

A COMPARISON OF CHILDREN'S LANGUAGE SAMPLES
COLLECTED IN FOUR ELICITATION PROCEDURES

by

Susan Grubb

B. A., University of Arizona, Tucson, Arizona 1971

A MASTER'S THESIS

submitted in partial fulfillment of the

requirements for the degree

MASTER OF ARTS

Department of Speech

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1973

Approved by:


Major Professor

LD
2668
T4
1973
G7
C.2
Doc.

TABLE OF CONTENTS

| | Page |
|--|------|
| I. Introduction. | 1 |
| II. Method. | 4 |
| Subjects | 4 |
| Design | 5 |
| Experimental Facility. | 6 |
| Treatments, Materials, and Procedure | 6 |
| Protocol Preparation | 7 |
| III. Results | 9 |
| IV. Discussion. | 15 |
| V. Appendices. | 21 |
| Review of Pertinent Literature | 21 |
| Age and IQ of Subjects | 27 |
| Peabody "I Wonder" Series. | 28 |
| Compound Word Lists. | 29 |
| VI. References. | 32 |

LIST OF TABLES

| Number | Page |
|---|------|
| 1. Assignment of the twenty-four subjects to the six Latin squares, according to Level and Age group characteristics. Treatment Order for each subject is also given. | 5 |
| 2. Mean squares from the analysis of variance for total number of words (TNW), mean length of utterance (MLU), Carroll type-token ratio (CTTR), and length-complexity index (LCI). | 12 |
| 3. MI-AB Level I and III mean scores on the total number of words (TNW), mean length of utterance (MLU), Carroll type-token ratio (CTTR), and length-complexity index (LCI) in the four Treatments. . . . | 13 |
| 4. Mean scores for all subjects on the total number of words (TNW), mean length of utterance (MLU), Carroll type-token ratio (CTTR), and length-complexity index (LCI) in the four Treatments. | 14 |

ACKNOWLEDGEMENTS

I sincerely thank the chairman of my committee, Dr. Thomas M. Longhurst, for his unfailing patience, guidance, and encouragement during the preparation of this thesis, and for his help in its design.

I would also like to thank the members of my committee, Dr. Norma Bunton, Dr. Leo F. Engler, and Dr. Bruce C. Flanagan, for their attention to my work.

Special appreciation is extended to Joan Lemon and Jan Fritz for their careful study of the MLU, to Barbara Sargent for her painstaking work on reliability, and to Meg Marmor for her assistance in the preparation of many protocols.

Thank you, Leslie Lea Grubb, for putting up with your student mother.

Most especially, I would like to thank those significant others whose persistent moral support has insured the completion of this thesis.

INTRODUCTION

An adequate assessment of a child's verbal maturity requires an evaluation of his expressive oral language. As yet no standardized method has been developed for collecting oral language samples. Some of the variables inherent in the elicitation situation include the verbal behavior of the examiner, the stimulus materials presented to the subject, and the type of situation in which the subject is induced to speak, i.e. picture or object description, play therapy, conversation, or free play. The possibility that different collection procedures produce measurable differences in the respondent's language must be considered. Results of independent investigations are noncomparable to the extent that stimulus variables in the elicitation situation have not been held constant. Furthermore, it is desirable that the standardized procedure be demonstrated as the most efficient in eliciting from the child a sample of his best expressive oral language.

Several recent investigators have directly measured the effect of stimulus variation upon the language samples produced by their subjects.

Cowan, Weber, Hoddinott, and Klein (1967) presented ten magazine cover pictures to normal grade school children. They found that the scores for mean length of response (MLR) of all responses to two of these pictures were significantly different from the scores for MLR

of all responses to the other eight pictures.

Mintun (1968) elicited language samples from thirty educable mentally retarded children whose mean age was 8-5 years. Subjects were randomly divided into three groups and each group was presented only one type of stimulus. Her three stimulus conditions were toys, color photographs of these toys, and twenty-second color films of the real item represented by the toys. For example, a film was taken of a real horse. Results showed that the length-complexity index (LCI) score of the group responding to photographs was significantly different and smaller than the LCI score of the group responding to films. The difference between the LCI scores of the photograph group and the toy group was not significant. No significant differences were found among the MLR scores of the three groups. The scores for total number of words (TNW) and number of different words (NDW) were significantly different and larger for the toys condition, than for either films or photographs. TNW and NDW were greater and significantly different in the film condition as compared to photographs.

Strandberg (1969) elicited language samples from thirty normal children with a mean age of 5-1 years. Her stimulus conditions, division of subjects, and presentation of materials were identical to Mintun. She found no significant differences in the LCI and MLR scores for the three groups. The scores for TNW and NDW were significantly different and smaller in the photograph elicitation condition than in the other two conditions, which were not significantly different from each other.

Ahmed (1973) obtained language samples from thirty-two mentally retarded children who were residents of a state institution. Sixteen

of these children were classified as trainable and sixteen as educable. The average age of all the children was 14-4 years. Her stimulus items consisted of two sets of multi-colored pictures, presented as slides, from the Peabody Language Development Kit (Dunn and Smith, 1966). One set was single object pictures and the other multi-object pictures. Subjects were randomly divided into two groups composed of equal numbers of educable and trainable children. Each treatment group was presented with only one set of pictures. She found that the multi-object pictures yielded significantly different and higher TNW, Carroll type-token ratio (CTTR), and LCI scores for both the trainable and educable children. There was no significant difference in the MLR scores of the two treatment groups.

The results of three of these investigations (Cowan et al., Mintun, and Ahmed) demonstrated that variations in stimulus conditions produce measurable differences in obtained language samples. The LCI, rather than the MLR, appeared to be the more sensitive measure of real differences between language samples for Mintun and for Ahmed. Although Cowan et al. found some significant differences among language samples using the MLR as their measure, these differences might have been even more outstanding had they used the LCI. In addition it is possible that the language responses of retarded children, as opposed to normal children, are more easily influenced by variations in stimulus conditions. Two of these investigations (Cowan et al.; Ahmed) have shown that differences between pictures are important, while the other two (Mintun; Strandberg) present conflicting evidence of the differential stimulus value of pictures as compared to toys and films.

The present study was designed to answer three questions:

1. Do retarded subjects of different age or intellectual level respond differently to unstructured and structured elicitation situations?
2. Are speech samples that are collected in structured situations comparable to samples collected in less structured situations?
3. When oral language samples are subjected to linguistic analysis, which linguistic measures best evaluate differences between the samples?

METHOD

Subjects

Twenty-four residents of a state institution for the mentally retarded were used as subjects. They were divided equally into two Levels, Level I and Level III, according to their measured intelligence and adaptive behavior (MI-AB) (Heber, 1959), as determined by previous psychological evaluation. The mean IQ for all Level I subjects was 79, and the mean IQ for all Level III subjects was 45. The chronological ages of the subjects ranged from 9-9 to 14-7 years. Within each Level the subjects were divided by age into three Age groups, so that for each MI-AB Level there were three Age groups with mean ages of 10-2, 11-3, and 14-0 years. Sex was not controlled as a main variable. Eight of the twenty-four children were girls. Individual age and IQ characteristics of each subject are given in Appendix B.

All of the subjects were given a hearing screening and were able to respond at 500, 1K, 2K, and 4K at 30 dB ISO 1964.

Prior to the experiment, each subject met briefly with the examiner in order to determine whether he would be cooperative, and to determine if his speech was sufficiently intelligible to allow for

**THIS BOOK
CONTAINS
NUMEROUS PAGES
WITH DIAGRAMS
THAT ARE CROOKED
COMPARED TO THE
REST OF THE
INFORMATION ON
THE PAGE.**

**THIS IS AS
RECEIVED FROM
CUSTOMER.**

later transcription.

Design

Each of the twenty-four subjects responded to four elicitation situations (Treatments) in a counter-balanced order, which was determined by a 4 X 4 Latin square design. In all there were six Latin squares, three Age group squares at each MI-AB Level. The assignment of the subjects to the Latin squares, according to Level and Age group, is presented in Table I. The order in which each subject responded to the Treatments is also shown.

TABLE 1

LATIN SQUARES

Assignment of the twenty-four subjects to the six Latin squares, according to Level and Age group characteristics. Treatment Order for each subject is also given.

| | | Levels | |
|------------|------|-------------------|---------------------|
| | | I (Mean IQ 79) | III (Mean IQ 45) |
| | | Treatment Order | Treatment Order |
| Age groups | 10-2 | S-1 A D B C | S-13 A D C B |
| | | S-2 D C A B | S-14 D C B A |
| | | S-3 B A C D | S-15 C B A D |
| | | S-4 C B D A | S-16 B A D C |
| | 11-3 | S-5 A D B C | S-17 C A D B |
| | | S-6 D C A B | S-18 D B A C |
| | | S-7 B A C D | S-19 A C B D |
| | | S-8 C B D A | S-20 B D C A |
| | 14-0 | S-9 A B C D | S-21 C A B D |
| | | S-10 B C D A | S-22 D B C A |
| | | S-11 C D A B | S-23 A C D B |
| | | S-12 D A B C | S-24 B D A C |

Experimental Facility

Language samples for the four Treatments were collected in two experimental rooms. Room 1, the "waiting" room, contained a table, two chairs, and a microphone hidden beneath the table. The tape recorder was located in an adjacent room. Room 2 also contained a table and two chairs, but the microphone was not hidden. The tape recorder and a 2 X 2 inch slide projector were housed in an adjacent room to minimize noise. The slides were projected through a one-way mirror into room 2 and were programmed by remote control.

Treatments, Materials, and Procedure

Treatment A (Object Elicitation): With the subject and the examiner both seated on the floor in the experimental room, the subject was presented with eleven objects, brought from storage behind a screen one at a time in random order. The objects were a spoon, a plate, a volleyball, a pair of scissors, a small red box, a watch, a pen, a pair of sunglasses, a man's shoe, a candy bar, and a whistle. The instructions were as follows:

We are going to play a game. I will show you an object and you are to tell me all you can about it. Don't just name the object, but try to string words together and tell me about it.

The examiner attempted to be vaguely encouraging, but did not give any specific reinforcement, or prompt the child to continue speaking by asking any questions.

Treatment B (Picture Elicitation): Each child was brought to experimental room 2 and seated at the table. He was then shown ten pictures, presented as slides, from the Peabody Language Development Kit (Level 2) (Dunn and Smith, 1967). These were colored, story

situation slides, which used a variety of subjects and settings (see Appendix C). The order of presentation of the ten pictures was the same for each child. The same type of instructions as in Treatment A were given. Again, the examiner offered no prompting or reinforcement.

Treatment C (Adult-Child): The subject and the examiner were seated in experimental room 2. No specific stimulus materials were present. The examiner asked the subject about a variety of topics including his family, his school activities, sports, TV programs, and other appropriate subjects in an effort to elicit conversation. All speech from the child, both spontaneous and as responses to the examiner's questions, was considered as part of the language sample.

Treatment D (Child-Child): Two subjects, of the same age and MI-AB Level, were brought by the examiner to experimental room 1. They were told that because of scheduling difficulties they would have to wait a few minutes there before they could meet with the examiner. The subjects were then left alone, and allowed to converse in any manner they wished. Magazines and other objects that might discourage conversation were purposely excluded from the room. After an adequate sample had been collected from both subjects, they were taken from the room.

Protocol Preparation

Four linguistic analyses were computed from the protocols. These were (a) the total number of words (TNW) in each protocol, (b) the mean length of the first fifty utterances (MLU), (c) a Carroll type-token ratio (CTTR), and (d) a length-complexity index score (LCI).

The guidelines presented by Siegel (1963) were used in counting the words for the TNW measure. In some cases an arbitrary decision was required concerning compound words (see Appendix D). The TNW was viewed as a quantitative measure of language output.

For MLU, an utterance was defined as a unit of spoken language preceded and followed by a breath or a pause, or terminated by some change in inflection. MLU has long been used as a measure of language ability and is one of the few language measures for which any normative data is available.

Some of the protocols contained less than fifty utterances, the number usually recommended for a reliable MLU (Darley and Moll, 1960). An independent study of the language protocols of one hundred children has shown that an MLU based on as few as fifteen utterances does not appear to differ substantially from an MLU based on fifty utterances (Lemon and Fritz, 1973). Their analysis demonstrated that an MLU computed on the first fifteen utterances has a part-whole correlation of .89 with the MLU computed on the first fifty utterances. Furthermore, the confidence interval of the MLU of the first fifteen utterances did not vary as much as a half of a word from the MLU based on fifty utterances.

In view of these findings, and in order to utilize data from all of the subjects, it was decided in the present investigation to compute each subject's MLU on whatever number of utterances was available for him, keeping within a range of fifteen to fifty utterances. Essentially this same procedure was followed by Sharf (1972).

A CTRR was computed for each entire protocol. This ratio may be considered as a measure of the diversity of expressive vocabulary.

Because each of the protocols contained a different number of words it was not possible to use a traditional type-token ratio (Johnson, 1944), which is highly dependent on sample size (Carroll, 1964, p. 54). The formula used to compute CTTR was: $CTTR = \text{types} / \sqrt{2X \text{ tokens}}$ and is approximately independent of sample size (Carroll, 1964).

Finally, the protocols were resegmented into "sentences" according to Miner (1969, p. 228) for the LCI. In this case segmentation was more dependent upon syntactic and semantic unity, than upon the surface features of pause and inflection. The first ten "sentences" segmented for LCI were discarded, and the measure was computed on the eleventh through twenty-fifth sentence of each protocol. Griffith and Miner (1969) established that fifteen sentences were sufficient to compute a reliable LCI score. Analysis by LCI provides for a composite evaluation of utterance length and utterance complexity. LCI score is the sum of noun phrase (NP) points plus verb phrase (VP) points plus additional points (AP) for each sentence divided by the number of sentences (NS). Written as a formula, $LCI = NP + VP + AP/NS$.

RESULTS

As a measure of inter-scorer reliability two protocols were chosen at random from each of the treatments. These eight protocols were resegmented and rescored for both MLU and LCI by a senior in Speech Pathology, after receiving verbal instructions from the examiner. The average percentage of segmentation agreement for MLU and LCI was 83% and 88% respectively. The percentage agreement for scoring both MLU and LCI was 100%. Counting words or different words

has been shown to be highly reliable by other authors (Longhurst and Siegel, 1973).

A total of ninety-six language samples were collected, each of the twenty-four subjects having responded in four elicitation situations. Because of the design of this experiment it was possible to analyze differences in the language samples resulting from Level, Age group, Order of Treatment presentation, and Treatment variables, and their interactions. Means from the four linguistic analyses for each Latin square were subjected to a series of analysis of variance (Cochran and Cox, 1957, p. 117). The mean squares from the analysis of variance are reported in Table 2. Whenever these analyses resulted in an F ratio significant at the .05 level or better a least significant difference procedure (LSD) (Fryer, 1966, p. 260) was used to compare the means at the .05 level.

The average performance of all Level I subjects was higher than the average performance of all Level III subjects regardless of the linguistic measure used to evaluate performance and regardless of the Treatment in which the language sample was collected. These differences were significant at the .01 level of confidence. Mean scores for Levels I and III are presented in Table 3.

The only significant Age difference was found in the CTTR scores, when Age group scores from Level I and Level III were combined. The mean CTTR score for 10 year olds was 4.22, the mean for 11 year olds was 5.03, and the mean for 14 year olds was 4.7. The difference between the mean of the 10 and 11 year olds was significant at the .05 level.

There was no significant interaction between Level or Age group

and Treatments. The effects of the Treatments were the same for both Levels and for all three Age groups.

There was no significant Order effect. Performance did not vary significantly according to the Order in which the subjects participated in the Treatments.

When the scores of all Age groups and Levels were combined, significant differences were found among the language samples collected in the four Treatments for all four linguistic measures. Table 4 presents the mean of all subjects in each Treatment for each linguistic measure. TNW showed no significant difference between adult-child and child-child, but both of these Treatment means were higher and significantly different at the .01 level from the pictures or objects Treatment means. MLU showed no significant difference between child-child and pictures, both of which were higher than adult-child and objects. This difference was significant at the .01 level. CTTR was significantly different at the .01 level of confidence in all Treatments, with the rank order of high to low CTTR being adult-child, child-child, pictures, and objects. LCI showed no significant difference between adult-child and child-child, both of which produced higher and significantly different scores than objects. The LCI score for the picture Treatment was not significantly different from any of the other three Treatments.

TABLE 2

MEAN SQUARES

Mean squares from the analysis of variance for total number of words (TNW), mean length of utterance (MLU), Carroll type-token ratio (CTTR), and length-complexity index (LCI).

| Source | df | Linguistic Measures | | | |
|----------------------|----|---------------------|-------------|-------------|-------------|
| | | TNW | MLU | CTTR | LCI |
| Squares | 5 | | | | |
| Level | 1 | ** 298608.00 | ** 37.54 | ** 21.67 | ** 99.67 |
| Age | 2 | 20568.00 | 1.01 | * 3.50 | 5.17 |
| Level X Age | 2 | 100912.00 | 6.40 | 3.29 | 13.46 |
| Subjects/ Squares | 18 | 36413.51 | 2.76 | 1.25 | 6.31 |
| Order/ Squares | 18 | | | | |
| Order | 3 | 30904.00 | 1.71 | .35 | 4.21 |
| Order X Square | 15 | 18521.00 | .47 | .34 | .79 |
| Treatment | 3 | ** 320529.51 | ** 2.00 | ** 23.11 | ** 10.43 |
| Treatment X Square | 15 | 21853.61 | .85 | .92 | 2.07 |
| Error | 36 | 13406.80 | .60 | .55 | 1.84 |

*p < .05

**p < .01

Table 3

MI-AB Level I and III mean scores on the total number of words (TNW), mean length of utterance (MLU), Carroll type-token ratio (CTTR), and length-complexity index (LCI) in the four Treatments.

| Treatment | Level | TNW | MLU | CTTR | LCI |
|------------------------|-------|--------|------|------|------|
| Child- Child | I | 439.33 | 4.22 | 5.31 | 5.87 |
| | III | 315.00 | 3.02 | 4.36 | 4.08 |
| Adult- Child | I | 403.58 | 3.62 | 6.02 | 5.25 |
| | III | 375.00 | 2.52 | 5.76 | 3.19 |
| Picture Elicitation | I | 318.33 | 4.28 | 4.95 | 6.19 |
| | III | 155.83 | 2.79 | 3.67 | 3.56 |
| Object Elicitation | I | 214.58 | 3.70 | 4.21 | 4.39 |
| | III | 83.83 | 2.48 | 2.90 | 2.72 |

Note.- Each Level I mean is significantly different from the corresponding Level III mean at the .01 level of confidence.

Table 4

Mean scores for all subjects on the total number of words (TNW), mean length of utterance (MLU), Carroll type-token ratio (CTTR), and length-complexity index (LCI) in the four Treatments.

| Treatment | TNW | MLU | CTTR | LCI |
|---------------------|-----------------------|------------------------------|---------|------------------------------|
| Rank of Means | <u>C</u> <u>D</u> B A | <u>D</u> B <u>A</u> <u>C</u> | C D B A | <u>D</u> B <u>C</u> <u>A</u> |
| Child-Child | 377.16 ^A | 3.62 ^A | 4.83 | 4.97 ^A |
| Adult-Child | 389.29 ^A | 3.07 ^B | 5.89 | 4.22 ^A |
| Picture Elicitation | 237.08 | 3.53 ^A | 4.31 | 4.87 ^{AB} |
| Object Elicitation | 149.21 | 3.09 ^B | 3.55 | 3.55 ^B |

Note.- Means within a column which share a common superscript are not significantly different.

DISCUSSION

The difference in performance between Level I and Level III subjects was found to be significant for each Treatment and for each linguistic measure. This difference was predicted on the basis of the difference in mean IQ between the two Levels. More importantly, there was no significant interaction between Level and Treatment. This finding concurs with that of Ahmed (1973).

A significant difference in performance among Age groups was found for only one linguistic measure, the CTTR. It is difficult to understand why performance improved from the 10 to 11 year Age group, and then fell below the 11 year Age group mean for the 14 year old group.

No significant Order effect was found. This finding gives support to the importance of the Treatments themselves, rather than the order in which they were presented.

One of the main questions this experiment was designed to answer was whether speech samples collected in structured situations differ from those collected in less structured situations. The four Treatments used in this research represent part of the continuum from completely structured to completely unstructured.

Treatment B (Picture Elicitation) and Treatment A (Object Elicitation) were highly structured in that the subjects were required to respond to specific stimulus materials and that they were given specific instructions prior to their verbalization. They were partially unstructured because the examiner did not attempt to modify or prolong their speech once they had begun responding. Treatment C (Adult-

Child) was less structured because although the subject was asked to respond to specific questions, he was also free to change the subject and talk about anything he wished. In this case the examiner could prolong the conversation by simply asking another question and this was a means of modifying the subject's total verbal output. Treatment D (Child-Child) was the least structured of all the Treatments. There were no specific instructions to the subjects other than asking them to stay in the room. There was no control over either the topics of conversation or the amount of verbalization from each subject. Although social interaction was not analyzed in this experiment it was interesting to note the relationships which developed between the paired subjects in Treatment D. In most cases they conversed as amiable peers, but in a few instances one of the subjects assumed an authoritative role, and would question, advise, and sometimes threaten the other subject. Treatment D was not completely unstructured (naturalistic) because of the confinement of the subjects in an unfamiliar room.

The analysis of results showed that there were significant differences in the oral language samples collected in these four situations. Each of these Treatments will be discussed individually in terms of the linguistic scores received by the subjects in the Treatment.

Treatment A (Object Elicitation) elicited a smaller TNW and CTR in comparison to each of the other Treatments. Utterances were shorter and less complex than in the other Treatments. Subjects tended to name the objects and give a brief description and statement of function. Perhaps verbalization beyond that requires an ima-

gination or experiential history which this Treatment did not tap or these subjects did not possess. (Unless you are a shoemaker or Cinderella there is just so much you can say about a shoe!) The objects were chosen because of their familiarity to all the subjects, and in most cases, their lack of identification with either sex. It could be concluded that the subjects simply were not sufficiently interested in these objects to discuss them thoroughly.

Treatment B (Picture Elicitation) elicited a larger TNW and CTTR than Treatment A. The MLU score was higher than either Treatments C or A, and not significantly different from Treatment D. The LCI score was not significantly different from any of the other Treatments. This Treatment, which is probably the most frequently used in collecting language samples, appears to elicit "average" performance from all subjects, in relation to the other Treatments used in this experiment.

Treatment C (Adult-Child) elicited a larger CTTR than any of the other Treatments. This is probably because in asking questions of the child, the examiner was modeling new vocabulary which the child could incorporate into his answers. The fact that the LCI score was not significantly different than in either Treatment B or D may give evidence that while specific vocabulary items can be immediately imitated from the model, grammatical patterns are not immediately imitated. It could be concluded that the examiner can readily modify the subject's expressive vocabulary, but not the grammatical complexity of his response. The larger TNW of this Treatment, although it was not significantly different from that of Treatment D, could also have resulted from the examiner's option to elicit more responses from the

subject by additional questioning. That many of these responses were "yes" and "no" is reflected in the low MLU score. Had these responses been eliminated from Treatment C protocols before the MLU was calculated, the MLU would probably not have been significantly different than that of either Treatments D or B, as was the case for the LCI score. The reduced MLU in Treatment C reflects the completely unedited population of responses upon which it was calculated.

Treatment D (Child-Child) elicited consistently high scores for the measures TNW, MLU, and LCI. The CTTR was lower than in Treatment C, but it was still higher and significantly different from either Treatments B or A. Although the subjects could act as vocabulary models for each other, there was less likelihood that they would use a word that the other child did not already know and use. It is interesting to note that both the MLU and the LCI scores were greater in this condition than in any other, which may indicate a tendency toward more advanced communication in a situation where no adult is present.

This experiment has demonstrated that real differences do exist in the language samples collected in various elicitation situations. Less structured, conversational settings generally elicited language of greater quantity and complexity than the more structured, task-oriented settings.

The linguistic measures applied to these samples were in every case sufficiently sensitive to indicate at least some differences between Treatments. TNW and CTTR seem to be more sensitive to situational variables than MLU and LCI. An interpretation of this finding is that vocabulary and total verbal output are most easily modified

by the situation. Note that the TNW more than doubles from Treatment A to Treatments D and C (see Table 4). MLU was also affected although no Treatment produced a mean gain of more than a half a word. Cowan et al. (1967) found that the MLU changed by as much as four words in language samples elicited by different pictures. From this they concluded that the MLU was not an inherent characteristic of the child, remaining fairly constant from situation to situation. The results for MLU presented in Table 3 however do not support their findings.

Contrary to previous research (Mintun, 1968; Ahmed 1973) of all the language measures LCI differentiated least between Treatments in the present experiment. It is concluded from this that the grammatical complexity of a child's speech is his most stable verbal characteristic, and is least subject to immediate modification by situational variables.

In comparing language samples collected in different situations greater changes would be expected in TNW and CTTR, than in MLU and LCI. A substantial change or difference in these latter two scores can more confidently be interpreted as indicating a real difference in language ability, either in one child over time, or between two groups of children. Fluctuation in CTTR and TNW might result from real differences in language ability, but they might also be simply a reflection of situational variables. Siegel and Harkins (1963) found that TTR was very sensitive to differences in experimental situations.

As a research or clinical tool the analysis of a language sample is one means of evaluating maturity in language skills. For this

purpose an analysis of the subject's best speech would give the greatest amount of information concerning which language skills he already possessed, which were emerging, and which had not yet been learned. The language samples elicited in Treatments D and C elicited the best performance for the subjects in this experiment. Some thought should be given however, to the efficiency of the collection procedure. Obtaining a language sample as in Treatment D requires the availability of a peer for the subject. Excellent recording devices are also necessary because later analysis depends upon precise transcription from the tape. These prerequisites may be impossible to satisfy. Therefore, Treatment C is recommended as the most efficient method of collecting an oral language sample, with the considerations that the CTTR may be inflated by imitation of adult vocabulary, and that the MLU may be decreased when "yes" and "no" responses are included in its calculation.

A question posed for further research is whether the subject's language remains constant in conversations with different examiners. If it were found that the subject's language did not vary as a function of the examiner, then Treatment C could also be recommended as a standardized method of collecting an oral language sample. Since it does not require specific materials, which may not be universally available, it could be easily replicated. Cowan et al. (1967) found that the MLU varied as much as four words for homogeneous populations tested by different examiners, but this was in a picture description, and not a conversational, situation. Their results are further confounded since each examiner scored his own protocols.

APPENDIX A

Review of Pertinent Literature

Investigators who are interested in the development of children are concerned with the language abilities of children because language is a fundamental component of intellectual and adaptive behavior.

Many investigations have begun with the collection of an oral language sample from a child. This sample was then analyzed in terms of its quantity, complexity, and frequently its social appropriateness or function.

The interest of early investigations was focused upon what type of language was characteristic of a certain population or group of children. For instance, McCarthy (1930) studied normal children from one and a half to four and a half years. Day (1932) studied twins from ages two to five. Davis (1937) compared the language development of twins, singletons, and only children from ages five and a half to nine and a half. Templin (1957) studied normal children from ages three to eight.

These investigators were not primarily concerned with the possibility that a variation in the procedure of eliciting a language sample would produce a variation in the language elicited. However, in order to compare the results of different investigations it is necessary to standardize the variables inherent in the elicitation situation. Obvious variables are the experimenter, the procedure, the in-

structions, and the materials. Each of these variables must be systematically studied in order to discover its significant effect upon the language produced by a particular subject population.

Menyuk (1963) was the first investigator to systematically vary the stimulus situation. The purpose of her investigation was to analyze and describe the language of nursery school and first grade children in terms of a generative model of grammar. Her three stimulus situations included spontaneous response to The Blacky Pictures (Blum, 1950), conversation with an adult, and participation in a role-playing activity with other children. Menyuk does not tell us why she used three stimulus situations for each child. It is likely that she wanted to collect as much language as possible. Variations in language produced by the different elicitation situations were not reported.

Minifie, Darley, and Sherman (1963) collected language samples from forty-eight normal five-year-old and forty-eight normal eight-year-old children. Their primary purpose was to investigate the temporal reliability of the language measures used to analyze the samples. Each child responded to three sets of pictures, two of which were compiled by the investigators. The other set was the Children's Apperception Test Cards (Bellak and Bellak, 1950). Results of this study were reported only in terms of temporal reliability. Separate means for the three elicitation times were not reported.

Wilson (1969) used the Picture Story Language Test (Myklebust, 1965) to obtain a language sample from forty normal subjects ages three to seventeen. The procedure and instructions were constant for all of the subjects, and samples were analyzed for length and complexity

of the language produced. Although Wilson has recommended this as a standardized procedure, it cannot be accepted immediately without some further comparison with other elicitation procedures as Griffith (1969) points out.

Barlow and Miner (1969) in a study of the temporal reliability of the length-complexity index (LCI) and mean length of response (MLR) used three sets of stimulus pictures which they had compiled. Although they reported the means of these language measures for the three elicitation times, they randomized the order of presentation of the three sets of pictures. Because of this randomization no comparison is possible of the differential stimulus value of these sets.

Cowan, Weber, Hoddinott, and Klein (1967) studied the effects of different experimenters and stimulus pictures upon the mean length of responses of language samples elicited from ninety-six normal children. These children comprised four subgroups of twenty-four children each for ages five, seven, nine, and eleven. The subgroups were equalized for sex and socio-economic status. The stimulus materials were ten pictures "taken from a popular magazine's cover paintings". After a warm-up period each child was shown all ten pictures, presented in random order with the instructions, "Tell me what you see in the picture", "Tell me what is happening", or "Tell me what the people are doing". Two-thirds of each subgroup were seen by one experimenter and one-third by the other. Both experimenters were experienced speech therapists. Results indicated that two pictures were significantly different from the rest, one of these eliciting a shorter average MLR, and the other a longer average MLR. The difference in the mean MLR for these two pictures was 4.49 words. There was also a significant

interaction between the variables of age, sex, and experimenter. The first experimenter elicited longer average MLR's for seven and nine year olds, while the second experimenter elicited longer MLR's from five and eleven year olds. In averaging MLR scores across all ages, boys gave longer MLR's with the first experimenter, while girls gave longer responses with the second.

Mintun (1968) elicited language samples from thirty educable mentally retarded children ranging in age from six years ten months to ten years four months. Her specific purpose was to determine whether three different stimulus situations elicited significantly different language samples as evaluated by MLR, total number of words (TNW), number of different words (NDW), and LCI. Her stimulus conditions were toys, single object color photographs of these same toys, and twenty second single concept color films of the actual items which the toys represented. For example, a movie was made of a real horse. Subjects were randomly assigned to one of these three conditions. Following a training session with the first three stimulus items, the other nine stimuli were presented in a prescribed order to each subject. Instructions were "Tell me all you can about this". Further encouragement was given to subjects by remarks such as "Can you tell me anything else", "That was a good story", and "You're doing fine". A significant difference in LCI scores was found between photographs and films, the photographs yielding lower LCI scores. The LCI score for toys was higher than, but not significantly different from, the LCI score for photographs. No significant difference was found in the MLR scores elicited in the three conditions. The scores for TNW and NDW were significantly different and larger in the toys condition.

Strandberg and Griffith (1969) utilized three stimulus conditions for their study of the effects of training in visual literacy on the expressive language of twelve preschool upper socio-economic status children. These children were divided into stimulus and control groups. The stimulus media were three sets of pictures which the children took themselves with an Instamatic camera. The first set was of individual toys, the second and third were taken of anything in the child's home environment. Between the shooting of the second and third set of pictures the experimental group received training in visual sequencing. Each child was presented with his own sets of pictures with the instruction, "Tell me about your pictures". Results showed significant differences in LCI scores for all three language samples of the control group, the lowest score obtained was in response to pictures of toys and the highest to the last set of pictures taken at home. The experimental group produced an increase in LCI from the first to the second set, and from the second to the third set. The difference in LCI from the second to the third set was significant.

Strandberg (1969) duplicated Mintun's study using thirty normal children ranging in age from four years zero months to five years eleven months. Her results showed no significant differences in LCI or MLR scores for language samples elicited by the three stimulus media. Both toys and films yielded higher and significantly different scores for the measures of TNW and NDW.

Ahmed (1973) elicited language samples from thirty-two mentally retarded residents of a state institution, whose ages ranged from ten to eighteen years. Their average age was fourteen years four months. Subjects were divided into two groups, each group containing an equal

number of trainable level and educable level children. The stimulus media she compared were color slides of single object and multi-object pictures taken from the Peabody Language Development Kit (Dunn and Smith, 1966). She found that multi-object pictures produced higher and significantly different LCI, Carroll type-token ratio, and TNW scores for both trainable and educable children. There was no significant difference in the MLR scores elicited by single object and multi-object pictures.

**THIS BOOK
CONTAINS
NUMEROUS PAGES
THAT WERE
BOUND WITHOUT
PAGE NUMBERS.**

**THIS IS AS
RECEIVED FROM
CUSTOMER.**

APPENDIX B

| <u>Subject</u> | <u>Age</u> | <u>IQ</u> |
|----------------|------------|-----------|
| S-1 Jesse | 10-1 | 77 |
| S-2 Mary | 10-5 | 78 |
| S-3 John | 9-10 | 81 |
| S-4 John | 9-9 | 92 |
| S-5 Tim | 11-0 | 82 |
| S-6 Dwayne | 11-2 | 86 |
| S-7 Ronald | 11-1 | 71 |
| S-8 Carl | 11-0 | 80 |
| S-9 Jim | 14-7 | 76 |
| S-10 Tony | 14-2 | 73 |
| S-11 Kim | 14-1 | 81 |
| S-12 Penny | 14-3 | 71 |
| S-13 Joyce | 10-2 | 49 |
| S-14 Crystal | 10-3 | 43 |
| S-15 Larry | 9-10 | 43 |
| S-16 Larry | 10-3 | 41 |
| S-17 Ricky | 11-0 | 52 |
| S-18 Nancy | 12-2 | 36 |
| S-19 Donald | 12-0 | 49 |
| S-20 Lyle | 12-5 | 48 |
| S-21 Carol | 14-4 | 46 |
| S-22 Mary | 13-3 | 45 |
| S-23 David | 13-8 | 47 |
| S-24 Lee | 13-6 | 45 |

APPENDIX C

Peabody "I Wonder" Series

- W1 - three children in a petshop with various animals in cages
- W2 - street scene with a fruit truck, a dog, and a man running down the street
- W3 - an airport
- W4 - futuristic scene, three children in a space car
- W5 - two deep sea divers, a porpoise, and a treasure chest
- W6 - a hospital emergency room, two children, nurse, doctor
- W9 - boy in bed in his room, broken window, footprints on the wall
- W10- boy in a space rocket, castle in the background
- W11- rodeo, boy riding a bull, girl roping a calf
- W12- flying saucer, two martians

APPENDIX D

Constructions and compound words which were arbitrarily counted as one word for the TNW measure.

| | |
|--------------------------|------------------|
| ahold | grade school |
| baldheaded | grape sour |
| ballgame | Henny Penny |
| ballpoint | hi diddle diddle |
| barbie doll | high school |
| baseball | home run |
| basketball | hot dog |
| bat boy | hot rod |
| birthday cake | ice cream |
| boo-boo | iced tea |
| boy scout | ink pen |
| boy wonder | Johnny Cycle |
| brownie scout | K-State |
| bubble gum | K.U. |
| Bugs Bunny | let's |
| candy bar | let's see |
| cat woman | living room |
| chicken pox | lookit |
| cornbread | merry-go-round |
| cottage cheese | Mickey Mouse |
| cowboy | nighttime |
| cracker jacks | oh no |
| cub scout | oh wow |
| Daffy Duck | oh yes |
| dining room | OK |
| diving board | one hundred |
| dress-up | "other where" |
| Ducky Lucky | panty hose |
| everyday (adjective) | peanut butter |
| FBI | pet farm |
| fire escape | pet shop |
| first name + middle name | play room |
| fish bowl | policeman |
| fish hook | Porky Pig |
| fish pole | potato chips |
| football | pussy cat |
| forty-seven | ray gun |
| french fry | record player |
| french toast | rifleman |
| giftwrap | road runner |
| Goosey Loosey | school bus |

school teacher
 seven-thirty
 shoestring
 shush-up
 shut-up
 silverware
 skin diver
 smoke stack
 smoke stainer
 softball
 spacecar
 space gun
 spacehouse
 spacejet
 spaceman
 space rider
 space rocket

space ship
 space stick
 space suit
 steamboat
 St. Mary's
 stoplight
 Sunday school
 swimming pool
 tape recorder
 tic tac toe
 toenail
 toothpaste
 tootsie roll
 T.V.
 volley ball
 you know (when uttered as a phrase)

Constructions and compound words which were arbitrarily
 counted as two words for the TNW measure.

ain't
 all right
 a lot
 bear cub
 bunny rabbit
 bus ride
 bus stop
 Captain Kangaroo
 Charlie Brown
 Cookie Man
 Cookie Monster
 craft gun
 Dr. Knickerbocker
 Elmer Fudd
 every day (adjective + noun)
 every night
 every time
 filling station
 first name + last name
 fried chicken
 flying saucer
 gonna
 gotta
 green bean
 grocery store
 hadda
 Hawaii Five-O
 hurry up
 jack rabbit
 Jackson Five
 Kansas City

kinda
 kitty cat
 mashed potatoes
 meat loaf
 Mimosa B
 Miss Smith
 Mod Squad
 orange juice
 orbit gun
 O.T.
 outer space
 Partridge Family
 plane station
 police car
 pork chops
 post office
 rec room
 railroad track
 rocket car
 rocket fire
 Sesame Street
 show-off
 South Dakota
 space deal
 Split Second (name of T.V. show)
 telephone pole
 tha'd
 traffic light
 treasure chest
 T.V. room
 Uncle Donny

wanna
watch

what's

Constructions and compound words which were arbitrarily
counted as three words for the TNW measure.

cream of wheat
Mr. Green Jeans

talking tender love
whatcha

REFERENCES

- Ahmed, S. E. Linguistic analysis of children's speech: Effects of stimulus media on elicited samples. Unpublished master's report, Kansas State University, 1973.
- Barlow, M. C. and Miner, L. E. Temporal reliability of length-complexity index. Journal of Communication Disorders, 1969, 2, 241-251.
- Bellak, L. and Bellak, S. S. An introductory note on the Children's Apperception Test (CAT). Journal of Projective Techniques, 1950, 14, 173-180.
- Blum, G. S. The Blacky Pictures. Psychological Corporation, 1950.
- Carroll, J. B. Language and thought. Englewood Cliffs, N. J.: Prentice-Hall, Inc., 1964.
- Cochran, W. G. and Cox, G. M. Experimental designs. (2nd ed.) New York: John Wiley and Sons, Inc., 1957, 117.
- Cowan, P. A., Weber, J., Hoddinott, B. A., and Klein, J. Mean length of spoken response as a function of stimulus, experimenter, and subject. Child Development, 1967, 3, 191-203.
- Darley, F. L. and Moll, K. L. Reliability of language measures and size of language sample. Journal of Speech and Hearing Research, 1960, 3, 166-173.
- Davis, E. A. The development of linguistic skill in twins, singletons with siblings, and only children from age five to ten years. Child Welfare Monographs, No. 14, Minneapolis: University of Minnesota Press, 1937.
- Day, E. J. The development of language in twins I. A comparison of twins and single children. Child Development, 1932, 3, 179-199.
- Dunn, L. M. and Smith, J. O. (Eds.), Peabody language development kit. Circle Pines, Minn.: American Guidance Service, 1966.
- Fryer, H. C. Concepts and methods of experimental statistics. Boston: Allyn and Bacon, 1966.

- Griffith, J. Comment on "A standardized method for obtaining a spoken language sample". Journal of Speech and Hearing Research, 1970, 13, 438-439.
- Griffith, J. and Miner, L. E. LCI reliability and size of language sample. Journal of Communication Disorders, 1969, 2, 264-267.
- Heber, R. F. A manual on terminology and classification in mental retardation. (Monogr. Suppl.) American Journal of Mental Deficiency, 1959, 64, 1-17.
- Johnson, W. Studies in language behavior. I. A program of research. Psychological Monographs, 1944, 56, 1-15.
- Lemon, J. and Fritz, J. Sample size reliability of MLR for mentally retarded children. Unpublished manuscript. Kansas State University, 1973.
- Longhurst, T. M. and Siegel, G. M. Effects of communication failure on speaker and listener behavior. Journal of Speech and Hearing Research, 1973, 16, 128-140.
- McCarthy, D. The language development of the preschool child. Child Welfare Monographs, No. 4, Minneapolis: University of Minnesota Press, 1930.
- Menyuk, P. Syntactic structures in the language of children. Child Development, 1963, 34, 407-422.
- Miner, L. E. Scoring procedures for the length-complexity index: A preliminary report. Journal of Communication Disorders, 1969, 2, 224-240.
- Minifie, F. D., Darley, F. L., and Sherman, D. Temporal reliability of seven language measures. Journal of Speech and Hearing Research, 1963, 6, 139-148.
- Mintun, S. An evaluation of three stimulus media for eliciting verbal language samples from educable mentally handicapped children. Unpublished master's thesis. Eastern Illinois University, 1968.
- Myklebust, H. Development and disorders of written language. Vol. I., Picture Story Language Test. New York: Grune and Straton, 1965.
- Sharf, D. J. Some relationships between measures of early language development. Journal of Speech and Hearing Disorders, 1972, 37, 64-74.
- Siegel, G. M. Appendix I: Prototypes of criteria for counting words. Monograph 10. Journal of Speech and Hearing Disorders, 1963, 10, 103.

- Siegel, G. M. and Harkins, J. P. Verbal behavior of adults in two conditions with institutionalized retarded children. Monograph 10. Journal of Speech and Hearing Disorders, 1963, 10, 39-46.
- Strandberg, T. E. An evaluation of three stimulus media for evoking verbalization from preschool children. Unpublished master's thesis. Eastern Illinois University, 1969.
- Strandberg, T. E. and Griffith, J. Study of the effects of training in visual literacy on verbal language behavior. Journal of Communication Disorders, 1969.
- Templin, M. C. Certain language skills in children: Their development and interrelationships. Child Welfare Monographs, No. 26, Minneapolis: University of Minnesota Press, 1957.
- Wilson, M. E. A standardized method for obtaining a spoken language sample. Journal of Speech and Hearing Research, 1969, 12, 95-102.

A COMPARISON OF CHILDREN'S LANGUAGE SAMPLES
COLLECTED IN FOUR ELICITATION PROCEDURES

by

Susan Grubb

B. A., University of Arizona, Tucson, Arizona 1971

AN ABSTRACT OF A MASTER'S THESIS

submitted in partial fulfillment of the

requirements for the degree

MASTER OF ARTS

Department of Speech

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1973

ABSTRACT

One of the most comprehensive means of evaluating the language of a child is by the analysis of a sample of his expressive oral language. Various linguistic measures can be applied to this sample to yield measurements of the quantity and complexity of the child's language. It is possible that these measurable characteristics of the child's language may be modified by the situation in which he is induced to speak. The results of various researches of child language are difficult to compare if different elicitation situations have been used to obtain language samples. This points out the need for a standardized elicitation procedure, which must also be demonstrated as an efficient procedure for collecting a sample of the child's best oral language.

This study compared language samples elicited in four different Treatments; peer conversation, adult-child conversation, picture description, and object description. Twenty-four children residing in a state institution for the mentally retarded participated as subjects. They were divided into two intelligence Levels, Level I with a mean IQ of 79, and Level III with a mean IQ of 45. They were also divided into three Age groups with means of 10-2, 11-3, and 14-0 years. Each child responded in all four Treatments in a counter-balanced Order. Language samples were analyzed by the measures of total number of words (TNW), mean length of utterance (MLU), Carroll type-token ratio

(CTTR), and length-complexity index (LCI).

Mean scores from each Treatment for these four language measures were subjected to an analysis of variance. Whenever the F ratio was significant at the .05 level or better a least significant difference (LSD) procedure was used to compare means at the .05 level.

Results of these analyses were that:

1. The average performance of Level I subjects was higher than the average performance of Level III subjects regardless of the linguistic measure used to evaluate performance and regardless of the Treatment in which the language sample was collected.

2. Combining scores across Levels, a significant Age group difference was found only in the CTTR scores, where performance improved from 10 to 11 years, and then fell slightly at 14 years. The difference in the mean for the 10 and 11 year olds was significant at the .05 level.

3. As indicated by the scores of the four linguistic measures there were significant differences in the language samples collected in the four Treatments. Peer conversation and adult-child conversation elicited relatively higher scores than picture and object description did. When scores of all subjects were combined it was found that the LCI and MLU were least influenced by a change in Treatment, and that CTTR and TNW were most influenced by changes in Treatment. LCI and MLU appeared to be more stable measurements of a child's verbal ability.

Of the Treatments compared in this study, adult-child conversation was recommended as the most efficient in collecting a sample of the child's best verbal performance. Although peer conversation also

produced high scores it is a more difficult Treatment to apply since it requires a peer, and because it necessitates excellent recording conditions. It was suggested that future research would be helpful to determine the stability of a child's language in conversation with different examiners.