWorks at University of Montana. For more information, please contact scholarworks@mso.umt.edu.
COPYRIGHT ACT OF 1976

This is an unpublished manuscript in which copyright subsists. Any further reprinting of its contents must be approved by the author.

Mansfield Library
University of Montana
Date: 1984
A COMPARISON OF AN ON-LINE PROCEDURE
FOR MLU CALCULATION WITH THE
TRADITIONAL MLU PROCEDURE

By

Kari L. Morris
B.A., University of Northern Iowa

Presented in partial fulfillment of
the requirements for the degree of
Master of Arts
UNIVERSITY OF MONTANA
1984

Approved by:

[Signature]
Chairman, Board of Examiners

[Signature]
Dean, Graduate School

[Date]
The purpose of this study was to compare an on-line procedure for determining MLU with the traditional procedure of transcribing utterances from a video-taped recording. Both the intra-clinician accuracy level and the differences in the amount of time involved were investigated.

Thirteen graduate students in communications sciences and disorders were subjects; ten received on-line training and three were a control group. The on-line training provided practice counting morphemes on-line while viewing video-taped language samples of normal children. Eight of the ten subjects who participated in the on-line training passed exit criterion of accurately obtaining two consecutive on-line MLUs and participated in the experimental condition. In the experimental condition subjects viewed tapes of three children and determined each child's MLU on-line. Next the same three tapes were scored using the traditional transcription method. The MLU scores obtained for each method were compared. Additionally, the time required to complete each method was also compared. The control group obtained MLUs traditionally for each of the three tapes and their scores and times were compared to those of the subjects.

Results indicate that there were no significant differences between the on-line and traditional MLU scores obtained by the subjects. The time, however, was significantly less for the on-line method. No significant differences were found between the subjects and the control group in either MLU scores or amounts of time involved. Results suggest that using an on-line method for obtaining MLU can increase the speech and language clinician's efficiency and effectiveness in the clinical setting.
## TABLE OF CONTENTS

ABSTRACT .............................................................................. ii

Chapter

I. INTRODUCTION .......................................................... 1

   Background Information and Rationale ......................... 2
   Observational Measures ............................................... 2
   Observational Measures in Speech and Language Pathology .. 4
   Efficiency of Speech and Language Measures ............... 6
   Productive Language Measures ................................. 8
   Mean Length of Utterance ...................................... 9
   Purpose of the Study ............................................. 17

II. METHODS ..................................................................... 18

   Subjects ..................................................................... 18
   Sample Tapes ........................................................... 18
   Training Program Procedures .................................. 20
   Experimental Condition ......................................... 23
   Measurement and Design ......................................... 24

III. RESULTS ................................................................... 26

   Entrance Criterion .................................................. 26
   On-line Training and Exit Criterion ........................... 26
   Experimental Condition ......................................... 28
   Summary ............................................................. 35

IV. DISCUSSION .............................................................. 36

   Clinical Implications ............................................... 36
   Future Research ..................................................... 39
   Summary ............................................................. 43

REFERENCES ................................................................. 44

APPENDICES ................................................................. 48
LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Brown's target values and upper bounds for linguistic stages</td>
<td>11</td>
</tr>
<tr>
<td>2</td>
<td>MLU stages outlined by de Villiers and de Villiers (1973b) and Miller and Chapman (1981)</td>
<td>12</td>
</tr>
<tr>
<td>3</td>
<td>Number of trails required to pass on-line training</td>
<td>27</td>
</tr>
<tr>
<td>4</td>
<td>On-line and traditional MLU scores</td>
<td>29</td>
</tr>
<tr>
<td>5</td>
<td>Stage assignments based on on-line and traditional scores</td>
<td>31</td>
</tr>
<tr>
<td>6</td>
<td>Time ranges and means for each method in minutes and seconds</td>
<td>34</td>
</tr>
</tbody>
</table>
ACKNOWLEDGEMENTS

Thanks and appreciation are extended to the two people responsible for directing me through this project; Chris Dollaghan for her help in the creation of this study and Barb Bain for her help in the completion of it. A special thanks goes to all of the graduate students in the University of Montana Communication Sciences and Disorders Department who so graciously gave of their time to participate as subjects, without them this study would not have been possible.
CHAPTER I

INTRODUCTION

When planning an assessment of a child's language skills the issue of efficient use of time, both in the gathering of data and in the analysis of these data, plays an important role in the choice of tests and measurement procedures to be used. A popular means of assessing a child's productive language is to analyze a spontaneous speech sample. Such a speech sample can be obtained by recording an adult-child interaction, transcribing the child's utterances and then analyzing these utterances (Shriner, 1969). A number of different procedures are designed for the analysis of these utterances. The amount of time involved in analysis can vary from a few minutes to several hours (Emerick & Hatten, 1979), depending on the analysis procedure chosen.

The analysis of spontaneous speech samples has several advantages which aid in the diagnosis of language (i.e. provides representative data on a child's expressive language skills; Bloom & Lahey, 1978), but some clinicians think that an overriding disadvantage of the time required to do this analysis limits the use of this procedure. The present study was designed to maintain advantages of spontaneous speech sample analysis while reducing the amount of time involved. To do this, the transcription step of a widely used measure of productive language (Mean Length of Utterance) was eliminated and observers were trained to obtain this measure "on-line", as the behavior occurred. The amount of time saved, as well as the accuracy with which observers were able to
perform this on-line task was investigated and compared with the more traditional means of recording, transcribing and analyzing the speech sample.

BACKGROUND INFORMATION AND RATIONALE

Background information has been divided into three main areas. The first concerns observational measures in general and how such measures have been used in speech and language pathology. Next the issue of efficient use of time in speech and language evaluations is discussed. The third area involves general measurement procedures used to evaluate productive language skills in children, and specifically Mean Length of Utterance (MLU).

Observational Measures

Methods for obtaining data on human behavior include controlled laboratory experiments, the use of standardized tests or questionnaires and direct observational measures (Sackett, 1978). MLU is considered to be a direct observational measure and therefore the advantages and disadvantages of such measures are presented.

Observational measures, as discussed by Sackett (1978) involve "measuring behavior when subjects are free to vary their individual and social responses in a large number of ways with few or no constraints imposed by the investigator" (p.2). Such measures should be used when automated measuring techniques are not available or when information from standardized tests is not appropriate (Sackett, 1978), as is often
the case in speech and language evaluations.

The use of observational measures introduces problems of observer reliability and bias that are minimized with use of automated instruments and standardized tests (Sackett, 1978). Despite the problems of reliability, observational measures are often more valid than other methods and considered by some to be the preferred method of obtaining data in certain situations (Bloom & Lahey, 1978; Sackett, 1978; Emerick & Haffen, 1979). They have the advantage of allowing the observer to judge a subject's abilities in "natural" contexts (Sackett, 1978; Miller & Dollaghan, in press). Sackett suggests that observational measures should be used when standardized tests or mechanical methods are unable to generate sufficient data. For example, when studying complex human behavior, the instruments currently available can provide only a limited amount of information when compared to the amount of information that can be gained by trained observers.

Observational data can be gathered by having the observer code behavior on-line, as it occurs, or by using video and/or audio tape recordings (Sackett, 1978). Observational measures rely on the use of human sense receptors and judgment abilities which have limitations in terms of reaction time and the number of discriminations that can be made in a short period of time (Sackett, 1978). These limitations must be considered when requiring observers to code behaviors on-line.
Many authors have been concerned with the issue of maximizing the reliability of observational measures. Duncan & Fiske (1977) suggest that one way of doing this is to narrow the observers' focus of attention and have them make discrete, moment-to-moment decisions. Rosenthal (1963) believes that eliminating sources of observer bias such as errors of omission, miscoding and those resulting from expectancies are ways of increasing the reliability of observational measures. Miller & Dollaghan (in press) suggest that to maximize the reliability of observational measures, the behaviors under study should be carefully defined and the observation of these behaviors should be well organized. Additionally, they discuss the importance of minimizing subjective input of the observers. Despite the reliability problems associated with observational measures such measures can be designed to give accurate results if caution is taken to guard against these problems both in the development of the measure and in the training of observers.

Observational Measures in Speech and Language Pathology

Observational measures have been devised and used as research tools in many areas of developmental research, and have been particularly important in research concerned with documenting language acquisition. "In few other areas of developmental psychology [as in developmental psycholinguistics] are observational measures so firmly established as the central methodology" (Dale, 1978. p. 219).
Speech and language clinicians have used observational measures both as tools for research and as aids in the evaluation of children's productive language skills. Many researchers interested in documenting normal language acquisition have chosen not to rely on the use of any standardized test, and therefore observational measures have been the major means for collecting data on early language acquisition (Dale, 1978), as is illustrated by the seminal longitudinal observational studies of Brown (1973), Bowerman (1973), and Bloom (1973).

Assessment procedures available for the evaluation of language behavior can be divided into four categories: standardized tests, developmental scales, nonstandardized tests and behavioral observations (Miller, 1978). The speech and language clinician must consider many factors when deciding which type of procedure or combination of procedures will be used in any given language evaluation. For example, the age of the child can influence the clinician's choice of assessment procedures. Rice (1978) states that standardized tests may not be the most appropriate means to assess young children's language skills because the children may be too young or unfamiliar with a testing situation to perform adequately. Therefore, the behavior sampled may not be representative of the child's skills and consequently the clinician may increase errors in clinical judgement or management.

Most tests of language abilities elicit only fragmented language samples and may represent only a fraction of the child's repertoire of responses to his environment (Emerick & Hatten, 1979). According to Miller (1978), the more structure or constraints that are imposed on a
child the less varied his language will become in terms of variety of structures used. Although a spontaneous speech sample may not provide the clinician with an opportunity to observe the child's entire repertoire of linguistic knowledge, it is less artificial and constraining than are standardized testing procedures and may provide the clinician with more representative data on a child's expressive language abilities (Dale, 1976; Bloom & Lahey, 1978).

A major advantage of standardized tests is that they can be administered and scored in a relatively short period of time (Emerick & Hatten, 1979), especially when compared to the time-consuming procedures of obtaining and analyzing a spontaneous speech sample. If the amount of time spent in spontaneous speech sample analysis, considered by some to be more valid than standardized tests, could be reduced perhaps more clinicians would consider using such procedures.

Efficiency of Speech and Language Measures

The efficient use of time in a speech and language evaluation is a major concern to many clinicians. To date the concept of efficiency in the field of speech and language pathology has not been well defined or researched. "Efficiency" can be defined as "the capacity to produce desired effects with a minimum of effort, expense, or waste" (Webster's New World Dictionary, 1957). Efficiency as discussed in reference to a statistical estimator can only be determined when it is compared to another estimator; according to Marascuilo (1971), efficiency is a relative concept.
The concept of an "efficient measurement procedure" for purposes of this study is drawn from both of these definitions. The speech and language clinician must consider and weigh the amount of time that is required to complete a measurement procedure as compared to the quantity and quality of the information that it provides. To help determine whether a particular measurement procedure is "efficient" the clinician must have a reference point with which to compare it (i.e. a previously used procedure designed to measure the same aspect of communication).
**Productive Language Measures**

The characteristics of the normal developmental sequence have been documented for three major areas of language production: syntax, semantics and phonology (Miller, 1981). Because chronological age has been found to be an unsatisfactory indicator of a child's linguistic ability (Bowerman, 1973; Brown, 1973), measurement procedures have been developed in each area which provide specific information that is used in determining linguistic abilities. Dromi and Berman (1982) state that such linguistic indicators have been needed for research and clinical use to compare groups to one another, to compare a subject to a group, and to document linguistic change over time.

Syntax is the area which has been best documented and therefore a majority of language measures have been designed to measure this aspect of language development (Klee & Paul, 1981). A number of procedures have been developed to measure expressive syntax using sentence length and/or complexity as an index of development; some of these measures include mean length of response (Nice, 1925; McCarthy, 1930; Templin, 1957; Darley & Moll, 1960); mean of five longest responses, number of one word responses and number of independent clauses (McCarthy, 1954); length-complexity index (Shriner & Sherman, 1967; Miner, 1969); developmental sentence score (Lee & Canter, 1971; Lee, 1974); and mean length of utterance (Brown, 1973; de Villiers & de Villiers, 1973a).
These measures differ in their depth of analysis as well as in the amount of time required to complete them. Miller (1981) suggests that prior to the initiation of a time-consuming, detailed linguistic analysis, the completion of a general measure that quantifies relatively quickly the entire sample is beneficial. Initially completing such a general measure can help the clinician in determining which of the more detailed analyses could provide the most useful information. One measure recommended by Miller (1981) to serve as this general indicator is Mean Length of Utterance in morphemes, which is one of the most widely used general measures of syntactic complexity (Dale, 1976).

Mean Length of Utterance

Historically, the length of children's utterances has been used to help determine a child's expressive linguistic maturity. Originally length was measured in number of words per utterance, and the average length of utterances for a sample was reported as the Mean Length of Response (MLR), (Nice, 1925; McCarthy, 1930; Davis, 1937; Templin, 1957). Brown (1973) refined the measure of MLR by introducing the idea of counting morphemes rather than words. This new measure was called Mean Length of Utterance, and the following sections review this measure.

Definition: A morpheme is defined as the smallest meaningful unit of language (Dale, 1976) and consists of an entire word or an inflectional or derivational marker added to a word which changes its meaning (i.e. plural (s), past tense (ed). etc.). Brown (1973) introduced the idea of counting morphemes rather than words because this
resulted in a more detailed reflection of increased utterance length due to the additional use of any new kind of grammatical structure (i.e. addition of obligatory morphemes, addition of negative forms and auxiliary, etc.).

**MLU Stage:** Brown's (1973) longitudinal observational study of the language development of three subjects occurred during a period when their MLUs ranged from 1.5 to 4.25. In the analysis of his data he arbitrarily divided this developmental range into five equidistant stages. He based these stages on MLU and "upper bound" (the longest utterance spoken by the child at a particular stage) and named these stages for the major new process that emerged during that period. (Appendix A). He originally chose specific MLU points to define each stage (Table 1), but found later that using MLU ranges was a more realistic way of describing these stages.
Table 1. Brown's target values and upper bounds for linguistic stages.

<table>
<thead>
<tr>
<th>Stage</th>
<th>MLU</th>
<th>Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>1.75</td>
<td>5</td>
</tr>
<tr>
<td>II</td>
<td>2.25</td>
<td>7</td>
</tr>
<tr>
<td>III</td>
<td>2.75</td>
<td>9</td>
</tr>
<tr>
<td>IV</td>
<td>3.50</td>
<td>11</td>
</tr>
<tr>
<td>V</td>
<td>4.00</td>
<td>13</td>
</tr>
</tbody>
</table>

Taken from Brown, 1973 (p. 56)

De Villiers & de Villiers (1973b), in their research investigating the development of the comprehension of word order, used Brown's MLU stages to divide their subjects into groups. They found some "important discontinuities in performance" for the subjects in MLU stages I and IV and therefore further divided these two stages. Miller and Chapman (1981), in their work with MLU and chronological age made an additional division in Stage V. Table 2 outlines these subdivisions made on Brown's original five MLU stages.
Table 2. MLU stages outlined by de Villiers and de Villiers (1973b) and Miller and Chapman (1981).

<table>
<thead>
<tr>
<th>Stage</th>
<th>MLU Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early Stage I</td>
<td>1.0 - 1.49</td>
</tr>
<tr>
<td>Late Stage I</td>
<td>1.5 - 1.99</td>
</tr>
<tr>
<td>Stage II</td>
<td>2.0 - 2.49</td>
</tr>
<tr>
<td>Stage III</td>
<td>2.5 - 2.99</td>
</tr>
<tr>
<td>Early Stage IV</td>
<td>3.0 - 3.49</td>
</tr>
<tr>
<td>Late Stage IV and Early Stage V</td>
<td>3.5 - 3.99</td>
</tr>
<tr>
<td>Late Stage V</td>
<td>4.0 - 4.49</td>
</tr>
<tr>
<td>Post Stage V</td>
<td>4.5+</td>
</tr>
</tbody>
</table>

Clinical Uses: Emerick & Hatten (1979) state that an underlying assumption in measuring MLU is that an increase in the complexity of a child’s utterances is reflected by an increase in the length of his utterances, as well. This is true when a child’s MLU falls between 1.0
and 4.5, but is not necessarily true when it exceeds 4.5 (Brown, 1973). Therefore, MLU as an indicator of syntactic development is most beneficial when it falls between 1.0 and 4.5 (Miller, 1981).

Dromi and Berman (1982) have outlined three properties that a general indicator of linguistic abilities should exhibit, and recommend the use of MLU because it meets these properties. The first is, "it should be easily and readily scored on a relatively small sample of spontaneous speech". Brown (1973) and de Villiers & de Villiers (1973b) agree that MLU is easily scored, and McCarthy (1930) states that using a speech sample of 50 utterances gives a "fairly representative sample of the child's linguistic development in a relatively short period of time, without tiring the child with a prolonged observation" (p. 32).

The second property Dromi & Berman discuss is "it should be fairly constant across different testing situations". One objection to the use of measures of spontaneous speech samples, as discussed by Dale (1976), is that results from such measures are "highly dependent upon the situation" in which the speech sample was obtained. However, he suggests that MLU is less affected by the situation than other measures because it considers all utterances in a sample instead of looking only at utterances of a particular construction, as other syntactic measures do.

The third property Dromi and Berman report is that "it should predict or correspond to other more complex or in-depth measures of linguistic abilities". Although controversy has existed for years concerning the value of the information an MLR/MLU provides (Emerick &
Hatten, 1979), most authors agree that MLU is one of the best general indices of a child's early productive language skills (de Villiers & de Villiers, 1973b; Brown, 1973; Dale, 1976; Bloom, 1978; Miller, 1981) and in fact during the past ten years researchers have demonstrated that the development and use of certain linguistic structures correspond highly to given MLU Stages.

Brown (1973) and de Villiers & de Villiers (1973a) have outlined the normal acquisitional sequence of fourteen grammatical morphemes in relation to the MLU stage in which they emerge. Paul (1981) studied language samples of 59 children between the ages of 2;5 and 6;11 and found MLU to be a better predictor of use of most types of complex sentences than either chronological age or cognitive level. Additionally, Miller (1981) has compiled the works of several authors and outlined the structures of noun and verb phrase elaborations, negation, yes/no and "wh" question development which one would expect to see at each MLU stage.

MLU has been used as a measure which provides the clinician with a general idea about the language structures that normally develop at any given stage. Determining a child's MLU and then evaluating the language structures normally seen at his linguistic stage has been used as a means to help determine whether a language impaired child's expressive skills are delayed or deviant (Leonard, 1979) Is the child developing language in the order normally seen but at a rate slower than most children, or is he "doing something atypical for that period of linguistic development?" (Leonard, 1979. p.207).
More recently, the results of a study by Klee and Fitzgerald (in press) questioned the validity of using an MLU score to predict grammatical complexity in children past Brown's Stage II. They found that frequency and diversity of grammatical morpheme usage increased predictably with increased MLUs, but no statistically significant differences were found in clause and phrase level grammatical organization. They concluded that too much weight has been given to using MLU as a predictor of grammatical complexity, but that it continues to have value as a screening instrument; especially when used with Miller and Chapman's (1981) data on MLU and age in normal children.

Whether using MLU as a screening device to isolate children needing further syntactic evaluation or using it as a general indicator of language complexity, obtaining an MLU score early in a language evaluation would be beneficial. Present procedures prevent the clinician from obtaining this information immediately during the evaluation.

**Traditional Procedures:** The traditional procedure for obtaining an MLU involves video and/or audio tape recording an adult-child interaction, transcribing 50 of the child's utterances (McCarthy, 1930), counting the morphemes in each utterance and calculating the average length in morphemes. The time spent interacting with a child to elicit 50 utterances usually takes from 15 to 30 minutes, depending on how verbal the child is as well as the clinician's skill in eliciting language from the child. The time required to transcribe these utterances following the evaluation takes at least as long as the
original interaction and usually longer because the tape may be stopped and/or replayed while an utterance is transcribed. Hadian (1978) states that transcribing a 50 utterance sample "usually takes from 15 to 60 minutes depending on the language being transcribed" (p. 69). Following the transcription step, time still is required to count the morphemes. The entire time spent in determining an MLU using this traditional means is estimated to range from 45 minutes to 1 and 1/2 hours per child.

If the transcription step could be eliminated from the MLU procedure, the time required to obtain an MLU could be reduced by at least 50%. Additionally, a quick determination of MLU at the start of an evaluation could help the clinician to more efficiently use the remaining portion of the evaluation. By knowing a child's MLU stage, the clinician could set up situations designed to elicit the language structures expected to be present in that stage.

Without knowing an MLU stage during the evaluation, clinicians have had to do one of two things; both of which are very time consuming. One option has been to test the development of all language structures (i.e. grammatical morphemes, structure of questions, etc.). A second choice has been to take a lengthy language sample and look for all types of structures to occur naturally. Not only does this require a long analysis of the data following an evaluation, but there is also some question as to the validity of the results. If a language sample does not contain a certain structure the clinician then has to determine whether this is because there was no opportunity for the child to use
it, or whether that structure is not part of the child's expressive language.

**Purpose of the Study**

The preceding review of literature has demonstrated that observational measures of spontaneous speech samples, and more specifically the determination of MLU, are useful methods in the evaluation of children's expressive language skills. The major drawback of such measures is that they are not as efficient as some of the other measurement techniques currently available in the field. In an effort to make the use of MLU more efficient, the present study was designed to answer the following research questions:

1. **Is there a significant difference in score between on-line MLUs and those obtained using the traditional method for subjects who have participated in a training program designed to teach the on-line method?**

2. **Is there a significant difference in the amount of time expended to obtain an MLU between the on-line method and the traditional means of taping and transcribing the utterances prior to counting morphemes?**
METHODS

Subjects

The subjects were eight graduate students in the Communication Sciences and Disorders department at the University of Montana. At some point in their education, all had previously been trained in MLU calculation. Their experience ranged from practice exercises for a class project to computation of 10-15 MLUs for clients. Prior to their participation in the MLU training program all subjects passed an audiological screening (ASHA, 1975) or reported having normal hearing based on hearing tests received within the previous six month period.

Sample Tapes

Thirteen video tapes were prepared as stimuli for the present study; ten for training and three for the experimental condition. The tapes were recorded using a Sony Videocorder AV-3400 portable video recorder, a Sony AVC-3450 camera, a Sony Cardiod microphone and 3/4 inch Sony and Scotch brand reel to reel video tapes.

All tapes contained adult-child interactions in play situations. The adult in each interaction was either the examiner interacting with a child to obtain a spontaneous speech sample using procedures similar to those outlined by Lund and Duchan (1983), (Appendix B) or one of the child's parents "playing" with him or her. Each tape contained at least 60 child utterances and ranged from 9 minutes and 5 seconds to 17 minutes in length (Appendix C).
Stimuli tapes were chosen based on the following criteria. The children represented a variety of ages (1 year 11 months to 4 years 3 months) and a variety of linguistic stages (MLUs 1.0 to 5.0); (Appendix C). Speech samples were judged to be intelligible using the Percentage Consonants Correct,(PCC) (Shriberg & Kwiatkowski, 1982); (Appendix D). Samples with PCC scores in the mild (85 to 100) and moderate (65 to 85) categories were used as stimuli. Approximately three seconds or more had to separate the child's utterances (filled with an adult utterance or a pause).

Using the traditional method the examiner calculated the MLU for the child in each tape to be used as standards in the training program. A second observer was used to establish interexaminer reliability. Three tapes were randomly chosen to be individually scored by the second observer. Differences in MLU between the two scorers were .00, .06 and .08; these indicate a 97 to 100% agreement between the two observers on the overall measure. Additionally, a point by point comparison of these data indicated 73 - 78% agreement between the two observers.
**Training Program Procedures**

The training program designed to teach subjects to obtain MLUs on-line was divided into four sections: reviewing MLU computation and segmenting rules, establishing entrance criterion, training and practice obtaining MLUs on-line, and meeting exit criterion.

**Reviewing procedures:** Since all subjects had prior knowledge of the procedures involved in calculating MLU, this portion of the training program was designed to review these procedures, as well as the rules for segmenting speech samples into utterances. First each subject was provided with a handout outlining MLU computation rules (Chapman, 1981; Appendix E) and segmenting rules (Johnson, Darley & Spriestersbach, 1963; Appendix F). These rules were discussed and thirty utterances (15 oral and 15 written) were presented for review (Appendix G). Each of the oral examples was discussed until a consensus of length in morphemes was reached between each subject and the investigator. Next each subject scored a written transcript of 15 utterances (Appendix H), and discrepancies were discussed until a decision was reached.

**Establishing Entrance Criterion:** Prior to participating in the training program subjects were required to pass the entrance criterion of accurately computing an MLU using the traditional method. Each potential subject individually viewed a video tape of an adult-child interaction, transcribed 50 utterances and calculated the child's MLU. No time limit was imposed so that each observer viewed the tapes as many times as was needed. After completing this task each observer submitted her transcription and calculated MLU to the examiner. Those with MLU
scores $\pm .2$ morphemes from the previously determined MLU for that sample met the entrance criterion and began the on-line training portion of this study.

In discussing the representativeness of samples, Chapman (1981) suggests that a difference of .5 morphemes constitutes a significant difference between two language samples. The criterion of $\pm .5$ was judged by the examiner to be too lenient for this study, based on the range of MLU scores considered to make up any given stage. If observers were allowed to be .5 morphemes from a child's true MLU, children would be frequently assigned to an inappropriate linguistic stage. By allowing a variance of $\pm .2$, such inappropriate MLU stage assignments would be greatly reduced.

Training for the On-line Method: This portion of the training program was designed to teach the potential subjects an on-line method of computing MLU and to give them practice using this method. In a pilot study, the examiner discovered that individual scorers differed in the specific way that they were best able to tally morphemes on-line (i.e. recording the number of morphemes in each utterance, or making a mark for each morpheme heard). Therefore, the first step in the training was to provide subjects practice using different scoring procedures. Subjects listened to an audio tape recording of an adult reading two sets of sentences of varying lengths with pauses of approximately three seconds separating utterances, and recorded the number of morphemes heard in each utterance using one method for each set. They then chose the method they thought was most effective.
Next, the subjects practiced using the scoring system they had chosen with a sample of 50 utterances produced by an adult. An audio recording was made of two adults reading a portion of a transcript of an adult-child interaction (Miller, 1981; pp.102-105). Using a form containing 50 blank spaces, the subjects counted the morphemes of the speaker who was reading the child's part. Each subject then calculated an MLU using her data and compared it to the previously determined MLU.

Next observers viewed the training video tapes and calculated MLUs using the on-line method. To allow the observers to become familiar with the child's manner of speaking prior to counting morphemes, the observers began by viewing a segment of tape containing ten of the child's utterances. After each utterance in this section the tape was stopped and the observer and the investigator discussed the number of morphemes present. Observers then viewed the remainder of the tape and individually counted the morphemes in the child's utterances. Next they calculated an MLU and compared it to the previously determined MLU for that sample. MLUs were obtained in this manner for as many of the ten training tapes as were necessary for the observers to meet the exit criterion.

Meeting Exit Criterion: After an observer had successfully calculated two consecutive MLUs that were within ± .2 morphemes of the predetermined MLU she was considered to have met the exit criterion and became a subject used to gather data in the experimental condition. The number of tapes viewed by each subject varied according to the number of trials required to reach exit criterion. An upper limit of ten trials
was chosen because the time required to complete ten trials (approximately 1 and 1/2 to 2 hours) was judged to represent an amount of time which a speech and language clinician might spend learning a new assessment procedure.

**Experimental Condition**

Subjects calculated MLUs for three children, first using the on-line method and then the traditional method. Subjects individually determined MLUs on-line for three children (MLU Stages I, III & V) using procedures similar to those outlined in the training program. The order of tape presentation was counterbalanced across subjects and the experimenter recorded the time necessary for each subject to complete each MLU calculation, including the time spent viewing the tape. The subjects then viewed the same three tapes individually and determined the MLU using the traditional method while the experimenter recorded the total amount of time required for each subject to determine the MLU for each tape.

A control group of three graduate students also determined MLUs for the three experimental tapes using the traditional procedure. This group participated in the review portion of the training program but not in the on-line training portion. A control group was included to determine if the training program influenced the subjects' ability to accurately determine MLUs using the traditional method. The performance of the control group was also used to determine whether the time involved by the subjects in scoring the traditional method improved by first having coded MLU on-line.
**Measurement and Design**

A one by two repeated measures design was used to answer both research questions. One group of eight subjects obtained MLU scores for each of three stimulus tapes using two different methods (on-line and traditional). A control group of three subjects also determined MLUs for the three stimulus tapes using only the traditional method. The specific measures obtained using the subjects in the experimental condition were:

1. On-line MLU scores for the three stimulus tapes.

2. MLU scores obtained using traditional procedures for the same three tapes.

3. Amount of time required to determine each MLU on-line.

4. Amount of time required to determine each MLU using traditional procedures.

The following measures were obtained on the control group's performance:

1. MLU scores obtained using traditional procedures for the three stimulus tapes.

2. Amount of time required to determine MLUs using traditional procedure.
T-tests for repeated measures were used to determine whether there were significant differences in the MLU scores obtained or the time involved obtaining them between the two measurement methods. Independent t-tests were used to compare the control group's scores and times to those of the subjects. Additionally, descriptive data was collected throughout the study to be presented in tabular and graphic forms.
RESULTS

Entrance Criterion

Eighteen graduate students were possible subjects for this study. Thirteen of these successfully passed the entrance criterion of accurately obtaining an MLU score using the traditional transcription method. The MLUs that were obtained by the 18 participants on the entrance tape ranged from 3.7 to 5.64 (Appendix I). Subjects who calculated an MLU that was within .2 morphemes from the predetermined MLU were eligible to begin the on-line training portion of this study. Thirteen students successfully passed this entrance criterion and five who were unable to calculate an MLU within this limit and were excluded from the study. Ten of the thirteen who passed entrance criterion participated in the on-line training and three comprised a control group for the experiment.

On-line Training and Exit Criterion

Eight of the ten possible subjects successfully met the exit criterion of correctly computing two consecutive on-line MLU scores that were within .2 morphemes from the predetermined MLU. The range of trials needed to meet this criterion was three to seven (Table 3). The eight who successfully met exit criterion became the subjects for the experimental condition. The two who were unable to meet criterion in ten trials were excluded from the study.
Table 3. Number of trials required to pass on-line training.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Trials</th>
<th>Pass/No Pass</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>Pass</td>
</tr>
<tr>
<td>2</td>
<td>9</td>
<td>No Pass</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td>No Pass</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>Pass</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>Pass</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>Pass</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>Pass</td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>Pass</td>
</tr>
<tr>
<td>9</td>
<td>7</td>
<td>Pass</td>
</tr>
<tr>
<td>10</td>
<td>3</td>
<td>Pass</td>
</tr>
</tbody>
</table>
Experimental Condition

MLU Differences: T-tests for repeated measures were used to determine whether significant differences existed between on-line and traditional MLU scores obtained by the subjects. Three separate t-tests were performed, one for each of the three children on the stimulus tapes. The data base consisted of 23 pairs of scores. Each of eight subjects generated three pairs of scores, an on-line and a traditional score for each child (Table 4). This should have resulted in 24 pairs, but one subject was unable to calculate an MLU for Child C because she had not obtained data for 50 utterances by the time the video taped language sample was completed. No significant differences were found between the on-line and traditional MLU scores (a = .05); Child A: t = .29, Child B: t = 2.28, Child C: t = .95. This indicated that trained observers were able to obtain MLUs on-line which were statistically similar to those they obtained with traditional procedures.
Table 4. On-line and traditional MLU scores.

(ON = online; TRAD = traditional).

<table>
<thead>
<tr>
<th>Subject</th>
<th>CHILD A</th>
<th>CHILD B</th>
<th>CHILD C</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ON</td>
<td>TRAD</td>
<td>ON</td>
</tr>
<tr>
<td>1</td>
<td>1.30</td>
<td>1.32</td>
<td>3.20</td>
</tr>
<tr>
<td>2</td>
<td>1.36</td>
<td>1.30</td>
<td>2.86</td>
</tr>
<tr>
<td>3</td>
<td>1.40</td>
<td>1.32</td>
<td>2.80</td>
</tr>
<tr>
<td>4</td>
<td>1.42</td>
<td>1.54</td>
<td>3.28</td>
</tr>
<tr>
<td>5</td>
<td>1.38</td>
<td>1.32</td>
<td>3.02</td>
</tr>
<tr>
<td>6</td>
<td>1.40</td>
<td>1.46</td>
<td>2.82</td>
</tr>
<tr>
<td>7</td>
<td>1.32</td>
<td>1.34</td>
<td>2.96</td>
</tr>
<tr>
<td>8</td>
<td>1.38</td>
<td>1.26</td>
<td>3.16</td>
</tr>
</tbody>
</table>

A Pearson Product Moment Correlation was used as another means of comparing on-line scores with traditional scores. When all 23 pairs of scores were used in the analysis the \( r \) value equaled .98; showing a very high correlation between on-line and traditional scores. This finding suggests that MLUs can reliably be used to assign children to different linguistic stages, and that the on-line and traditional
methods are equally effective ways of doing so. When analyzing each child's tape separately the correlations were greatly reduced, as would be expected because of the limited range of scores for each child and the small number of comparisons in each group. Results for Child A (r = .51) and Child B (r = .46) showed no significant correlation between on-line and traditional scores. The on-line scores for Child C correlated moderately with the traditional scores (r = .77). The significant correlation for Child C may be explained by the fact that this child had the longest MLU and the resulting scores had the most variation between them.

The clinical significance between MLU scores was also examined. Specifically, comparisons were made to determine whether on-line MLU scores differed in stage assignments from traditional MLU scores. Because all clinicians do not use the same stage divisions, the data were analyzed in relation to stage assignments twice; once using Brown's (1973) original five stages and then using Miller and Chapman's (1981) eight divisions of these five stages.

Table 5 summarizes the differences in scores and stage assignments based on the MLUs obtained. The on-line score was subtracted from the traditional score and the differences are presented. These differences ranged from .02 to .7 morphemes; with 78% of them being ±.2 morphemes of each other. There were no patterns indicating that either method consistently resulted in a higher or lower MLU score.
Table 5. Stage assignments based on on-line and traditional MLU scores. (ON = on-line; TRAD = traditional)

<table>
<thead>
<tr>
<th>Subject</th>
<th>Child Differences in Scores (TRAD - ON)</th>
<th>Brown's Stages ON</th>
<th>Brown's Stages TRAD</th>
<th>Miller &amp; Chapman's Stages ON</th>
<th>Miller &amp; Chapman's Stages TRAD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A .02</td>
<td>I</td>
<td>I</td>
<td>EI</td>
<td>EI</td>
</tr>
<tr>
<td></td>
<td>B .34</td>
<td>IV *</td>
<td>V</td>
<td>EIV</td>
<td>IV-V</td>
</tr>
<tr>
<td></td>
<td>C --</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>2</td>
<td>A -.06</td>
<td>I</td>
<td>I</td>
<td>EI</td>
<td>EI</td>
</tr>
<tr>
<td></td>
<td>B .06</td>
<td>IV</td>
<td>IV</td>
<td>III</td>
<td>III</td>
</tr>
<tr>
<td></td>
<td>C .2</td>
<td>V+</td>
<td>V+</td>
<td>LV</td>
<td>LV</td>
</tr>
<tr>
<td>3</td>
<td>A -.08</td>
<td>I</td>
<td>I</td>
<td>EI</td>
<td>EI</td>
</tr>
<tr>
<td></td>
<td>B .6</td>
<td>IV</td>
<td>IV</td>
<td>III *</td>
<td>EIV</td>
</tr>
<tr>
<td></td>
<td>C .3</td>
<td>V+</td>
<td>V+</td>
<td>LV *</td>
<td>PV</td>
</tr>
<tr>
<td>4</td>
<td>A .13</td>
<td>I</td>
<td>I</td>
<td>EI</td>
<td>* LI</td>
</tr>
<tr>
<td></td>
<td>B .08</td>
<td>IV</td>
<td>IV</td>
<td>EIV</td>
<td>EIV</td>
</tr>
<tr>
<td></td>
<td>C -.02</td>
<td>V+</td>
<td>V+</td>
<td>PV *</td>
<td>LV</td>
</tr>
<tr>
<td>5</td>
<td>A -.06</td>
<td>I</td>
<td>I</td>
<td>EI</td>
<td>EI</td>
</tr>
<tr>
<td></td>
<td>B .06</td>
<td>IV</td>
<td>IV</td>
<td>EIV</td>
<td>EIV</td>
</tr>
<tr>
<td></td>
<td>C .1</td>
<td>V+</td>
<td>V+</td>
<td>PV</td>
<td>PV</td>
</tr>
<tr>
<td>6</td>
<td>A .06</td>
<td>I</td>
<td>I</td>
<td>EI</td>
<td>EI</td>
</tr>
<tr>
<td></td>
<td>B -.08</td>
<td>IV</td>
<td>IV</td>
<td>III</td>
<td>III</td>
</tr>
<tr>
<td></td>
<td>C -.26</td>
<td>V+</td>
<td>V+</td>
<td>PV *</td>
<td>LV</td>
</tr>
<tr>
<td>7</td>
<td>A .02</td>
<td>I</td>
<td>I</td>
<td>EI</td>
<td>EI</td>
</tr>
<tr>
<td></td>
<td>B .02</td>
<td>IV</td>
<td>IV</td>
<td>III</td>
<td>III</td>
</tr>
<tr>
<td></td>
<td>C -.7</td>
<td>V+</td>
<td>V+</td>
<td>PV *</td>
<td>LV</td>
</tr>
<tr>
<td>8</td>
<td>A -.12</td>
<td>I</td>
<td>I</td>
<td>EI</td>
<td>L1</td>
</tr>
<tr>
<td></td>
<td>B -.04</td>
<td>IV</td>
<td>IV</td>
<td>EIV</td>
<td>EIV</td>
</tr>
<tr>
<td></td>
<td>C .2</td>
<td>V+</td>
<td>V+</td>
<td>PV</td>
<td>PV</td>
</tr>
</tbody>
</table>

Total number of same stage assignments 22 (96%) 16 (70%)

(E = early stage; L = late stage; P = post stage; * = different stage assignment)
Table 5 also shows specific stage assignments for each score. One column uses Brown's assignments and the other uses Miller and Chapman's assignments. Using Brown's more general categories 22 of the 23 pairs of scores resulted in the same stage assignment. Using Miller and Chapman's divisions, which have a smaller range (.5 morphemes), 16 of the 23 pairs of scores resulted in the same stage assignment. The stage assignments for all of the pairs that were not similar were however, all in adjacent stages. Additionally, two pairs showing different stage assignments barely crossed stage boundaries with score differences less than .2 morphemes. In general, trained observers were able to obtain on-line MLU scores that resulted in stage assignments that were identical to, or one stage away from those they obtained using the traditional method.

Time Differences: The amount of time required for each subject to determine both an on-line and a traditional MLU was recorded separately for each child; again resulting in 23 pairs. The amount of time involved in the traditional method exceeded that involved in the on-line method for 22 of the 23 pairs of scores (Appendix J). One subject's traditional time for Child A was 15 seconds less than her on-line time for that child.

Three t-tests for repeated measures were used to determine whether the amounts of time involved for each method were significantly different from one another. Times involved for all of the children were found to be significantly different (α=.05); Child A: t = 2.21, Child B: t = 5.20 and Child C: t = 5.85, with the times for traditional
method being consistently greater than for the on-line method.

Table 6 contains both the ranges of times required for subjects to complete MLU measures using the two methods and the means for each group. Child A was in Early Stage I and required the least amount of time for subjects to determine MLUs, both on-line (mean = 8:31) and traditionally (mean = 11:25). More time was required to calculate MLUs for Child B (Stage III) and Child C (Stage V). The mean on-line times for both children were approximately 11 minutes and the mean traditional times were approximately 18 minutes (Child B) and 19 minutes (Child C). Additionally the differences between on-line and traditional times for Children B and C were greater than for Child A by approximately 4 to 5 minutes. In summary, a significantly shorter period of time was required to gather MLU information on-line than traditionally, and the amount of time saved increased as the children became more linguistically advanced.
Table 6. Time ranges and means for each method in minutes and seconds.

<table>
<thead>
<tr>
<th>Method</th>
<th>Child A</th>
<th>Child B</th>
<th>Child C</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>8:00 - 9:50</td>
<td>9:23 - 13:00</td>
<td>9:19 - 13:42</td>
</tr>
<tr>
<td>On-line</td>
<td>mean = 8:31</td>
<td>mean = 11:14</td>
<td>mean = 11:12</td>
</tr>
<tr>
<td></td>
<td>mean = 11:25</td>
<td>mean = 18:09</td>
<td>mean = 19:18</td>
</tr>
</tbody>
</table>

Differences

between methods  | 1:01 - 12:00 | 6:41 - 12:45 | 4:46 - 13:04 |
Mean differences | 2:54         | 6:55         | 8:06         |

Subjects vs. Control Group: Results of the traditional procedure obtained by the experimental subjects were compared to those results obtained by the control subjects to determine whether participation in the on-line training influenced the subjects ability to calculate MLUs in the traditional manner. No significant differences were found between the traditional MLU scores obtained by the two groups (Child A: t = .409; Child B: t = .718; Child C: t = .476). Neither were significant differences found in the amount of time required to complete the task of traditionally determining MLU scores between the experimental subjects and control subjects (Child A: t = 1.542, Child B: t = .330, Child C t = .430).
These results indicated that calculating an on-line MLU prior to using the traditional method did not significantly increase or decrease the traditional MLU scores. Results also indicated that neither the practice during on-line training, nor first viewing the tape and determining the MLU on-line significantly decreased the amount of time subjects needed to determine MLUs using the traditional method of calculating MLUs.

**Summary**

Subjects who were successfully able to pass the on-line training portion of this study were able to obtain on-line MLU scores that were similar to those obtained with the traditional method. The results were similar statistically as well as clinically, based on the linguistic stage assignments. The amount of time required to obtain MLUs on-line was consistently and significantly less than that required for the traditional method. Based on the comparisons made between experimental subjects and the control group, the data generated by the experimental subjects was judged to be a representative sample of what would be expected by a group of clinicians who were able to pass the entrance criterion of accurately determining an MLU score using the transcription method.
DISCUSSION

The purpose of this study was to compare results of MLUs obtained on-line with MLUs computed in the traditional manner using subjects who had received training on the on-line method. For subjects who passed the on-line training at the criterion level of ± .2 morphemes for two consecutive tapes, there were no significant differences between the MLU scores obtained using the two methods. In addition, less time was required to compute MLUs using the on-line method than was required using the transcription method.

The rest of this chapter has been divided into two major sections. The first deals with the clinical implications of this study and suggests ways in which determining MLUs on-line can help increase the speech and language clinician's efficiency and effectiveness in dealing with their communicatively disordered clients. Next, suggestions for future research are presented as means of overcoming some of the limitations of the present study and to extend and further define the findings of this study.

Clinical Implications

A review of literature found no MLU training nor reliability studies; consequently, the results of this study can not be discussed in comparison to other studies. The usefulness of the present study, however, will be discussed in relation to ways on-line MLU calculations can be used in a clinical setting.
An MLU can be used as a guideline to determine whether or not a child's expressive language skills are developing normally. Miller and Chapman (1981) have outlined age ranges at which children normally pass through linguistic stages (Early I - Post V). By comparing a child's MLU to these age norms a clinician can determine whether or not the child is acquiring language at a rate similar to other children. An MLU obtained on-line during the initial part of an evaluation can be used as a screening device to determine whether a child's expressive language skills are within normal limits, or whether further evaluation is warranted.

Quickly obtaining on-line MLU scores early in an evaluation would help clinicians make decisions on-line which would enable them to more effectively use the evaluation time. First, by using an MLU score as a screening device, the clinician can determine whether further evaluation of expressive language is necessary. If further evaluation is needed, by knowing the child's MLU stage the clinician will be able to evaluate the specific structures that may be beneficial in diagnosing the type of linguistic language disorder.

MLUs are also used to assign children to general linguistic stages. Researchers have been able to document some specific language structures which normally occur during specific stages (Brown, 1973; de Villiers and de Villiers, 1973a; Miller, 1981). By comparing a child's MLU stage assignment with his/her use of these language structures (i.e. grammatical morphemes) a clinician can determine whether or not a child's expressive language is progressing in the normal sequence. The
on-line method of determining MLUs is judged to be clinically similar to the traditional method in that stage assignments based on on-line scores were identical to, or in the stage adjacent to that derived from the traditional score. If certain structures are not used by a child during a language sample the clinician, by knowing the child's MLU stage, can set up situations immediately in the diagnostic setting to determine if a child can produce the structure given the opportunity. The procedures presently available for calculating MLU do not allow clinicians to make these on-line decisions and a second diagnostic session is often required to obtain information regarding the use of specific language structures.

By using a child's MLU score, Miller and Chapman's age norms, and information gained on the usage of specific language structures a clinician can begin making decisions about the diagnosis. First, is the child developing expressive language within normal limits? If not, is the child's language acquisition pattern delayed or deviant? A child with delayed language skills would be using the structures expected to occur in his MLU stage but his MLU would be less than expected for a child his age. A child with a deviant language pattern would not have some or all of the specific structures expected at his MLU stage; his MLU may or may not match the age norms.

Previously, clinicians have not determined a child's MLU until after the evaluation was completed. Not knowing the child's MLU stage during the evaluation required the clinician to sample behaviors across several stages or to take lengthy language samples hoping that all the
structures needed for diagnosis occurred. Much time could be saved by determining MLUs on-line and then focusing the remainder of the evaluation gathering information on the specific behaviors that would be most beneficial in diagnosing the type of disorder.

MLUs have also been used in the therapeutic process to monitor a child's progress or to evaluate the successfulness of a specific type of intervention. A disadvantage in using MLUs obtained traditionally as a monitoring device is the amount of time required to transcribe utterances and calculate the MLU score. Eliminating this 45 minutes to one hour expenditure of time makes using MLUs to monitor progress during therapy more reasonable.

In summary, the determination of MLUs on-line has been demonstrated to be as reliable as the use of the traditional method and the time involved is significantly less. The previous section has discussed some of the advantages of determining MLU on-line. This procedure can be used to increase the efficiency of language evaluations and to measure the effectiveness of therapeutic intervention.

**Future Research**

This study was designed to determine whether clinicians could be trained to reliably gather MLU data on-line under "ideal" conditions. Since clinicians were able to obtain accurate on-line MLUs under "ideal" circumstances, reliability in less than ideal circumstances must be investigated. All children in this study were normal children, in that they had intelligible, fluent speech and a MLU within the appropriate
range for a child his or her age. One question that needs to be answered is "Will clinicians be as reliable when determining MLUs on-line for communicatively disordered children?" To answer this question research must focus on a variety of communicative disorders (delayed language, impaired articulation, dysfluency, etc.) as well as children with varying degrees of severity.

The stimulus tapes were constructed such that pause times of approximately three seconds or more separated a majority of the child's utterance. If there had been less pause time, would clinicians still have been able to obtain reliable on-line MLUs? Or, was this pause time necessary to allow clinicians time to count and record morphemes? Further research is needed to assess the limits under which reliable MLUs can be obtained.

Because this study was conducted under "ideal" conditions, caution must be used when generalizing the results to a clinical setting. No definite statements on the clinical usefulness of on-line MLUs can be made until future research investigates two areas; the reliability of this measure with disordered children and the reliability with language samples which have not met strict selection criteria.

Another topic for future research is in the area of intra-clinician reliability in general, rather than just for clinicians demonstrating a certain level of accuracy during training. In order for subjects to participate in the experiment they had to calculate two consecutive "accurate" MLUs on-line during training. For their scores to be considered accurate they had to closely match those obtained by the
examiner (inter-subject reliability). Subjects who were unable to match the examiner were excluded from the study and did not participate in the experimental condition. In the experimental condition subjects' scores were compared to their traditional scores (intra-subject reliability) rather than to the examiner's standards. The present study only investigated the intra-clinician reliability of clinicians who were shown as a group to perform similarly to a predetermined standard. Not all clinicians in the field are trained to meet such a standard and therefore intra-clinician reliability in general must be investigated. Additionally, clinicians must investigate their own internal consistency before choosing to use an on-line measure of MLU.

Results of the entrance criterion portion of this study suggested that the inter-subject reliability for MLU measures obtained traditionally is another area that needs further investigation. The present study selected a group of subjects who were accurate in determining MLUs using the transcription method. Eighteen subjects transcribed the entrance tape and only thirteen of them were within .2 morphemes of the predetermined MLU. Prior to transcribing the entrance tape each possible subject reviewed the rules for computing MLUs and for segmenting utterance. During this review, which contained practice exercises, discrepancies arose between subjects in counting morphemes of particular grammatical structures. These discrepancies were discussed and agreements were reached prior to the transcription of the entrance tape. Even after this review procedure the MLUs obtained varied as much as 1.94 morphemes between transcribers. Using Miller and Chapman's stages this range of scores spanned three MLU stages. This means that
using the exact same language sample clinicians computed MLUs that resulted in placing the child in three different stages. The discrepancies did not appear to be in the actual counting of morphemes, but rather in the transcription of utterances. For example, one specific utterance was judged to be unintelligible by eight observers and the other ten judged the same utterance to contain two to six morpheme (Appendix K).

These results suggested that, although students in speech and language pathology are exposed to the measure of MLU, little consistency may exist among them in the calculation of MLU scores. Many diagnostic and therapeutic decisions are made based on MLU measures and the wide range in MLU scores obtained for the entrance tape suggest that these decisions may differ from clinician to clinician depending on who determined the MLU. These discrepancies indicate a need for further research investigating the inter-examiner reliability of determining MLUs traditionally and the effectiveness specific training has on the determination of MLU scores. The variability in clinicians' ability to obtain MLUs traditionally must be documented. Then specific training programs designed to increase consistency among clinicians should be implemented and tested as to their effectiveness at increasing inter-examiner reliability. Investigating inter-clinician and further investigating intra-clinician reliability will provided needed insight on the overall usefulness of MLU as a measure of expressive language.
Summary

For clinicians who have demonstrated a high level of inter-examiner reliability, the use of an on-line procedure for determining MLU has been demonstrated to be as accurate as the traditional method. Clinicians are able to obtain accurate MLU scores for normal children. The question still remains as to whether on-line MLU calculations can be done accurately for speech and/or language disordered children. If similar research results are obtained with children who are communicatively impaired, the use of language sampling to determine MLUs will become a more efficient means for evaluating language skills and monitoring therapeutic progress. The time involved in diagnostic evaluations can be greatly reduced and more effectively used. The success of intervention strategies can be more easily monitored making intervention more effective also. Overall, clients will be better served if clinicians are able to improve both their efficiency and effectiveness.
References


Davis, E. A. The development of linguistic skills in twins, singletons with siblings and only children from five to ten years. *Institute of Child Welfare Monographs Series*, Minneapolis: University of Minnesota Press, 1937.


APPENDIX A

BROWN'S LINGUISTIC STAGES
(MLU 1.01-4.0)

I. Semantic Roles and Syntactic Relations

II. Grammatical Morphemes and Modulation of Meaning

III. Modalities of the Simple Sentence

IV. Embedding of one sentence with Another

V. Coordination of Simple Sentence and Propositional Relations
APPENDIX B

LANGUAGE SAMPLE ELICITATION GUIDELINES
(taken from Lund and Duchan, 1983, 19-20)

1. Keep focus off your attempt to get the child to talk. You should comment on what you are doing and allow for, but not directly request, the child's verbal participation.

2. Do not talk too much and do not be afraid to allow silent pauses during the conversations. Do not fill up every empty space with a question. This encourages the child to let you take the lead.

3. Select materials appropriate to the child's level of functioning.

4. If the child will initiate conversation about your materials, let him or her take the lead and ask questions or comment briefly on what the child is saying.

5. If the child does not initiate, make comments yourself about the materials and ask open-ended leading questions.

6. If statements or questions produce no response, demonstrate what you expect of the child.
## APPENDIX C

### SAMPLE TAPE AND CHILD INFORMATION

<table>
<thead>
<tr>
<th>Child</th>
<th># of Utterances</th>
<th>Sample Length</th>
<th>Age (yr-mo)</th>
<th>MLU</th>
<th>MLU Stage</th>
<th>PCC</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Training Tapes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>76</td>
<td>11 min 58 sec</td>
<td>3-2</td>
<td>4.25</td>
<td>Late V</td>
<td>94%</td>
</tr>
<tr>
<td>2</td>
<td>72</td>
<td>14 min 14 sec</td>
<td>2-11</td>
<td>2.12</td>
<td>II</td>
<td>81%</td>
</tr>
<tr>
<td>3</td>
<td>65</td>
<td>11 min 2 sec</td>
<td>3-11</td>
<td>4.46</td>
<td>Late V</td>
<td>99%</td>
</tr>
<tr>
<td>4</td>
<td>72</td>
<td>12 min 41 sec</td>
<td>3-11</td>
<td>4.94</td>
<td>Post V</td>
<td>94%</td>
</tr>
<tr>
<td>5</td>
<td>75</td>
<td>10 min 4 sec</td>
<td>3-2</td>
<td>3.86</td>
<td>IV-V</td>
<td>83%</td>
</tr>
<tr>
<td>6</td>
<td>74</td>
<td>14 min 44 sec</td>
<td>2-2</td>
<td>2.56</td>
<td>III</td>
<td>85%</td>
</tr>
<tr>
<td>7</td>
<td>62</td>
<td>17 min 0 sec</td>
<td>3-5</td>
<td>3.68</td>
<td>IV-V</td>
<td>92%</td>
</tr>
<tr>
<td>8</td>
<td>67</td>
<td>13 min 23 sec</td>
<td>2-8</td>
<td>3.82</td>
<td>IV-V</td>
<td>91%</td>
</tr>
<tr>
<td>9</td>
<td>75</td>
<td>14 min 30 sec</td>
<td>4-0</td>
<td>3.96</td>
<td>IV-V</td>
<td>95%</td>
</tr>
<tr>
<td>10</td>
<td>91</td>
<td>9 min 5 sec</td>
<td>3-8</td>
<td>3.80</td>
<td>IV-V</td>
<td>89%</td>
</tr>
<tr>
<td><strong>Experimental Tapes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>72</td>
<td>9 min 26 sec</td>
<td>1-9</td>
<td>1.32</td>
<td>I</td>
<td>69%</td>
</tr>
<tr>
<td>B</td>
<td>79</td>
<td>13 min 26 sec</td>
<td>2-11</td>
<td>2.88</td>
<td>III</td>
<td>94%</td>
</tr>
<tr>
<td>C</td>
<td>73</td>
<td>11 min 24 sec</td>
<td>4-3</td>
<td>4.24</td>
<td>V</td>
<td>93%</td>
</tr>
</tbody>
</table>
APPENDIX D

PROCEDURES TO CALCULATE PERCENTAGE CONSONANTS CORRECT (PCC)

Sampling rules

1. Consider only intended (target) consonants in words. Intended vowels are not considered.
   a. Addition of a consonant before a vowel is not scored because the target sound is a vowel.
   b. Post-vocalic /r/ is a consonant, but stressed and unstressed vocalics are vowels.

2. Do not score target consonants in the second or successive repetitions of a syllable, score only the first occurrence.

3. Do not score target consonants in words that are completely or partially unintelligible or whose gloss is highly questionable.

4. Do not score target consonants in the third or successive repetitions of adjacent words unless articulation changes. For example, the consonants in only the first two words of the series /k t/, /k t/, /k t/ are counted. However, the consonants in all three words are counted if the series were /k t/, /k k/, /k t/.

Scoring rules

1. The following six types of consonant sound changes are scored as incorrect:
   a. deletions of a target consonant;
   b. substitutions of another sound for a target consonant, including replacement by a glottal stop or a cognate;
   c. partial voicing of initial target consonants;
   d. distortions of a target sound, no matter how subtle;
   e. addition of a sound to a correct or incorrect target consonant;
   f. initial /h/ deletion and final /n/ for / / substitutions are counted as errors only when they occur in stressed syllables; in unstressed syllables they are counted as correct.

2. Observe the following:
   a. The response definition for children who obviously have speech errors is "score as incorrect unless heard as correct". This response definition assigns questionable speech behaviors to an "incorrect" category.
   b. Dialectal variants should be glossed as intended in the
child's dialect.

c. Fast or casual speech sound changes should be glossed as the child intended.

d. Allophones should be scored as correct.

Calculation of PCC

\[
PCC = \frac{\text{Number of Correct Consonants}}{\text{Number of Correct Plus Incorrect Consonants}} \times 100
\]
APPENDIX E

COMPUTING MLU


1. Stuttering is marked as repeated efforts at a single word; the word is counted once in the most complete form produced. In the few cases where a word is produced for emphasis, or the like (no, no, no!) each occurrence is counted separately.

2. Such fillers as mmm or oh are not counted, but no, yeah, and hi are.

3. All compound words (two or more free morphemes), proper names and ritualized reduplications count as single words. Some examples are: birthday, rackety-boom, choo-choo, quack-quack, night-night, pocketbook, see saw. The justification for this decision is that there is no evidence that the constituent morphemes function as such for these children.

4. All irregular pasts of the verb (got, did, went, saw) count as one morpheme. Again, there is no evidence that the child relates these to present form.

5. All diminutives (doggie, mommie) count as one morpheme because these children do not seem to use the suffix productively. Diminutives are the standard forms used by the child.

6. All auxiliaries (is, have, will, can, must, would) count as separate morphemes as do all catenatives (gonna, wanna, hafta, gotta). The catenatives are counted as single morphemes, rather than as going to or want to, because evidence is that they function as such for children. All inflections, for example, possessive (s), plural (s), third person singular (s), regular past(ed), and progressive (ing), count as separate morphemes.

Additional Rules (not discussed by Chapman)

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
1. False starts or incomplete sentences followed by a complete sentence are not counted (the boy --- the girl wants one). (Tyack & Gottsleben, 1974; Miller, 1981).

2. Negative contractions (can't, don't, doesn't, etc.) count as two morphemes (Brown, 1973; Miller, 1981).

3. Words such as lookit, which are made of two words that are run together and are used as one word are counted as one morpheme (McCarthy, 1930; Miller, 1981).

4. If an utterance contains one or two syllables that are unintelligible to the observer and judged to represent a single word it is counted as one morpheme (Tyack and Gottsleben, 1974; Miller, 1981).
APPENDIX F

SEGMENTING UTTERANCES

(Johnson, Darley and Spriestersback, 1963)

1. An utterance is considered a separate unit if it is marked off from the proceeding and succeeding remarks by pauses.

2. An utterance is considered finished if the speaker comes to a full stop, either letting the voice fall, giving interrogatory or exclamatory inflection, or indicating clearly that he does not intend to complete the sentence.

3. When one simple sentence is followed immediately by another simple sentence with no pause for breath, the two are considered to comprise one utterance if the second statement was clearly subordinate to the first.

4. Remarks connected by interjections and conjunctions, such as "and", "um", "er", etc. are considered as separate utterances if the remarks appear to be clearly enumerative.
APPENDIX G

ORAL EXAMPLES AND NUMBER OF MORPHEMES

1. H-h-hi hi truck. (2)
2. Help, help, help me! (4)
3. Oh, I see. (2)
4. Hi there. (2)
5. Cookie Monster eat it. (3)
6. Bye-bye airplane. (2)
7. Doggie wanted it. (4)
8. I had mommie’s coat. (5)
9. I hafta get books. (5)
10. He comes toady. (4)
11. I put the . . . you put it away. (4)
12. It don’t work. (4)
13. The bluk is gone. (4)
14. Hey, lookit. (2)

WRITTEN EXAMPLES AND NUMBER OF MORPHEMES

1. Mmm, you like that? (3)
2. When are we gonna do that? (6)
3. I can’t. (3)
4. No, my too little. (4)
5. Get outta here. (3)
6. He flied in airplane. (5)
7. She's trying to get it out. (8)
8. I, uh that, he is here. (3)
9. I said no, no! (4)
10. But you, y-you have it. (4)
11. I go night-night with Raggedy Ann. (5)
12. Daddy went home. (3)
13. Sandy's car has two doors. (7)
14. He goes to smoot. (5)
15. It a no-no. (3)
APPENDIX H

ON-LINE PRACTICE SENTENCES

Set 1
1. I get on it.
2. Oh, she on.
3. I wanted that one.
4. This is Bryan's.
5. I went in.
6. The dog is running here.
7. Hi.
8. Who that girl?
9. I gotta leave it up there now.

Set 2
1. Hi truck.
2. I am playing with friend now.
3. Where it go?
4. Stop!
5. I ate birthday cake all up.
7. She wants one.
8. I gonna give you one.
9. Mmm, I know.
10. Where you going?
APPENDIX I

MLU SCORES OBTAINED FOR ENTRANCE CRITERION

<table>
<thead>
<tr>
<th>Possible Subject Number</th>
<th>Subject Number</th>
<th>MLU</th>
<th>Error</th>
<th>Pass/No Pass</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>4.00</td>
<td>+.20</td>
<td>Pass</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>3.96</td>
<td>+.16</td>
<td>Pass</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>3.72</td>
<td>-.08</td>
<td>Pass</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>3.84</td>
<td>+.04</td>
<td>Pass</td>
</tr>
<tr>
<td>5</td>
<td>-</td>
<td>4.86</td>
<td>+1.06</td>
<td>No Pass</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>3.98</td>
<td>+.08</td>
<td>Pass</td>
</tr>
<tr>
<td>7</td>
<td>6</td>
<td>3.74</td>
<td>-.06</td>
<td>Pass</td>
</tr>
<tr>
<td>8</td>
<td>-</td>
<td>5.64</td>
<td>+1.84</td>
<td>No Pass</td>
</tr>
<tr>
<td>9</td>
<td>-</td>
<td>4.25</td>
<td>+.45</td>
<td>No Pass</td>
</tr>
<tr>
<td>10</td>
<td>-</td>
<td>5.50</td>
<td>+1.70</td>
<td>No Pass</td>
</tr>
<tr>
<td>11</td>
<td>7</td>
<td>3.90</td>
<td>+.10</td>
<td>Pass</td>
</tr>
<tr>
<td>12</td>
<td>-</td>
<td>4.92</td>
<td>+1.12</td>
<td>No Pass</td>
</tr>
<tr>
<td>13</td>
<td>8</td>
<td>3.94</td>
<td>+.14</td>
<td>Pass</td>
</tr>
<tr>
<td>14</td>
<td>9</td>
<td>3.92</td>
<td>+.12</td>
<td>Pass</td>
</tr>
<tr>
<td>15</td>
<td>C1</td>
<td>3.70</td>
<td>-.10</td>
<td>Pass</td>
</tr>
<tr>
<td>16</td>
<td>C2</td>
<td>3.96</td>
<td>+.16</td>
<td>Pass</td>
</tr>
<tr>
<td>17</td>
<td>10</td>
<td>4.0</td>
<td>+.2</td>
<td>Pass</td>
</tr>
<tr>
<td>18</td>
<td>C3</td>
<td>4.0</td>
<td>+.2</td>
<td>Pass</td>
</tr>
</tbody>
</table>

C = control group
### APPENDIX J

**TIME INVOLVED IN ON-LINE AND TRADITIONAL MLU CALCULATIONS**  
(minutes & seconds)

<table>
<thead>
<tr>
<th>Subject</th>
<th>Child</th>
<th>On-line</th>
<th>Traditional</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>8:44</td>
<td>11:31</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>13:43</td>
<td>22:40</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>---</td>
<td>22:21</td>
</tr>
<tr>
<td>2</td>
<td>A</td>
<td>8:35</td>
<td>9:36</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>10:03</td>
<td>20:22</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>9:23</td>
<td>16:39</td>
</tr>
<tr>
<td>3</td>
<td>A</td>
<td>8:53</td>
<td>10:45</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>10:55</td>
<td>18:00</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>13:10</td>
<td>21:50</td>
</tr>
<tr>
<td>4</td>
<td>A</td>
<td>8:00</td>
<td>12:00</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>10:26</td>
<td>23:30</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>10:17</td>
<td>18:30</td>
</tr>
<tr>
<td>5</td>
<td>A</td>
<td>9:50</td>
<td>11:04</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>11:41</td>
<td>16:40</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>10:05</td>
<td>22:48</td>
</tr>
<tr>
<td>6</td>
<td>A</td>
<td>8:00</td>
<td>10:00</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>9:19</td>
<td>14:05</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>10:44</td>
<td>17:25</td>
</tr>
<tr>
<td>7</td>
<td>A</td>
<td>7:38</td>
<td>10:12</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>12:26</td>
<td>13:30</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>14:07</td>
<td>15:30</td>
</tr>
<tr>
<td>8</td>
<td>A</td>
<td>8:33</td>
<td>8:18</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>11:17</td>
<td>16:24</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>10:44</td>
<td>22:30</td>
</tr>
</tbody>
</table>
## APPENDIX K

### TRANSCRIPTION FOR A SINGLE UTTERANCE FROM THE ENTRANCE TAPE

<table>
<thead>
<tr>
<th>Transcription</th>
<th>Number of Morphemes</th>
<th>Number of Transcribers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unintelligible</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>Yeah, [unintelligible].</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Yeah, that was.</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Yeah, that [unintellibible].</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Yeah, that was it.</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Yeah, that what it is.</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>Yeah, that was the thing.</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>Yeah, that what it says.</td>
<td>6</td>
<td>2</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.