1978

Comparison of the Elicited Language Inventory with the Developmental Sentence Scoring method

Terrylea Madsen Reagor
The University of Montana

A COMPARISON OF THE ELICITED LANGUAGE INVENTORY WITH THE DEVELOPMENTAL SENTENCE SCORING METHOD

By

Terrylea Reagor

B.A., Adelphi University, 1970

Presented in partial fulfillment of the requirements for the degree of

Master of Arts

UNIVERSITY OF MONTANA

1978

Approved by:

Chairman, Board of Examiners

Dean, Graduate School

6-2-78

Date

Reagor, Terrylea, M.A., June 1978 Communication Sciences and Disorders

A Comparison of the Elicited Language Inventory with the Developmental Sentence Scoring Method (48 pp.)

Director: Lynda Miller

The purpose of this study was to determine if the Elicited Language Inventory and the Developmental Sentence Scoring procedure are assessing similar features of syntax. Both the ELI and the DSS were administered to thirty-three normal boys ranging in age from 4 years 3 months to 6 years 11 months. A Pearson product moment correlation revealed that a nonsignificant correlation exists between the ELI and the DSS with this population. This study indicates the importance of using both types of procedures in assessing children's abilities to use syntax. Neither procedure alone represents the child's entire ability to expressively use syntax.

ACKNOWLEDGEMENTS

The author would like to express her grateful appreciation to Lynda Miller for her assistance in directing this study and to Mike Wynne for his help in establishing the reliability of the scoring.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

INTRODUCTION

BACKGROUND OF THE PROBLEM

Longhurst and Schrandt (1973, p. 240) assert that "oral language is frequently the most important single factor used to evaluate a child's growth and development." Consequently, it is necessary that the speech clinician obtain a good representation of the child's language performance. Many authorities agree that after an assessment the clinician needs to be able to describe any communication problems, provide suggestions for remediation, and estimate prognosis. For example, Johnson, Darley, and Spriesterbach (1963) state that the speech clinician seeks to describe adequately any communication problem through an appropriate examination. After interpreting the findings, suggestions for therapy and a prognosis are made. During the examination standardized tests are useful in analyzing communication skills. Weiner and Hoock (1973, p. 616) state:

> The use of standardized tests in clinical examinations or in research has a number of advantages over that of more impressionistic methods. Perhaps the value of such tests can be best summarized by the term objective. The carefully explicit directions for administration allow for replication of the examination with the same or different individuals. Unwanted and uncontrolled variation can be largely eliminated. Informal measures, usually constructed by the individual examiner, may have the advantage of flexibility. All too often, however, this advantage is outweighed by the subjectivity involved in variable methods

> of presentation of tasks. Another major advantage of standardized instruments rests in their norms. When they are appropriate, such norms provide a basis for meaningful and objective interpretation of the test result.

According to Lee (1974), normative data obtained from standardized measures are necessary for diagnostic purposes, for evaluating the effectiveness of therapy, and for estimating a child's progress in training.

The area of communication assessment that is the focus of this study is the production of syntax by children. Two major procedures utilized in the assessment of syntax will be discussed in this section:

1. Testing of syntax based on a spontaneous sampling procedure.
2. Testing of syntax based on an elicited imitation procedure.

Testing of Syntax Based on a Spontaneous Sampling Procedure

Spontaneous sampling, a procedure preferred by many researchers and clinicians, is useful because it enables the clinician to obtain a more natural example of the child's language (Dale, 1972). It also allows the examiner to evaluate a child's consistency and frequency of usage of linguistic forms (Lee and Canter, 1971). For example, Lee and Canter (1971, p. 316) state, "Conversational speech places a grammatical 'load' upon a child's performance which cannot be evaluated by selective testing." Wilson (1969) asserts that an oral language sample is useful for the clinical evaluation of language development and for research on language development and on language proficiency measures.

However, several problems can occur with spontaneous sampling procedures. One problem concerns numerous variables which can influence the quality of the language sample. For example, stimulus and examiner variables may account for differences in mean length of response found between studies (Cowan, Weber, Hoddinott, and Klein, 1967). The environmental situation and the topic have been found to affect the length of the child's sentences and his total response as well as the completeness of his sentence structure (Hahn, 1948). Wilson (1969) determined that it is beneficial to control stimulus and elicitation variables when obtaining a language sample. Shriner and Sherman (1967) assert that various factors including time of day, emotional and physical condition of the child or of the examiners, practice effect, and stimuli may affect the nature of samples taken from the same children at different times. Two additional problems which may occur with spontaneous sampling are (1) all the behaviors of interest may not appear in a given sample and (2) the sample may not be representative of the child's usual performance (Siegel and Broen, 1976).

A wide variety of measures and indexes have been utilized in the analysis of a language sample. The earliest measures dealt mostly with vocabulary and length of utterance, but little was done to determine the relative importance of the various aspects of linguistic development or to provide a single index of language development (McCarthy, 1954). In one of the early studies, McCarthy (1930) researched the changes in sentence length, the

complexity of sentence structure, and the proportion of the various parts of speech that occurred in language samples of 140 children 18 to 54 months of age. Sentences were measured in stages of approximation to adult usage and were classified according to (1) complete responses which included six stages ranging from "functionally complete but structurally incomplete" to "elaborated" and (2) incomplete responses which included omissions of verbs, subjects, prepositions, conjunctions, verbs and subjects, or miscellaneous. The procedure for obtaining a language sample, which was first devised by McCarthy (1930), was further developed and used by numerous authors including Templin (1957), Winitz (1959), and Wilson (1969). Guidelines for taking the language sample developed by some of these authors may be found in the books by Johnson, Darley, and Spriestersbach (1963) and Lee (1974).

Another important study was conducted by Templin (1957) who studied language samples from 480 children ages 3 to 8 years. She analyzed the utterances according to mean length of response (MLR), grammatical complexity, grammatical accuracy, and the parts of speech used. Length refers to the mean number of words per remark. A descriptive and a quantitative method were used to classify the grammatical complexity. The descriptive category involved dividing all remarks into complete and incomplete sentences. Complete sentences were further classified into the type of sentence and type of subordination. Incomplete sentences were classified according to the type of incompleteness (eg. subject omitted).

Percentages of occurrence of specific sentence types were calculated for the quantitative method. The structural complexity score (SCS) was obtained by assigning scores of 0 for all incomplete remarks, 1 for simple sentences, 2 for simple sentences with two or more phrases or a compound subject or predicate, 3 for compound sentences, and 4 for complex and elaborated compound sentences. The grammatical inaccuracies were classified according to those resulting in incomplete sentences and according to specific grammatical errors. Each word in the remarks of the subjects was categorized according to the following parts of speech: noun, verb, adjective, adverb, pronoun, conjunction, preposition, article, interjection, and miscellaneous. Norms are provided for each of the above four categories in Templin's book (1957) and the book by Johnson et. al. (1963). Templin found the quantitative complexity score to be less stable than the other language measures used in her study. Her complexity score has been questioned for a lack of efficiency (Shriner and Sherman, 1967), reliability (Darley and Moll, 1960; Minifie, Darley, and Sherman, 1963), and validity (Minifie et. al., 1963). The MLR has been called the "most objective and reliable single index" of language development (McCarthy, 1954, p. 574), but it has been criticized on grounds of validity (Cowan et. al., 1967; Minifie et. al., 1963, Shriner, 1969) and reliability (Minifie et. al., 1963; Shriner, 1967, 1969). Miner (1969, p. 226) contends that both the MLR and the SCS provide "relatively scant information about morphological and syntactical developmental changes which occur with

increased chronological age."

Shriner (1967) combined two previous language measures, the MLR and the SCS, to form the Length-Complexity Index (LCI). Sentence length and complexity are scored concurrently based on a numeric weighting system to assess children's developmental changes in grammatical rules (Miner, 1969). Miner (1969) provides a detailed description of this procedure. The LCI score resulted from adding noun phrase points and verb phrase points plus additional points (eg. for negatives, modifiers) for each sentence divided by the number of sentences. In order to utilize this system a familiarity with base structure rules for generating sentences (Chomsky, 1965) is necessary. Besides the numeric analysis, which makes comparisons within and across groups possible, a linguistic analysis can be made. The linguistic analysis, useful in remediation, can reveal the kind and frequency of generative rules utilized by the child. Shriner (1967) described the LCI as being the best single indicator of language development for children younger than five years. Sharf (1972) also found it to be a valid indicator of language development. However, no normative data were established for the LCI enabling comparisons to be made between a child and his peers.

Linguistic analysis is also used in a language measure developed by Tyack and Gottsleben (1974) to discover the grammatical rule systems used by children. Their procedure provides information for collecting a language sample, analyzing the sample, establishing a training program based on the analysis, and measuring change from the training program. Analyzing the sample involves the following

steps: counting words and morphemes, calculating the word-morpheme mean, assigning the sample to a linguistic level, sorting the forms and constructions, and entering the forms and constructions into appropriate categories. The forms and constructions are sorted into categories such as pronouns, prepositions, modals, verb phrase + noun phrase, and questions. They are also classified according to categories based on the child's assigned level. For example, Category B lists those appearing correctly above the child's assigned level. A sequence of language acquisition is provided which lists the linguistic developmental level based on the word-morpheme index for specific forms and constructions. For example, the verb + modifier + noun construction occurs at Level II which has a word-morpheme index of 2.5 to 3.0. Instructions for counting words, morphemes, and sentences and for categorizing constructions are also provided. This procedure provides a detailed analysis of a language sample with specific guidelines for utilizing the resulting information in therapy; however, the measure has not been standardized on a large population and no reliability or validity data are available at present.

Besides Shriner (1967) and Tyack and Gottsleben (1974), Lee (1974) was also influenced by psycholinguistic research in her development of two procedures, Developmental Sentence Types (DST) and Developmental Sentence Scoring (DSS). Both procedures are constructed according to developmental stages of language acquisition and can be used by the clinician as part of the assessment of children with language problems. They are useful in analyzing

spontaneous language samples, they provide progressive goals for planning remediation, and they allow for estimations of progress to be made during therapy. The DST is a method for studying and evaluating a child's grammatical development while he is still speaking mostly in pre-sentences. The pre-sentences are classified according to length and variety of types. The length classification contains three horizontal levels: (1) single words, (2) two-word combinations, and (3) multiword constructions which are not complete sentences. The type classification shows the development of the following vertical levels: (1) the noun phrase, (2) the designative sentence which names the topic of conversation, (3) the predicative sentence which names the topic and comments on it, (4) the subject-verb sentence, and (5) fragments which are appendages to the three main sentence types. No single score is given to the DST classification to measure verbal maturity. However, normative data are provided on 40 children ages 2-0 to 2-11 for comparison purposes.

The results of Sharf's study (1972) indicate a significantly close relationship between DST and MLR and LCI in determining growth of expressive language. Although Longhurst and Schrandt (1973) did find the DST to accurately discriminate between the language of a normal and a disordered speaker, they felt that it gave no more information than just classifying children's utterances according to levels of development such as word, phrase, construction, and sentence in describing the difference. Rather than being transformational, they stated that it just followed the developmental

sequence from sound to sentence without regard to the function of the utterance.

Lee's (1974) second procedure, the DSS, assesses a higher developmental level of a child's grammatical development than the DST. Only complete sentences which contain a noun and verb in subject-predicate relationship are included in a DSS analysis. The following eight categories of grammatical forms are analyzed: (1) indefinite pronoun or noun modifier, (2) personal pronoun, (3) main verb, (4) secondary verb, (5) negative, (6) conjunction, (7) interrogative reversal in questions, and (8) wh-question. Weighted scores are assigned to the forms within each category so that grammatical growth can be compared both within and across categories. Higher scores are given to later developing forms. A structure is not given a score unless all the adult syntactic and morphological rules are followed. If the sentence meets all of the adult grammatical standards, one additional sentence point is given. Norms for 160 children, ages 3-0 to 6-11, were provided with the original presentation of the DSS (Lee and Canter, 1971). The speech samples for these children were later rescored utilizing the Reweighted DSS procedure (Lee, 1974). The new norms for the 3-0 to 6-11 age group plus norms for 40 children, ages 2-0 to 2-11, are provided in Lee's book (1974). Norms are given for the over-all mean DSS score but not for each of the grammatical categories. Four length-maturity measures were also computed for number of words, number of DSS entries, weighted-sum scores, and mean developmental scores. These four measures are based on two inferences concerning

the DSS mentioned by Lee (1974): (1) the DSS measures the developmental maturity of syntactical usage as well as the length of utterance and (2) the DSS provides "quantification of the grammatical load of all structures combined within separate utterances (p. 238)."

Koenigsknecht (1974) reports on a number of studies which support the usefulness of the DSS and its reweighted scoring system as an objective measure of the development of syntax in children. For example, its validity was indicated by a significant increase in over-all scores and in grammatical category scores with an increase in age level. Significantly high reliability coefficients were also obtained. These include an over-all internal consistency of 0.71, a split-half reliability of 0.73, and a positive correlation between each of the grammatical categories and the over-all DSS score.

Results of studies of interviewing-clinician differences, stimulus material differences, temporal reliability, and sentence sequence effects also give evidence to the stability of the DSS procedure (Koenigsknecht, 1974). No significant differences were found between DSS scores obtained by two clinicians across three age levels. There were no significant differences obtained on the over-all DSS scores when using different stimulus materials. But the following four categories did show significant score changes with different materials: indefinite pronouns, personal pronouns, secondary verbs, and interrogative reversals. Significant increases in the total DSS score and in five categories (main verb, conjunction,

indefinite pronoun, personal pronoun, and negative) were found with four repeated applications during a two week period, indicating a possible practice effect (for example, over familiarity with the stimulus materials). However, increases over longer time intervals of four and eight months were consistent with increases in the developmental patterns in the normative data for the over-all score and for the categories. Performances on the first half of a language sample were found to be consistent with performances on the second half for the total DSS score and for each of the categories. This consistency between halves indicated that the effects of warm-up and general adjustment to the conversational setting did not result in significantly higher syntactic forms on later utterances in the sample. Although the above studies lend credence to the high reliability and high validity of the DSS, Johnson and Tomblin (1975) found that the sample size may affect the reliability of the DSS. They stated that a sample size of approximately 175 sentences is needed before a limited reduction in error can be achieved.

In summary, spontaneous sampling allows the clinician to obtain a more natural sample of language and to evaluate consistency and frequency of usage of linguistic forms. But this procedure has three major problems: (1) many variables can influence the quality of the language sample; (2) all the syntactical features which the clinician wants to assess may not appear in one sample; and (3) one sample may not be typical of the child's usual performance. In spite of these problems, numerous measures, which were described above, use spontaneous sampling. One of the more prevalent methods,

the DSS, possesses both good and bad attributes. All of the validity measures and several of the reliability measures have been significantly high. But some reliability measures, such as temporal reliability for short periods, were low. The small sample size of the DSS may also lower the reliability. Additional problems with the DSS are discussed under "PURPOSE OF THE STUDY."

Testing of Syntax Based on an Elicited Imitation Procedure

Several authors including McNeill (1970), Menyuk (1964), and Berry-Luterman and Bar (1971) have proposed a second method for providing useful information regarding language in children. This method utilizes sentence imitation rather than spontaneous sampling for testing the production of syntax. The focus is on verbatim repetition, and since the examiner can choose the sentences to be imitated, this method can assess a large range of sentence types (McNeill, 1970). Although Clark (1977) argues that some imitations used by children may be superior to spontaneous productions in syntactic structure, Erwin (1964) states that there is no difference in the grammar utilized by a child in his spontaneous speech and in his imitations of adult utterances. The child will imitate only what he utilizes in his spontaneous speech. According to McNeill (1970), if the model sentence exceeds the child's short term memory, the child will alter it to fit his grammar. In contrast, if the model sentence is within the child's immediate memory, Dale (1972) states that the child can imitate it perfectly without needing to process the meaning of the sentence. He provides this explanation for

why the children's imitation performance was superior to production in the Fraser, Bellugi, and Brown study (1963). Fraser, Bellugi, and Brown (1963) agree with Dale's interpretation in that they conclude that the children did not utilize their meaning systems to imitate sentences in their study. McNeill (1970) also asserts that if a child does not have a grammatical form in his speech, he will not imitate it in the speech of adults. He says that it is this inability to imitate what he does not utilize that allows imitation to be used as a test of children's language production.

Elicited imitation has been used in a number of studies of language development. For example, Menyuk (1964) used it as a procedure for comparing the grammar of 10 deviant and 10 normal speakers in an attempt to describe the term, "infantile speech." It was found that the grammar of the deviant speakers was not similar to that of the normals at any age. In another study Berry-Luterman and Bar (1971) suggested sentence repetition as a diagnostic test of grammatical performance. They assessed the performance of four subjects under the following conditions: (1) the repetition of grammatically incorrect sentences which were taken from the child's own production; (2) the repetition of the grammatically correct versions of these sentences; and (3) the repetition of the reversed word order of the grammatically correct versions. The authors discussed clinical and theoretical aspects of the linguistic performance of these subjects such as trends in the types of errors found.

Only two clinical tests were found which use elicited imitation as a procedure for assessing grammar in children. These are

Lee's (1969) Northwestern Syntax Screening Test (NSST) and Carrow's (1974a) Elicited Language Inventory (ELI).

According to Lee (1970), the NSST is only to be used for screening purposes rather than as a measure of a child's general language development or syntax capabilities. It was developed to identify children 3 to 8 years of age who are delayed enough in syntactic development to need further evaluation.

The second test, Carrow's (1974b) ELI, is the only diagnostic procedure which measures the child's productive control of grammar through elicited imitation. The stimuli are one two-word phrase and 51 sentences (two to ten words in length) which include the folowing grammatical categories: articles (41), adjectives (9), nouns (59), noun plurals (8), pronouns (41), demonstratives (2), conjunctions (7), verbs (103), negatives (13), contractions (12), prepositions (14), and adverbs (12). Errors are recorded according to category (verbs, negatives, etc.) and type (substitution, omission, addition, transposition, and reversal). Norms are provided for total error score, for grammar error score, and for type error score for 475 children, ages 3-0 to 7-11.

Carrow (1974a), along with research assistants and students, conducted a number of studies on the ELI. A significantly high internal and external reliability was found. In one of Carrow's studies the total error score had significant positive correlations with each error type and grammar subscore. Test-retest reliability was 0.98 and interexaminer reliability for transcription and scoring was 0.98. According to Carrow, the finding of significant differences

in total error scores between age groups is indicative of high validity. Validity was also shown in a study (Cornelius, 1974) where a significant difference was found on the ELI in total error score between a group of language disordered children and children with normal language. The ELI was found to have a significantly high correlation of -0.79 with the DSS (Cornelius, 1974).

In summary, advocates of elicited imitation procedures assert that a child will only imitate what he uses in his spontaneous speech. Examples of studies of language development using this procedure were discussed above. Although two clinical tests which use elicited imitation were presented, only one of these tests, the ELI, is a diagnostic procedure for assessing language. The results of several studies conducted with the ELI revealed significantly high validity and reliability values.

Comparison of Spontaneous Sampling Procedures with Elicited Imitation Procedures

The two methods of spontaneous sampling and elicited imitation were directly compared in four studies (Menyuk, 1969; Prutting, Gallagher, and Mulac, 1975; Cornelius, 1974). Menyuk (1969) used elicited imitation in two studies to determine how various grammatical utterances were understood and reproduced by children at different ages. She made a comparison between the types of structures imitated and those spontaneously produced by these children. The results of these studies indicated that it was the structure of a sentence which determined whether or not it was repeated, not its

length. Significant correlations were found between structures imitated and those spontaneously produced for all groups in both studies. In a third study Prutting et. al. (1975) compared syntactic structures produced on the expressive portion of the NSST with those produced in a spontaneous language sample. They found that 30% of the grammatical structures incorrectly produced on the NSST were correctly produced in a spontaneous language sample. It was concluded that the expressive portion of the NSST does not accurately represent the child's spontaneous language performance and, therefore, cannot be used for purposes other than as a screening instrument.

In the fourth study Cornelius (1974) utilized the ELI to examine the following criteria: (1) validity as demonstrated by the ability to separate language disordered and normal children, (2) interexaminer reliability, (3) efficiency based on time required for administration, transcription and scoring, and (4) ability to provide specific information concerning the child's language. The results were compared to the results obtained utilizing the DSS. It was found that the ELI does meet all of the above important criteria. A significantly high correlation of -0.79 ($p < .005$) was found between the total scores on the ELI and the DDS, indicating that both procedures provide similar information. However, low interexaminer reliability, excessive time requirements, and limited control of item sampling found with the DSS procedure were not evident with the ELI.

Two types of procedures utilized to assess the production of

syntax in children were described above. One of these procedures utilizes spontaneous sampling while the other is based on elicited imitation. Studies and clinical tests using these two methods separately were briefly discussed. Two widely used standardized tests are the DSS which incorporates spontaneous sampling and the ELI which is based on elicited imitation. Only four studies were found which directly compare spontaneous sampling with elicited imitation, and only one of these studies dealt with the DSS and the ELI. Since the DSS and the ELI are widely used clinical tests, there is a need for additional research comparing the two types of language measures.

PURPOSE OF THE STUDY

Spontaneous sampling and elicited imitation are two procedures frequently used in the assessment of the production of syntax in children. Two widely used clinical tests representing these procedures are the DSS, which incorporates spontaneous sampling, and the ELI, which utilizes elicited imitation. Several problems revealed in the review of the literature indicate that the DSS may be of questionable value. The first set of problems is inherent in the sampling of spontaneous speech in general. These problems include the effect of variables (for example, stimulus, examiner, and environment) on the quality of the language sample. Also, there is no guarantee that all the syntactical forms of interest will occur in a given sample. And one sample may not be typical of the child's usual language performance. Although some variables (eg. mood of the child) can also affect elicited imitation results, it is

easier to control many of them. For example, stimuli and instructions can be controlled by using the same items and directions with each presentation. With elicited imitation the examiner has control over item sampling because he determines what will be sampled. And since he picks the items, he can assess not only which syntactical forms the child uses but also which ones the child is capable of using.

The second set of problems includes those found with the DSS procedure itself. Although significantly high validity values have been found for the DSS (Koenigsknecht, 1974), not all of the reliability measures have been high. For example, low temporal reliability was indicated for short intervals of two weeks (Koenigsknecht, 1974) and interexaminer reliability was low (Cornelius, 1974). Johnson and Tomblin (1975) found that the sample size needs to be 3½ times larger than Lee recommended before only a very small reduction in the large error variance is achieved. On the other hand, validity and reliability values have all been significantly high for the ELI. The reliability measures include a high temporal reliability for a two week period (Carrow, 1974a) and a high interexaminer reliability (Cornelius, 1974). Another problem with the DSS procedure is that it is very time consuming. Cornelius (1974) found the average time for administering, transcribing, and scoring the DSS to be approximately 90 minutes, while it took only about 40 minutes to do the same with the ELI. Besides the reliability and the time element problems, fewer normative data are supplied for the DSS. Standardized norms are available for the total DSS

score but not for the individual grammatical categories. The clinician needs to depend on visual inspection alone for determining specific problem areas. Standardized norms are also provided for the total ELI error score, and although standard scores are not available for each subcategory, percentile ranks are given for each grammatical error score and each type error score according to age. These percentile ranks do need to be interpreted with caution (Carrow, 1974a); however, they give the clinician additional information other than visual inspection for determining specific problem areas. In addition to normative data, the DSS assesses fewer categories than the ELI. All of the grammatical categories of the DSS (see Table 1) except for the two question categories (interrogative reversals and wh-questions) are found in a similar classification on the ELI. Several sentences on the ELI do contain interrogative reversals and wh-questions so that the child's use of these two categories may be determined by visual inspection of those sentences. In addition the ELI also directly assesses articles, nouns, noun plurals, prepositions, adverbs, and contractions. The DSS uses the sentence point to partially account for those grammatical structures not included in the eight categories. However, this sentence point does not allow the clinician to determine additional problem areas such as noun plurals and prepositions. The errors on the ELI are classified on the basis of the grammatical category and the type of error (for example, substitution or omission). The verb category on the ELI can be further analyzed according to verb type (modal, auxiliary, copula, main verb, infinitive, and gerund),

verb context (declarative, interrogative, negative, and affirmative), and error type (tense, person, and number). No comparable procedure exists for the DSS. In addition to the above problems, the DSS is more difficult to score than the ELI. For the DSS the clinician has to look at each word in the sentence, determine if it fits in one of the grammatical categories, and then locate the word in the supplied chart before it can be scored. Occasionally a word cannot be found on the chart which involves attributing points to it subjectively. But on the ELI a matrix is provided as part of the scoring form which lists every word in each sentence under its corresponding category. All the clinician needs to do is mark the error under the category, determine the type of error, and count one point for each error.

The ELI does not have the above problems which relate to sampling, low reliability, excessive time requirements, and lack of enough analytical information, all of which indicate that the ELI may be a more beneficial tool than the DSS for the clinician to use in the assessment of the production of syntax in children. The ELI would be more efficient for diagnostic purposes if both tests are assessing the same abilities since the DSS has been shown to be a valid instrument for assessing the production of syntax in children. Koenigsknecht (1974) states that one way to determine the validity of a test is to compare it with other tests. Cornelius (1974) found a significantly high correlation of -0.79 between the total scores of the DSS and the ELI, indicating that they do provide

similar information. However, the total scores were derived from adding the scores for all the grammatical categories of each test even though the ELI has several categories (eg. nouns and prepositions) not found on the DSS. Therefore, this investigation involved comparing the total scores of the DSS and the ELI and the modified total scores based on only those categories (eg. verbs and negatives) which are produced on both tests to assess validity by discovering whether the procedures are measuring the same aspects of syntax.

The DSS (Lee, 1974) is a procedure used to analyze syntax produced by children, 2-0 to 6-11 years, in a spontaneous language sample which is tape recorded. A sample of 50 consecutive utterances containing both subjects and predicates is analyzed according to a weighted scoring system for the 18 grammatical categories listed in Table 1 with later developing structures receiving higher points. If a sentence meets all the adult standards of grammar, one sentence point is also given. The total mean score may be compared to supplied norms. The ELI (Carrow, 1974b) is a procedure used to analyze the grammar produced by children, 3-0 to 7-11 years, utilizing elicited imitation. A child is asked to imitate one phrase and 51 sentences which contain a wide range of syntactical complexity. The sentences for imitation are produced by the clinician and the procedure is tape recorded. The child's imitations are analyzed according to the category errors and the type errors made. The grammatical categories are listed in Table 1. The total error score, each grammar error score, and each type error score may be compared to the provided norms.

It was hypothesized that the DSS and the ELI would be found to be assessing similar features of syntax. If this hypothesis is valid, it is argued that the ELI would be considered a preferred clinical tool for assessing the production of syntax in children. The ELI would be preferred over the DSS because it does not have the problems of sampling, low reliability, excessive time requirements, and limited analytical information associated with the DSS.

CHAPTER II

MATERIALS AND PROCEDURE

DESCRIPTION OF SUBJECTS

Thirty-three boys participated in the present study. There were five boys ranging in age from 4-3 to 4-9, thirteen boys from 5-2 to 5-9, and fifteen boys from 6-0 to 6-11. The subjects were enrolled in Target Range Elementary School or in various preschools in Missoula. Each boy met the following criteria:

(1) obtained an IQ score above 85 on the Slosson Intelligence Test for Children and Adults (Slosson, 1963).

(2) passed a hearing screening test given by a certified speech clinician (ASHA CCC or Montana License), a certified audiologist, or the examining speech clinician within the present school year.

(3) was free from upper respiratory infection according to the child's mother or the examining speech clinician at the time this study was conducted.

(4) was judged to have language and cognitive skills appropriate to his age level by the child's teacher.

(5) was not receiving language therapy or resource room help at the time this study was conducted.

(6) had intelligible speech based on the judgment of the examining speech clinician.

(7) was from a monolingual home where Standard American English is spoken.

DESCRIPTION OF MATERIALS

All children were individually tested by the same examiner in a quiet room in the public school or preschools. The Slosson Intelligence Test for Children and Adults, the Developmental Sentence Scoring procedure, and the Elicited Language Inventory were administered to each child. The DSS and the ELI were recorded on high fidelity equipment. The order of presentation of the tests was the intelligence screening test, the DSS, and then the ELI with each subject. The DSS was given before the ELI because it was assumed that talking to the child about toys and pictures would not influence his ability to produce sentences on the ELI. On the other hand, if the ELI was presented first, the child may use more complex syntactical forms in the DSS sample as a result of being exposed to the ELI sentences. It was also assumed that the intelligence screening test would not influence the results of the DSS since the last 50 utterances were analyzed from the language sample. Each child was given a warm-up period of conversational speech to shift from the previous activity of giving short answers to questions before the sample was analyzed.

The DSS (Lee, 1974) was used to analyze the syntax produced by the subjects in a spontaneous language sample of 50 consecutive and different utterances. Each utterance must contain a noun and verb in subject-predicate relationship. First, the syntactic

structures were analyzed according to the 8 grammatical categories (see Table 1) specified in the standardized test. Weighted scores were assigned to the forms in each category with higher scores given to later developing forms. If a sentence met all of the adult grammatical standards, one additional sentence point was given. This analysis was done to obtain the total DSS score by adding the points for the 8 grammatical categories and the sentence points. The second procedure involved obtaining a modified total DSS score by adding the points for only 7 of the grammatical categories (see Table 2). The guidelines presented in Chapter II of Lee's book were utilized for the administration and scoring of the DSS. The stimulus materials for obtaining the language samples included toys (a doll family with furniture, a transport truck with removable cars, and a barn with farm animals), eleven story-action pictures taken from the preprimer series, We Read Pictures, We Read More Pictures, and Before We Read (Robinson, Monroe, and Artley, 1962), and the pictures of "The Three Bears" from What's Its Name (Utley, 1950). These materials were presented in the above order which is the way they were used in the standardization of the test. The language sample consisted of the last 50 sentences produced by each child, in accordance with the instructions for administering and scoring the standardized samples.

The ELI (Carrow, 1974a) was used to analyze the production of syntax in the subjects through elicited imitation. In accordance with Carrow's administration instructions, the stimuli which were produced by the examiner consisted of one two-word phrase and 51

sentences ranging in length from two to ten words with an average length of six words. The sentences were first analyzed according to the 12 grammatical categories (see Table 1) specified in the standardized test. The child's responses were compared to the model sentences produced by the examiner, and errors were classified based on these 12 categories. In accordance with the standardized procedure, one point was given to each error. This first procedure was followed to get the total ELI error score by adding all of the error points for each of the 12 grammatical categories. A second procedure involved obtaining a modified total ELI score by adding the points for 6 grammatical categories (see Table 2). The directions specified by Carrow in her manual were followed for the administration and scoring of the ELI.

TABLE 1. Classification of grammatical categories on the DSS and the ELI.

DSS	ELI
1. indefinite pronouns or noun modifiers	1. demonstratives
2. personal pronouns	2. pronouns
3. main verbs	3. verbs
4. secondary verbs	4. negatives
5. negatives	5. conjunctions
6. conjunctions	6. articles
7. interrogative reversals	7. adjectives
8. wh-questions	8. nouns
	9. noun plurals
	10. contractions
	11. prepositions
	12. adverbs

TABLE 2. Reclassification of grammatical categories on the DSS and the ELI for obtaining modified total scores.

DSS	ELI
1. indefinite pronouns or noun modifiers and personal pronouns	1. demonstratives and pronouns
2. main verbs and secondary verbs	2. verbs
3. negatives	3. negatives
4. conjunctions	4. conjunctions
5. wh-questions	5. wh-questions (wh-words in sentence #34, 38, 39, and 44)

STATISTICAL PROCEDURES

The Pearson product moment correlation was used to compare the total scores and the modified total scores for the ELI and the DSS. It was hoped that a higher correlation would be found between the modified total scores than between the total scores once the dissimilar categories were removed from the two procedures. Since the split-half reliability for the DSS is 0.73, the highest possible correlation coefficient that could be found between the DSS and the ELI is 0.85. This possible correlation was found by taking the square root of the lowest reliability of the two procedures (Nunnally, 1967).

The following guidelines (Silverman, 1977, p. 193) were used to interpret the strength of the relationship of the correlation coefficients found in this study:

1. Pearson r coefficients of less than 0.30 usually indicate that for most practical purposes no linear relationship exists between the attributes.
2. Pearson r coefficients between 0.30 and 0.50 usually indicate a weak linear relationship between the attributes.
3. Pearson r coefficients between 0.51 and 0.85 usually indicate a moderate linear relationship between the attributes.
4. Pearson r coefficients between 0.86 and 0.95 usually indicate a strong linear relationship between the attributes.
5. Pearson r coefficients higher than 0.96 usually indicate an extremely strong linear relationship between the attributes.

Using Silverman's guidelines it can be seen that due to the unreliability of the DSS, the maximum relationship that can be found to exist between the ELI and the DSS would be one of moderate degree. This author estimated that in order to say that the two procedures are measuring similar abilities, a correlation of at least 0.70 should be found. The computer at the University of Montana was used to determine the correlation coefficients.

A correction for attenuation formula (Nunnally, 1972) was utilized to estimate what the correlation of the ELI and the DSS would be if measurement error was not reducing the relationship.

Inter-rater reliability was achieved by comparing the scores obtained by 2 raters on 6 subjects. Both raters scored DSS and ELI samples until a minimum of 95% agreement was met on each procedure before this study was conducted. Then the second rater separately but concurrently scored every sixth subject plus the last subject to make certain a minimum of 95% agreement was maintained throughout the study. The Pearson product moment correlation was used to compare

the raters' scores for both procedures.

CHAPTER III

RESULTS

A Pearson product moment correlation was used to determine the strength of the relationship between the total scores and between the modified total scores on the DSS and the ELI. The correlation between the total scores was -0.1087 (nonsignificant), and the correlation between the modified total scores was -0.0969 (nonsignificant). After applying the correction for attenuation formula (Nunnally, 1972) to remove measurement error, the correlation between the total scores was -0.1303 (nonsignificant) and the correlation between the modified total scores was -0.1161 (nonsignificant).

The mean for the DSS total mean scores was 10.9697 with a standard deviation of 2.0914. For the ELI total scores the mean was 9.0909 with a standard deviation of 7.1430. The majority of the subjects' total scores on both the DSS and the ELI were skewed toward the high end of the standardized normative data curve (see Appendix B).

Since nonsignificant correlations were found between the total scores and between the modified total scores, Pearson product moment correlations were used to determine whether the similar grammatical categories of the DSS and the ELI also had nonsignificant correlations. All of the correlations between the grammatical categories of the DSS and the ELI were found to be nonsignificant (see Table 3).

TABLE 3. Pearson product moment correlations between grammatical categories on the DSS and the ELI.

DSS	ELI	Pearson r
1. indefinite pronouns or noun modifiers and personal pronouns	1. demonstratives and pronouns	1. -0.1816 (nonsignificant)
2. main verbs and secondary verbs	2. verbs	2. -0.2790 (nonsignificant)
3. negatives	3. negatives	3. 0.2373 (nonsignificant)
4. conjunctions	4. conjunctions	4. -0.0233 (nonsignificant)
5. wh-questions	5. wh-questions (wh-words in sentence #34, 38, 39, and 44)	5. coefficient unable to be computed*

*Since no errors were made on the wh-words in the ELI, a correlation coefficient could not be computed.

The Pearson product moment correlation was also used to determine inter-rater reliability. The two examiners obtained a correlation of 0.999 ($p < .001$) for the DSS and 1 for the ELI. More than 95% agreement was consistently maintained throughout the study for both procedures.

Appendix A contains the DSS and the ELI scores for all subjects. Appendix B contains a chart of the location of the subjects' total scores (ELI) and total mean scores (DSS) according to the standardized norms for the DSS and the ELI.

CHAPTER IV

SUMMARY AND DISCUSSION

SUMMARY

The purpose of this study was to determine if the Developmental Sentence Scoring procedure and the Elicited Language Inventory are assessing similar features of syntax. Thirty-three four, five, and six year old boys with normal language abilities, hearing within normal limits, and at least average intelligence participated in the study. The DSS and then the ELI were administered to each child. A Pearson product moment correlation revealed that a nonsignificant correlation exists between the two procedures with the above population. Inter-rater reliability was significantly high for the scoring of both the DSS and the ELI.

DISCUSSION

The results of this study do not support the hypothesis that the Developmental Sentence Scoring procedure and the Elicited Language Inventory are assessing similar features of syntax. Several possible explanations may be given for the low correlation found between these two procedures.

First, a nonsignificant correlation would exist between the two procedures if they were assessing different abilities. This low correlation may only appear when using a population with normal language skills because language disordered children may perform

similarly on both procedures due to their deficient language skills. In order to explain why this may be possible, language acquisition may be viewed as the development of a finite set of syntactic, semantic, and phonological rules from which an infinite number of utterances can be generated. The process appears to involve the child abstracting generalizations from the language he hears, forming hypotheses about language, testing these hypotheses through his own use of language, and then modifying the hypotheses based on the feedback he receives when his speech does not match the adult model. A language disordered child may be viewed as having an inadequate or incomplete set of rules for generating utterances. The child may be unable to perform the above processes of abstracting, hypothesizing, utilizing feedback, and re-hypothesizing. With a limited language system a language disordered child may respond in a similar manner to apparently dissimilar tasks. His language system may be limited enough so that he has not adequately developed the different abilities required for the DSS and the ELI and, therefore, responds on both procedures in a similar manner. If language disordered children did perform similarly on each of these procedures, the degree of correlation between the two would be stronger.

Two abilities that each procedure measures differently are competence and performance. The ELI samples the child's competence (what he is capable of using) in predetermined areas whereas the DSS measures the child's performance (what he uses) in a spontaneous manner. Several children did much better on the ELI than on the DSS which may have happened because these children were capable of using

more complex grammatical structures than they used in their conversational speech. This variability in performance on the two tests may have contributed in part to the low correlation. But there may be additional differences in abilities operating because several other children performed substantially higher on the DSS than on the ELI. If competence and performance were the only aspects which were operating differently on the DSS and the ELI, the child would most likely score about the same on each procedure (performance close to his capability to use grammatical forms) or score higher on the ELI than the DSS (performance below his capability to use grammatical forms). If the child used a substantially higher level of grammatical structures in his spontaneous speech as measured by the DSS, when his capability was relatively low as measured by the ELI, some other factors besides performance and competence are probably involved. For example, one factor which may also operate differently on each procedure and which may result in a lower score on the ELI is imitation which is further explained in the following paragraph.

Two additional abilities which may affect the score on the ELI are the ability to imitate and the auditory memory span required to repeat the sentences. The type of imitation that the ELI utilizes is elicited imitation which involves an adult asking the child to imitate him. Auditory memory span refers to the amount of information one can retain in the proper sequence. Two types of auditory memory are necessary to imitate the ELI sentences. One is short-term memory or the immediate recall of the words produced by

the examiner. The other is long-term memory which involves the memory of everything that is retained for more than a few minutes. The rules of language or, in this case, the knowledge of grammatical structures, are stored in long-term memory. The child uses this knowledge to impose a structure on what he has heard which makes it easier to recall the sentence. Therefore, the ability to repeat sentences on the ELI requires an interaction between short-term and long-term memory. If a child were deficient in any of these abilities (imitation and short or long-term memory), he may obtain a lower score on the ELI than the DSS. In contrast, Clark (1977) asserts that a child's imitations can be superior to his spontaneous productions. A child may be able to imitate grammatical structures that he does not understand well enough to use spontaneously. One explanation for imitation exceeding production is related to auditory memory. Dale (1972) states that if the sentence to be imitated is within the child's short-term memory span. it can be imitated perfectly because the meaning of the sentence does not need to be processed. Although there is controversy in the literature as to whether auditory memory span is auditorily based or linguistically based (Rees, 1973), the role auditory memory plays in the ELI has not been adequately investigated. Several children performed substantially higher on the DSS than the ELI, while several others performed substantially higher on the ELI than the DSS. It is possible that the abilities of imitation and auditory memory span may have contributed to some of these differences in performance on the two

procedures.

A second explanation for the results of this study is that the DSS and the ELI may be measuring different abilities with normal children than with language disordered children. It was explained previously that language disordered children may have an inadequate set of rules for generating utterances and forming hypotheses, etc. Their language system may be so different from the language system of normal children that the DSS and the ELI may be measuring different abilities with each population. Since these two procedures were standardized on normal populations, it does not necessarily mean that the normative data can be used with language disordered children although this is frequently done by clinicians. It's possible that inferences cannot be made from one population to the other because their language systems are very different. Thus if one study used all normal children and another study used half normal children and half language disordered children, different correlations could be found, especially if only small sample sizes were used. The fact that this study and Cornelius' study (1974) did find different correlations could be taken as an indication that the means for interpreting the responses of normal children are not necessarily transferable to language disordered children.

A third explanation for the low correlation found in this study is that one procedure may be sampling deep structures and the other may be sampling surface structures. The deep structure refers to the underlying component of a sentence which aids in determining the meaning of the sentence. The surface structure is the result

after the transformations (grammatical rules) have been applied to the deep structure. In the spontaneous utterance of a sentence the child needs to generate the surface structure of the sentence by applying transformations to the deep structure. His rule system and meaning system are involved in creating sentences. Therefore, the DSS may be sampling deep structures because the child needs to utilize the deep structures in his generation of sentences. On the other hand, the ELI may only be sampling the surface structures because if the sentences are within the child's immediate memory, he may be able to imitate them without processing the meaning of the sentences. The child may not need to understand the sentence, begin again with the deep structure, and apply his grammatical rules to reformulate the sentence. This explanation may explain why some children performed substantially higher on the ELI than on the DSS. If the ELI sentences were within their short-term memory span and they didn't need to process the meaning of the sentences in order to imitate them, they might be able to repeat the sentences in a tape recorder-like fashion. The ELI sentences range in length from two to ten words with an average of six words, but children are usually able to imitate more words when they are in well-formed sentences than when they are in strings of unrelated words. Because these subjects' scores did vary in opposite directions for the DSS and the ELI, they would contribute in part to the low correlation found between the two procedures.

A final explanation for the results of this study relates to problems with the DSS procedure discussed in Chapter I. Some of the

DSS samples may not have been typical of the child's conversational speech because of the inherent problem in sampling spontaneous utterances. And many forms which the child was capable of using on the ELI (eg. wh-question words and negatives) were not used or were rarely used in the DSS sample. Also, external variables (for example, the stimuli used) may have affected the DSS samples. Since several children did substantially better on the DSS than on the ELI, the above problems with the DSS procedure would probably not influence the results enough to cause such a low correlation. However, the effect of external variables could also have influenced how the children did on the ELI. For example, since the ELI was administered last, some children may have become fatigued or bored. But several boys performed substantially better on the ELI than the DSS so that it is quite unlikely that these variables would result in the low correlation between the two procedures.

Based on the results of this study it cannot be stated that the DSS and the ELI are assessing similar features of syntax in normally developing boys between the ages of 4 years 3 months and 6 years 11 months. The nonsignificant correlation and the extreme variability between several subjects' performance on the two procedures raise the question that different language abilities may be operating in each procedure even though each procedure may utilize apparently similar syntactical features (eg. verbs and pronouns). In other words, each procedure may be measuring similar features of syntax but through different abilities of the child to use these features.

Examples of abilities which may operate differently on the DSS and the ELI are competence and performance, imitation, and short-term or long-term auditory memory. Also, the ELI and the DSS may be measuring different abilities with language disordered children than with normal children since their language systems are very different. Or each procedure may be measuring different aspects of the language system. For example, the DSS may sample deep structures while the ELI may measure surface structures. The low correlation found in this study may be the result of a combination of the above explanations.

While the results of this study do not support the hypothesis that the DSS and the ELI are measuring similar features of syntax, Cornelius' (1974) results apparently do support this hypothesis. There are several factors which may account for the discrepancy between the two studies. The first factor is Cornelius' grouping of language disordered and normal children to run her correlation. Since Cornelius grouped the two populations together, her high correlation may have resulted from the extreme heterogeneity of the two populations. In other words, the large amount of variance between the two groups may have accounted for the high correlation between the DSS and the ELI. The present study used a homogeneous population and found nonsignificant correlations between the DSS and the ELI (see Table 3). If Cornelius had analyzed the correlations of each population separately, she may have also found nonsignificant correlations between the DSS and the ELI. The second factor

is the nature of the population tested in each study. The current study used 33 normally developing children, while Cornelius' study used 10 language disordered and 10 normal children. Different correlations could be found when using two different populations. For example, if the DSS and the ELI do assess different abilities but language disordered children perform similarly on both procedures, a study like Cornelius' study, which accumulated the scores for normal and language disordered children, could erroneously find a significantly high correlation. Besides the nature of the population, the size of the population could be another contributing factor. Although the sample size of this study was small (33), Cornelius' sample was even smaller (20, but with only ten from each of two populations). This author found some extreme variability among her 33 subjects in their performance of the two procedures. Since Cornelius used only ten normal children, she may have only sampled a portion of the variability found among normal children. A fourth factor which may cause some differences in the results between the two studies relates to the problems with the DSS procedure discussed in Chapter I. Examples of these problems include obtaining samples which are not typical of the child's conversational speech and the influence of external variables on the sample. External variables may also influence results on the ELI. Due to the wide variability among some of the subjects on the DSS and the ELI in this study, it is unlikely that these problems would cause the low correlation found. However, the effect these problems may have had on the results of Cornelius' study are not

known. The last factors which may partially account for the different results of the two studies relate to the differences in sex and ages between the subjects. The subjects who participated in this study were all boys ranging in age from 4 years 3 months to 6 years 11 months. Cornelius used 14 boys and 6 girls between the ages of 4 years 4 months and 5 years 9 months.

Two conclusions can be drawn from the results of this study. One is that it is important to use both a procedure that samples spontaneous speech and one that samples elicited imitation in assessing children's abilities to use syntax. It is important because each procedure may measure different abilities of the child or different aspects of language, and therefore, a child may perform significantly differently on each procedure. The second conclusion is that neither procedure is a definitive measure of a child's expressive use of syntax. Since numerous problems exist with the DSS procedure and since the procedures may be assessing different abilities of the child or different aspects of language, further research would need to be done in order to assert that one procedure can routinely be used in perference to the other procedure.

SUGGESTIONS FOR FUTURE RESEARCH

Since this study found a nonsignificant correlation between the DSS and the ELI, but another study (Cornelius, 1974) using a different population found a significantly high correlation between the two procedures, further research needs to be done to determine whether or not the two procedures provide similar information

regarding the child's expressive use of syntax. It is recommended that Cornelius' study be replicated with a large enough sample size of each population (normal children and language disordered children) so that correlations can be made for the combined population and for each individual population. This would enable comparisons to be made between the correlations for each group.

Since the possibility exists that the two procedures may be assessing different abilities for using syntax, more research needs to be done to discover what these abilities might be. For example, additional research on the processes of imitation and spontaneous production in general may provide some informatin. Detailed grammatical analysis of each would help. And the relationship between auditory memory span and the child's performance on the ELI needs to be better understood. Further research needs to be carried out to find out how procedures such as the ELI and the DSS differ as well as how they are similar.

BIBLIOGRAPHY

Berry-Luterman, L. and A. Bar. "The Diagnostic Significance of Sentence Repetition for Language-Impaired Children," Journal of Speech and Hearing Disorders, 36, 29-39, 1971.

Bloom, L. M. "A Comment on Lee's Developmental Sentence Types: A Method for Comparing Normal and Deviant Syntactic Development," Journal of Speech and Hearing Disorders, 32, 294-296, 1967.

Carrow, E. Elicited Language Inventory. Austin, Texas: Learning Concepts Inc., 1974a.

Carrow, E. "A Test Using Elicited Imitations in Assessing Grammatical Structure in Children," Journal of Speech and Hearing Disorders, 39, 437-444, 1974b.

Chomsky, N. Aspects of the Theory of Syntax. Cambridge, Mass.: The M. I. T. Press, 1965.

Clark, R. "What's the Use of Imitation?" Journal of Child Language, 4, 341-358, 1977.

Cornelius, S. "A Comparison of the Elicited Language Inventory with the Developmental Sentence Scoring Procedure in Assessing Language Disorders in Children," Master's thesis, Univ. of Texas, Austin, Tex., 1974.

Cowan, P. A., J. Weber, B. A. Hoddinott, and J. Klein. "Mean Length of Spoken Response as a Function of Stimulus, Experimenter, and Subject," Child Development, 38, 191-203, 1967.

Dale, P. S. Language Development: Structure and Function. Hinsdale, Illinois: The Dryden Press Inc., 1972.

Darley, F. L. and K. L. Moll. "Reliability of Language Measures and Size of Language Sample," Journal of Speech and Hearing Research, 3, 166-173, 1960.

Erwin, S. M. "Imitation and Structural Change in Children's Language." In E. H. Lenneberg (Ed.), New Directions in the Study of Language. Cambridge, Mass.: The M. I. T. Press, 1964.

Fraser, C., U. Bellugi, and R. Brown. "Control of Grammar in Imitation, Comprehension, and Production," Journal of Verbal Learning and Verbal Behavior, 2, 121-135, 1963.

Hahn, E. "Analyses of the Content and Form of the Speech of First Grade Children," Quarterly Journal of Speech, 34, 361-366, 1948.

Johnson, M. R. and J. B. Tomblin. "The Reliability of Developmental Sentence Scoring as a Function of Sample Size," Journal of Speech and Hearing Research, 18, 372-380, 1975.

Johnson, W., F. L. Darley, and D. C. Spriestersbach. Diagnostic Methods in Speech Pathology. New York: Harper and Row, 1963.

Koenigsknecht, R. A. "Statistical Information on Developmental Sentence Analysis." In L. L. Lee, Developmental Sentence Analysis: A Grammatical Assessment Procedure for Speech and Language Clinicians. Evanston: Northwestern University Press, 1974.

Lee, L. L. Developmental Sentence Analysis: A Grammatical Assessment Procedure for Speech and Language Clinicians. Evanston: Northwestern University Press, 1974.

Lee, L. L. "Developmental Sentence Types: A Method for Comparing Normal and Deviant Syntactic Development," Journal of Speech and Hearing Disorders, 31, 311-330, 1966.

Lee, L. L. The Northwestern Syntax Screening Test. Evanston, Ill.: Northwestern University Press, 1969.

Lee, L. L. "A Screening Test for Syntax Development," Journal of Speech and Hearing Disorders, 35, 103-112, 1970.

Lee, L. L. and S. M. Canter. "Developmental Sentence Scoring: A Clinical Procedure for Estimating Syntactic Development in Children's Spontaneous Speech," Journal of Speech and Hearing Disorders, 36, 315-340, 1971.

Leonard, L. B. "What is Deviant Language?" Journal of Speech and Hearing Disorders, 37, 427-466, 1972.

Longhurst, T. M. and T. A. M. Schrandt. "Linguistic Analysis of Children's Speech: A Comparison of Four Procedures," Journal of Speech and Hearing Disorders, 38, 240-249, 1973.

McCarthy, D. "Language Development in Children." In L. Carmichael (Ed.), Manual of Child Psychology. New York: John Wiley & Sons, Inc., 1954.

McCarthy, D. The Language Development of the Preschool Child. Minneapolis: The University of Minnesota Press, 1930.

McNeill, D. The Acquisition of Language: The Study of Developmental Psycholinguistics. New York: Harper & Row, 1970.

Menyuk, P. "Comparison of Grammar of Children with Functionally Deviant and Normal Speech," Journal of Speech and Hearing Research, 7, 109-121, 1964.

Menyuk, P. Sentences Children Use. Cambridge, Mass.: The M. I. T. Press, 1969.

Miner, L. E. "Scoring Procedures for the Length-Complexity Index: A Preliminary Report," Journal of Communication Disorders, 2, 224-240, 1969.

Minifie, F. D., F. L. Darley, and D. Sherman. "Temporal Reliability of Seven Language Measures," Journal of Speech and Hearing Research, 6, 139-148, 1963.

Nunnally, J. C. Educational Measurement and Evaluation. New York: McGraw-Hill, Inc., 1972.

Nunnally, J. C. Psychometric Theory. New York: McGraw-Hill, Inc., 1967.

Prutting, C. A., T. M. Gallagher, and A. Mulac. "The Expressive Portion of the NSST Compared to a Spontaneous Language Sample," Journal of Speech and Hearing Disorders, 40, 40-48, 1975.

Rees, Norma S. "Auditory Processing Factors in Language Disorders: The View from Procrustes' Bed," Journal of Speech and Hearing Disorders, 38, 304-315, 1973.

Robinson, H. M., M. Monroe, and A. S. Artley, Before We Read. Chicago: Scott, Foresman, 1962.

Robinson, H. M., M. Monroe, and A. S. Artley. We Read More Pictures. Chicago: Scott, Foresman, 1962.

Robinson, H. M., M. Monroe, and A. S. Artley. We Read Pictures. Chicago: Scott, Foresman, 1962.

Sharf, D. J. "Some Relationships Between Measures of Early Language Development," Journal of Speech and Hearing Disorders, 37, 64-74, 1972.

Shriner, T. H. "A Comparison of Selected Measures with Psychological Scale Values of Language Development," Journal of Speech and Hearing Research, 10, 828-835, 1967.

Shriner, T. H., "A Review of Mean Length of Response as a Measure of Expressive Language Development in Children," Journal of Speech and Hearing Disorders, 34, 61-68, 1969.

Shriner, T. H. and D. Sherman. "An Equation for Assessing Language Development," Journal of Speech and Hearing Research, 10, 41-48, 1967.

Siegel, G. M. and P. A. Broen. "Language Assessment." In L. L. Lloyd (Ed.), Communication, Assessment and Intervention Strategies. Baltimore: University Park Press, 1976.

Silverman, F. H. Research Design in Speech Pathology and Audiology. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1977.

Slosson, R. L. Slosson Intelligence Test for Children and Adults. Los Angeles: Western Psychological Services, 1963.

Templin, M. C. Certain Language Skills in Children: Their Development and Interrelationships. Minneapolis: The University of Minnesota Press, 1957.

Tyack, D. and R. Gottsleben. Language Sampling, Analysis, and Training. Palo Alto, Calif.: Consulting Psychologists Press 1974.

Utley, Jean. What's Its Name? Urbana, Illinois: University of Illinois Press, 1950.

Weiner, P. S. and W. C. Hoock. "The Standardization of Tests," Journal of Speech and Hearing Research, 16, 616-626, 1973.

Wilson, M. E. "A Standardized Method for Obtaining a Spoken Language Sample," Journal of Speech and Hearing Research, 12, 95-102, 1969.

Winitz, H. "Language Skills of Male and Female Kindergarten Children," Journal of Speech and Hearing Research, 2, 377-386, 1959.

APPENDIX A

ELI and DSS Scores

Subject	Age	DSS Total $\bar{X}$ Score	DSS Modified Total $\bar{X}$ Score	ELI Total Score	ELI Modified Total Score
1	4-3	11.60	10.00	7	1
2	4-6	10.80	9.70	2	0
3	4-7	9.50	8.30	6	5
4	4-7	11.94	9.90	21	17
5	4-9	10.04	8.48	15	13
6	5-2	11.14	10.20	20	16
7	5-3	8.54	7.80	12	10
8	5-4	11.72	10.56	11	8
9	5-6	7.80	7.04	11	6
10	5-6	9.78	9.06	32	26
11	5-7	10.14	9.40	9	3
12	5-7	14.94	14.26	20	15
13	5-8	12.04	11.22	9	6
14	5-8	11.00	10.20	4	3
15	5-8	9.68	8.84	11	7
16	5-9	11.92	11.00	11	8
17	5-9	9.90	9.16	4	4
18	5-9	9.40	8.44	8	7
19	6-0	8.48	7.72	12	6
20	6-1	11.04	10.26	5	5
21	6-2	11.22	10.34	20	10
22	6-2	16.70	15.84	2	2
23	6-2	10.38	9.94	6	4
24	6-4	7.30	6.40	4	4
25	6-4	12.00	11.34	12	8
26	6-4	9.28	8.34	8	6
27	6-4	10.08	8.96	3	2
28	6-5	11.62	10.64	5	4
29	6-6	11.88	10.94	0	0
30	6-7	9.80	9.06	3	1
31	6-8	12.24	11.26	3	3
32	6-10	11.50	10.60	3	3
33	6-11	16.60	15.84	1	1

APPENDIX B

ELI Total Scores and DSS Total Mean Scores
Compared to Standardized Norms

-2 SD	-1 SD		Mean		+1 SD		+2 SD		+3 SD
23	13	9	25	12	19	17	8		
3	21	18	11	2	24	1	16		
				5	29	7	27		
				22	6	10	28		
				31	30	14	4		
				33		15	20		
				26		32			
7		2		16		8		17	13
18		33		23		12		25	
		10		24		5		14	
		11		27		19		20	
		26		28		21			
		32		29		22			
				31					
				1					
				3					
				4					
				6					
				9					
				15					
				30					

NOTE: The subjects' total scores were ranged according to the standardized norms for the DSS and the ELI in order to show the large amount of variance present. The numbers correspond to the subjects in the following age groups:

4-3 to 4-9: #8, 12, 16, 17, 25
5-2 to 5-9: #2, 5, 13, 19, 21, 22, 23, 24, 27, 28, 29, 31, 33
6-0 to 6-11: #1, 3, 4, 6, 7, 9, 10, 11, 14, 15, 18, 20, 26, 30, 32

The effect of the age variable is removed from the sample in this chart, and the scores still fluctuate among the subjects in each of the three age groups.