

A COMPARISON OF THREE LANGUAGE COMPREHENSION TESTS

by

MARY LOU EMERY

B. A., Kansas State University, 1973

A MASTER'S THESIS

submitted in partial fulfillment of the

requirements for the degree

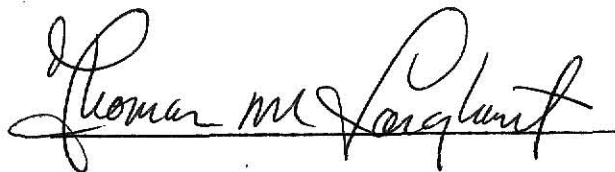
MASTER OF ARTS

Department of Speech

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1975

Approved by:

A handwritten signature in cursive script, reading "Thomas M. Langford", written over a horizontal line.

LD
2668
T4
1975
E46
C.2
Document

TABLE OF CONTENTS

	Page
LIST OF TABLES	iii
INTRODUCTION	1
METHOD	3
Subjects	
Experimental Facility	
Procedure	
Comparison Procedure	
RESULTS AND DISCUSSION	6
<u>Test for Auditory Comprehension of Language</u>	
<u>Miller-Yoder Test of Grammatical Comprehension</u>	
<u>Northwestern Syntax Screening Test</u>	
Comparison of linguistic structure in the three tests	
Chi-square analysis	
Analysis of variance comparing language ages and percentiles	
REFERENCES	20
APPENDIX	21

LIST OF TABLES

Table	Page
1. Chronological age, mean length of utterance, and <u>Peabody Picture Vocabulary Test</u> language age and vocabulary intelligence quotients for the eight subjects	4
2. Linguistic features across three language tests according to test question	16
3. Chi-square analysis of the proportion of correct response on the three tests by linguistic feature	18
4. Subject scores and means for language age and percentile scores on the Carrow and NSST	19

ACKNOWLEDGEMENTS

I wish to express my appreciation to Dr. Thomas M. Longhurst for his guidance and patience. I also wish to thank the members of my committee: Dr. Bruce C. Flanagan, Dr. James Armagost, and Dr. Norma D. Bunton.

For statistical advice, I am grateful for the assistance of Dr. Arthur Dayton of the KSU Statistics Department.

For providing moral support and sympathy, I wish to thank Megan King.

And since she dedicated her book to me, I would like to dedicate this thesis to my mother, Janet Pease Emery.

Evidence suggests that the acquisition of language skills, both comprehension and production, follows a relatively set sequence (Brown, 1973). This knowledge has made it possible to construct tests that probe the child's linguistic rules at various stages of development. Most of the formal tests that have been constructed have attempted to assess the comprehension of linguistic features by children. These comprehension tests are often used as predictors of later productive language skill development. This is reasonable since it is a widely held belief among developmental psycholinguists that passive control (comprehension) precedes active control (production) of the same linguistic features (McNeill, 1966).

A test designed to probe comprehension skill should: (a) require minimal or no productive language from the child and (b) accurately reflect the child's specific comprehension strengths and weaknesses. This second requirement of an effective comprehension test is probably the most difficult to assure. If a child misses an item, it is virtually impossible to know whether the child lacks the knowledge to respond correctly, or whether he is distracted by a drawing or by a poorly constructed stimulus item. If, on the other hand, the child passes an item, the examiner may be fairly certain that the child has correctly comprehended (Carrow, 1968).

There are now several tests on the market which purport to accurately test language comprehension skills, and yet they vary greatly among themselves in terms of structures tested, presentation of items, scoring, and interpretation. The discerning speech and language clinician should evaluate each of these areas before using a test as a diagnostic tool. Unfortunately, many clinicians have neither the time nor the test population

to make such judgments. There is, therefore, a need to provide speech and language clinicians with a critical review of the tests which are commonly used to assess receptive language skills.

The purpose of this study was to evaluate and compare three well-known comprehension tests, specifically, the Test of Auditory Comprehension of Language (Carrow, 1968, 1973); the Miller-Yoder Test of Grammatical Comprehension, Experimental Edition, (Miller & Yoder, 1975); and the Northwestern Syntax Screening Test (Lee, 1969).

The tests were compared on the following criteria: test instructions; stimulus pictures and arrangement of pictures; scoring and interpretation; ease of administration; use as a diagnostic tool; normative data; and reliability and validity.

Carrow stated that the two functions of the Test for Auditory Comprehension of Language were to measure the auditory comprehension of structure, and to allow the examiner to determine the areas of linguistic difficulty. The test consists of a set of 101 plates each containing from one to three line drawings designed to represent vocabulary items, grammatical categories, and morphological and syntactic structures. A correct response consists of the child correctly pointing to a word, phrase, or sentence presented by the examiner. Each item is presented in a specific manner, and tag phrases such as "show me" are not used unless specifically indicated. Each correct response counts as one point and the total score is the total number of items passed. The raw score is divided into sub-scores for broad category scores. Percentile scores and language ages can be derived for comparison with norms ranging from 3-0 to 6-11 years of age.

The Miller-Yoder Test of Grammatical Comprehension (MY) was intended to measure children's grammatical comprehension. The test consists of 42 sentence pairs to be identified by pointing in response to stimulus sentences. There are 42 plates, each plate containing 4 line drawings, 2 foils and 2 pictures identifiable from test sentences. Responses are recorded on the data sheet, which are then transferred to a Comprehension Test Results Form which indicates the categories that were missed. No raw score is obtained, and thus no percentile score or language age may be derived.

The Northwestern Syntax Screening Test (NSST) by Lee was intended to test not only a child's comprehension of grammatical features, but also his production of them and to compare the two. For this reason, the NSST consists of 40 sentence pairs with 20 to be identified receptively and 20 to be repeated productively. Since for the purposes of the current study the expressive portion of the test is not appropriate, discussion will be limited to the receptive section. It consists of 20 plates, each of which contains 4 pictures (again 2 foils and 2 pictures identifiable from test sentences). A correct response consists of the child pointing to the appropriate pictures one at a time as the examiner repeats the two sentences. Each of the test pairs yield perfect scores of 2, scores of 1 if one sentence is missed, and 0 if both are missed. Norms range from 3-0 to 7-11 years of age, and are expressed on charts from which a child's percentile score may be determined.

METHOD

Subjects

The subjects were selected from the children in one class of a university child development laboratory school. Four children were eliminated

from consideration because of bilingualism. The children's teachers rated the language abilities of the remaining twelve on the basis of a Guide for Teacher Ratings (Appendix A). Children in Group I were the four rated highest in language abilities while Group II children were the four rated lowest. A middle group of four children was excluded from the study. Teacher ratings were compared to results from the Peabody Picture Vocabulary Test, Form A (Dunn, 1959) and by mean length of utterance (MLU) computed from a 25-utterance sample. Descriptive information for the subjects is presented in Table 1.

Table 1
Chronological Age, Mean Length of Utterance, and Peabody
Picture Vocabulary Test Language Age and Vocabulary
Intelligence Quotients for the Eight Subjects

<u>Child</u>	<u>Chronological Age</u>	<u>Mean Length of Utterance</u>	<u>Language Age</u>	<u>Vocabulary I.Q.</u>
A	4-2	5.59	6-10	126
B	5-1	5.52	6-10	118
C	4-0	5.53	6-3	122
D	4-2	5.29	6-1	120
E	3-9	4.00	3-10	97
F	5-0	4.71	5-2	101
G	4-4	6.29	5-5	103
H	3-6	4.14	3-9	105

Children A-D comprised the high group for this study; children E-H comprised the low group.

The mean vocabulary intelligence quotient (VIQ) of Group I was 121.05 with a range of 118 to 126 and the mean VIQ of Group II was 101.50 with a range of 97 to 105. The chronological ages of the subjects ranged from 3-6 to 5-1, with a mean age of 4-11. The age range of Group I was 4-0 to 5-1 with a mean age of 4-4, and the age range of Group II was 3-6 to 5-0 with a mean age of 4-1. Analysis of variance at the .05 level of confidence revealed no significant difference between chronological ages of the two groups. Sex was not controlled as a main variable, and three of the eight subjects were female.

Experimental Facility

The three language tests were administered in the testing room of the child development laboratory school. The room contained two chairs and a table, and was free of distracting stimuli.

Procedure

Each subject was brought individually to the experimental room and seated at the table. After gaining rapport with the subject, the experimenter gave instructions for the particular test according to the test directions. Reinforcement in the form of praise was given only for attending to the task unless instructions specified otherwise.

The order of test administration to each child was randomized to minimize the statistical possibility of one test influencing the results of another. Each child was seen once a day for three consecutive sessions. No one session lasted more than 45 minutes. Each of the tests was administered a total of eight times, and only after the examiner had practiced enough to feel confident in the procedure.

Comparison Procedure

Although it was impossible to make irrefutable criticisms of the Carrow, the MY, and the NSST, it was possible to discuss difficulties which may be encountered by a clinician administering one of these tests. The comparisons were based on subjective judgments made at the time of test administration by one examiner, and later with two other examiners who had experience administering the tests. The tests were compared on the following criteria: test instructions; stimulus pictures and arrangement of pictures; scoring and interpretation; ease of administration; use as a diagnostic tool; normative data; and reliability and validity.

A chi-square analysis was conducted to compare the proportion of correct responses of the children on individual grammatical structures on each of the three tests.

RESULTS AND DISCUSSION

Test for Auditory Comprehension of Language (Carrow)

Carrow's test seemed to be the easiest of the three to administer, although this does not necessarily indicate any superiority in test quality. The directions in the test manual for administering the test were concise and clear. The examiner was instructed to repeat the stimulus items exactly as they appeared in the test booklet, and to give the stimulus only once. The examiner was instructed to refrain from making remarks or using facial expressions to encourage or discourage the child. This part of the instructions resulted in a tendency to slow down testing process, and to make it necessary for the examiner to take several breaks in testing in order to refocus the child's attention on the test pictures.

Had the test allowed for some reinforcement, the children may have found the test more enjoyable.

Each of the 101 test plates of the Carrow contained three line drawings--one representing a test item read by the examiner, and two foils. For the most part, the pictures were clear and interesting, and the foils were not distracting. However, several of the foil drawings were so distracting for the younger children that the child concentrated on the foil ignoring the other two less interesting pictures. For example, item 60 was designed to probe knowledge of the pronoun "he" and consisted of three pictures of a girl, a boy, and a monkey. Nearly all of the children were distracted by the monkey, and two children actually appeared to fail the item because they insisted upon pointing to and talking about the drawing of the monkey. A similar problem was noted with item 61, which tested "she." The pictures were of a single boy, a single girl, and a girl and boy together, all dressed in the same type of clothing. Several children pointed to the girl that was with the boy instead of the girl presented alone. Technically, the child missed the item, yet he demonstrated knowledge of the structure.

Another criticism of a specific test item concerned item 57 which probed the derivational suffix "ist" in the word "bicyclist." The three test items were one bicycle, several bicycles, and a bicycle rider. All eight children that were tested pointed to "several bicycles." One explanation of this was that none of the eight children possessed this linguistic feature. Another explanation, based upon the fact that all eight children pointed to the same plate, was that the test items which preceded item 57 probed the concepts "small," "taller," and "fattest."

The inflections "est" and "ist" sound alike, and the concept of superlatives was evidently carried through the next item. Placing "ist" in another test location might alleviate this problem.

The arrangement of the stimulus pictures was by grammatical category, not by level of difficulty which varied throughout the test. This factor tended to assure at least a minimal amount of reinforcement for the child throughout the test.

The score sheet consisted of a face page, two pages of stimulus sentences, and a one-page analysis section. On the response section, the examiner wrote the child's response (1, 2, or 3) in the appropriate column. It was unnecessary to score an item during administration. After the responses had been recorded and scored, the examiner transferred the items to the analysis section in one of the three subscores which included vocabulary subscore, morphology subscore, and syntax subscore. This simple procedure was intended by Carrow to provide an indication of the areas of difficulty, and also to serve as a method for judging maturation or change in previous performance. The total raw score was then converted to an age score which was compared to mean and median performances of each age group. Finally, a percentile ranking was computed through the use of tables contained in the test manual.

The normative data, test reliability and validity included with the test were more complete for the Carrow than the other two tests. The norms were based on 200 middle class Black, Mexican-American, and Anglo children ranging in age from three through six. There were 50 children in each age group.

The validity of the Carrow was demonstrated in three ways. First, increase in age was paralleled by an increase in score indicating that the test followed the same developmental trends that language follows in children. Second, the test was shown to distinguish between children who have known language disorders and those who do not. Third, the Carrow demonstrated reliable test-retest results in language disordered children.

There were several analyses cited which provided reliability data. The first data were obtained from a study which included 51 children, half of whom were Mexican-American, who all received the English version of the test. In addition, 32 Mexican-Americans were administered the Spanish version. All children were retested after a two-week interval. All of the reliability measures were significantly different from zero. There was high (.94 for the English, .93 for the Spanish) test-retest reliability for total scores.

Another study by Marquardt and Saxman (1972) retested five subjects with poor articulation and five with good articulation four weeks after initial testing. The rank order correlation between error scores for the first and second test administrations for 10 subjects was .92.

A study using the Kuder-Richard Reliability Test to measure the consistency of performance on items within the test was conducted with middle and lower class Black and Anglo children. The total test reliability was .77.

In a study measuring performance of dysphasic children tested at one year intervals, Weiner (1972) showed that the Carrow had high predictive ability. A statistically significant rank correlation was obtained ($r=.79$; $p=.02$).

A pre- and post-test analysis for both versions of the Carrow included data from various ethnic groups. Results for the English version showed high pre- and post-test reliability within a group of 495 children.

The Miller-Yoder Test of Grammatical Comprehension (MY)

This test differed in format and in presentation from the Carrow. For example, the instructions to the examiner tended to be less rigid. The examiner was instructed to proceed with two practice plates in order to assure that the child understood the task. Then the examiner proceeded through each of the 42 plates reading the first in a pair of test sentences. After completing this part of the test, the examiner returned to the first test plate, and read the second half of the sentence pairs. The examiner was encouraged to reinforce the child for attending and responding.

The stimulus pictures were four line drawings depicting two test sentences and two foils. Unlike the Carrow, the test items were paired sentences, usually comparing structures such as is/is not and he/she. The plates themselves were not bound and it was necessary to rearrange the plates before administering the test. Since the plates were not used in numerical order, it was difficult to quickly locate a specific plate.

The drawings had a simple child-like quality which the children seemed to enjoy. However, some of the drawings were cluttered in appearance and seemed to distract some of the younger children. Perhaps a greater problem was that no matter how lively and interesting the pictures were, none of the children tested enjoyed going through all 42 plates twice. In fact, one of the younger children left the room when the stack of test plates was turned over to begin the second half.

Recording the child's responses entailed simply writing down the number of the "box" to which the child pointed. The responses were then transferred to the Comprehension Test Results Form which merely listed each of these items in numerical order. This part of the MY was awkward and time consuming. There did not seem to be a simple way to transfer the results to this form. The Comprehension Test Results Form was compared to the Sentence Description Form which provided the sentences for each of the foil pictures. There were no normative data included, and no provisions for ranking a child according to language age or percentile. The examiner also used a chart included with the test (Miller & Yoder, 1973) which gave the number of children who passed each item on the test at ages three through six.

Since the MY did not provide a means for easy comparison of one child to another, and since it was a long test, it should not be used as a screening tool. It did provide some useful information about the children's receptive language skills, which was effectively presented by the analysis sections and the table.

At the time the MY was published, a study had just been completed by N. O. Owings which used 120 subjects ranging in age from three to six years of age in order to investigate several aspects of the MY. These were internal reliability, item analysis, item difficulty index, item criterion correlation, point of maximum discrimination on the criterion scale for each item, and the discrimination power of each item. Also, tentative age norms were established and compared with Carrow's 1968 comprehension data. These norms revealed that children comprehend grammatical forms later on the MY than the Carrow. The results of the item analysis showed the MY was highly reliable (.93 Hoyt Reliability Index).

The Northwestern Syntax Screening Test (NSST)

The NSST differed from the Carrow and the MY in several regards. First, Lee did not claim that the NSST possessed the diagnostic capabilities that were claimed by the authors of the other two tests. It was designed as a screening test. Additionally, the test was divided into a receptive section and an expressive section. However, the NSST's receptive section was similar in construction and administration to the other two tests.

There were no major difficulties in understanding or carrying out the test instructions. The examiner was instructed to first familiarize the child with a demonstration plate to assure he understood the task. Then the examiner administered the 20-item test. Each item was composed of a pair of sentences, as was the MY. Instead of reading one sentence of the pair and then returning to the beginning to read the second sentence of the pair, the examiner read and scored both sentences in one administration of the test plates. For this reason, and because the receptive section consisted of only 20 items, the NSST took only 5 to 10 minutes to administer.

One criticism of the NSST concerned the stimulus pictures. Like the MY, there were four line drawings on each plate. The drawings appeared to be adequate and uncluttered, but most of the children seemed to be bored with them. Although this test required the least amount of time to administer, several children expressed a desire to stop. Perhaps this was a fault of the drawings themselves, or of the sentences, most of which did not seem terribly interesting. Most of the drawings clearly depicted the sentences for which they were intended, although two plates intended

to probe interrogatives proved to be confusing for the children. Items 14 and 17 contained the sentences, "Is the dog in the box?" and "Has Daddy finished dinner?" Instead of pointing to the appropriate drawing, several children answered, "Yes."

The NSST was scored by assigning one point to each correct sentence making a total of 2 points possible for each test item, and 40 total points possible for a receptive raw score. The score sheet provided a space for recording the points. The test instructions did not tell the examiner to mark which sentence the child missed in the pair which should be more relevant information than the points earned.

From the raw score, the examiner simply computed a percentile score from the table included in the test booklet. This was the only measure provided.

There were no test validity or reliability data included in the test booklet. The normative data used to derive percentile scores was based on testing 242 children between the ages of 3-0 and 7-11 (Lee, 1970). The children were from middle or upper income communities, and from homes where a standard American dialect was spoken. These data were collected in 1969.

Although the number of subjects was 242, which appeared to be sufficiently large on which to do statistical analyses, it should be noted that the norms were age specific. In other words, although there were 242 total children, each age group ranged in number from as low as 13 to as high as 62. This means that for the age range 3-0 to 3-5, the normative data were based on the test results of only 13 children. Because the normative data were based on only white middle-class children, and because they were based on a small number of trials, the percentile scores are questionable.

The NSST was easy to administer, yet it would not be particularly useful beyond screening and would not be very helpful for therapy planning. As Lee stated, it would be applicable for screening large numbers of children. And yet, as will be shown later, one child who received high scores on the PPVT and the Carrow scored below the tenth percentile on the NSST.

Comparison of Linguistic Structure in the Three Tests

Table 2 compares the linguistic features across the three language tests according to test item. This table attempted to include all of the linguistic features found in the MY, the Carrow, and the NSST, and to indicate which of the tests probed each feature. Certain observations are evident in examining this chart. For instance, the MY did not test any constructions using impersonal pronouns; the NSST did not test the use of adjectives; and the Carrow did not include a probe for object pronouns. Although all three tests included items which were not included in the other two, the Carrow proved to be especially complete. For example, it included a lexicon of words which appeared later in the test to assure that a child did not fail an item due to lack of vocabulary.

Chi-square Analysis

A summary of a chi-square analysis comparing the proportions of correct responses on the three tests by linguistic feature is presented in Table 3. The chi-square analysis revealed that there were significant differences in the proportion of correct answers among twenty of the twenty-six linguistic features common to all three tests. Theoretically, if the child possessed a certain linguistic feature, he should be able to correctly answer any of the equivalent test items. These results indicate

that children did not correctly answer comparable items from these tests. In fact, out of the 26 linguistic features, only 6 showed similar results. Thus, it is difficult to determine whether or not a child truly possessed a linguistic feature and whether or not a specific test is valid.

Analysis of Variance Comparing Language Ages and Percentiles

The language age in months and percentiles for the Carrow and the NSST are presented in Table 4. Analyses of variance were used to compare the language age means and the percentile means for the Carrow and the NSST. The MY was not included in this statistical analysis because the test did not provide language age or percentile scores. The results of the analysis of variance indicated significant differences between the Carrow and the NSST on language age at the .10 level of significance and on percentile scores at the .05 level of significance. This result again brings into question the validity of the tests.

Table 2
Linguistic Features Across Three Language Tests
According to Test Question

Linguistic Feature	MY	Carrow	NSST
Prepositions			
under	3, 4	45	1
on	3, 4	44	3
in	5, 6	46	3
beside	5, 6	-	-
behind	-	-	1
between	-	48	-
at the side of	-	47	-
in front of	-	49	-
Nouns			
Singular	24, 25	69	6
Plural	24, 25	66, 67, 68	6
Tense			
Past	31, 32	74	8
Pres. Prog.	30, 31	64, 72, 73	12
Future	30, 32	76	12
Pres. Perf. Prog.	-	77	-
Pres. Perf.	-	75	-
Noun/Verb Inflections			
is	28, 29	70	5
are	28, 29	71	5
Noun/Verb Agreement			
Singular	26, 27	91, 92	6, 15
Plural	26, 27	-	6, 15
Pronouns			
he	18, 20	60	2
she	19, 20	61	2
they	18, 19	59	-
we	-	65	-
her	15, 16	-	10
them	15, 17	-	-
him	16, 17	-	-
his	-	63	-
their	-	-	10

Table 2 (cont'd)

Linguistic Feature	MY	Carrow	NSST
Active Reversibles			
Subj./Obj.	1, 2	78, 79	9
Passive Reversibles			
Subj./Obj.	39, 40	80, 81	18
Affirm./Negative			
is	13, 14	85	4
is not	13, 14	86, 87, 98	4
has/doesn't have	9, 10	-	-
can/can't	11, 12	-	-
neither/nor	-	88	-
no + noun	-	94	-
Possessive	7, 8	-	11
Modification			
object	33, 34	96	-
subject	35, 36	97	-
obj./subj.	37, 38	-	-
Indirect Object	-	95	20
Reflexives	41, 42	-	7
Impers. & Interrog.			
Pronouns	-	82	13, 16
who	-	84	13
what	-	-	16
where	-	83	-
when	-	-	19
this	-	42	19
that	-	43	-
these	-	-	-
Question Reversal	-	-	14, 17
Lexicon			
nouns	-	1-10	-
color	36, 37, 34	29, 1-10, 30	-
quality	-	11-13	-
quantity	-	14-20	-
adverbs	-	21-28	-
Derivational suff.	-	39-41	-
		50-58	-

Table 3
Chi-square Analysis of the Proportion of Correct
Response on the Three Tests by Linguistic Feature

Linguistic Feature	<u>X²</u>	<u>df</u>	Linguistic Feature	<u>X²</u>	<u>df</u>
under	3.097	3	are	*50.42	3
on	3.097	3	sing. n/v agree.	*14.654	5
in	3.097	3	pl. n/v agree.	*67.098	3
sing. noun	*75.35	3	she	*29.936	3
plural noun	2.534	5	he	*44.462	3
past	*169.26	3	they	4.210	2
pres. prog.	*25.39	5	sub./obj. act. revers.	*3.998	4
future	*324.89	3	obj. mod.	1.947	2
is	*185.55	3	subj. mod.	*6.634	2
sub/obj pass.	*33.992	4	ind. obj.	*48.369	1
neg/aff is	*7.227	3	reflexive	*8.817	2
neg/aff is not	*6.316	5	who	*7.360	2
possessive	*4.44	2	what	*64.309	1

*Significant at the .05 level.

Table 4
Subject Scores and Means for Language Age and Percentile
Scores on the Carrow and NSST

Child	Language Age in Months		Percentile	
	<u>Carrow</u>	<u>NSST</u>	<u>Carrow</u>	<u>NSST</u>
A	71	75	95	92
B	77	36	93	8
C	64	63	79	75
D	82	76	99	90
E	64	36	68	10
F	73	49	78	30
G	78	72	97	77
H	39	43	57	51
Mean	68.50*	56.25*	83.25**	54.13**

*Significantly different at the .10 level.

**Significantly different at the .05 level.

REFERENCES

- Brown, R., A first language--the early stages. Cambridge, Mass.: Harvard University Press, 1973.
- Carrow, E., Test of Auditory Comprehension of Language (Fifth Edition). Austin, Tex.: Learning Concepts, 1973.
- Carrow, Sr. M. A., The development of auditory comprehension of language structure in children. Journal of Speech and Hearing Disorders, 1968, 33.
- Dunn, L. M., Peabody Picture Vocabulary Test. Circle Pines, Minn.: American Guidance Service, 1959.
- Lee, L., A screening test for syntax development. Journal of Speech and Hearing Disorders, 1970, 35, 103-112.
- Lee, L., Northwestern syntax screening test. Chicago, Ill.: Northwestern University, 1969.
- McNeill, D., Developmental psycholinguistics. In F. Smith & G. Miller, The genesis of language. Cambridge, Mass.: MIT Press, 1966.
- Marquardt, T. P. & Saxman, J. H., Language comprehension and auditory discrimination in articulation deficient kindergarten children. Journal of Speech and Hearing Research, 1972, 15, 382-389.
- Miller, J. F. & Yoder, D. E., Assessing the comprehension of grammatical form in mentally retarded children. Unpublished manuscript, University of Wisconsin, 1973.
- Miller, J. F. and Yoder, D. E., Miller-Yoder test of grammatical comprehension (experimental edition), 1975.
- Owings, N. O., Internal reliability and item analysis of the Miller-Yoder test of grammatical comprehension. Unpublished Masters Thesis, University of Wisconsin: Madison, 1972.
- Weiner, P. S., The perceptual level functioning of dysphagic children: a follow-up study. Journal of Speech and Hearing Research, 1972, 15, 423-438.

Appendix A

Guide for Teacher Rankings

The teacher rankings of the most advanced and the least advanced will be made according to the teachers' judgments of the child's receptive and expressive language abilities.

Receptive language involves the ability to follow simple instructions; demonstration of an understanding of color, shape, time, and number; and demonstration of use and/or function of different objects and vocabulary words. Receptive language does not require any verbal response but rather a gesture in the form of pointing, touching, picking up, giving, or manipulating the appropriate object or task material.

However, in viewing expressive skills, the verbal response that the child produces is evaluated whether it is a spontaneous utterance or one made in an interaction with another person. Expressive language includes general talkativeness; sentence structure and length; vocabulary; and use of correct verb tenses, prepositions, and pluralization of nouns.

A COMPARISON OF THREE LANGUAGE COMPREHENSION TESTS

by

MARY LOU EMERY

B. A., Kansas State University, 1973

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

1975

ABSTRACT

The purpose of this study was to evaluate and compare three well-known comprehension tests used by speech and language clinicians. The tests compared were (a) the Test for Auditory Comprehension of Language (Carrow, 1968); (b) the Miller-Yoder (MY) Test of Grammatical Comprehension (Miller & Yoder, 1975); and (c) the Northwestern Syntax Screening Test (NSST) (Lee, 1969). Eight subjects ranging in chronological age from 3-6 to 5-1 were divided into a high group and a low group according to their teachers' judgments of each child's language skills, and according to results of the Peabody Picture Vocabulary Test (Dunn, 1959) and their average length of utterance. Each test was administered to all eight subjects. A chi-square analysis of the proportion of correct responses indicating comprehension of specific linguistic structures revealed significant differences among the three tests on 20 of the 26 linguistic features the tests had in common. Analyses of variance comparing the mean language age and percentile score for the Carrow and the NSST indicated significant differences between these two tests on language age at the .10 level, and on percentile scores at the .05 level of confidence. In addition to test results, an anecdotal critique of the three tests was included based on the following criteria: test instructions; stimulus pictures and arrangement of pictures; scoring and interpretation; ease of administration; use as a diagnostic tool; normative data; and, reliability and validity.