

A COMPARISON OF THREE PRESCHOOL LANGUAGE TESTS

by

MEGAN LUETHA KING

B.S., 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:


Major Professor

ACKNOWLEDGEMENTS

I want to express my appreciation to Dr. Thomas M. Longhurst for his interest, patience, and support in the preparation of this thesis.

I would also like to thank the members of my committee: Dr. Bruce Flanagan, Dr. James Armagost, and Dr. Norma Bunton.

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

A thanks also goes to Larry for his support throughout my graduate study and Mary for listening and encouragement during the preparation of this thesis.

LD
2668
T4
1975
K55
C.2
Document

TABLE OF CONTENTS

Introduction	1
Method	3
Subjects	3
Pretest	4
Experimental Setting	5
Procedures	5
Test Comparison Procedures	6
Results and Discussion	7
Statistical Analysis	7
Test Instructions	8
Test Items	9
Stimulus Materials	11
Response Criteria and Scoring	13
Data Analysis and Diagnostic Information	15
Normative Data	18
Reliability and Validity	19
Summary and Conclusions	21
References	23
Appendices	25

LISTING OF TABLES

Table 1, Pretest Means and Differences Between the Two Groups of Subjects	5
Table 2, Comparison of the Test Items on the Hannah-Gardner, Utah, and Houston	31

INTRODUCTION

There has been a growing interest in recent years in the young preschool child's language acquisition (Brown, 1973). In conjunction with this interest is the increasing interest in the language-deviant or language-delayed child (Berry, 1969; Emerick & Hatten, 1974; Schiefelbusch & Lloyd, 1974). This interest has been heightened by the realization that significant language delay or deviance in preschool children is highly predicative of later academic difficulties. This realization has precipitated a movement toward early identification of these children and initiation of early intervention procedures (Bereiter & Engelmann, 1966; Friedlander, 1970; Lavatelli, 1971; Bricker & Bricker, 1974).

There are three global or general tests that may be used by speech and language clinicians to assess the language abilities of preschool children. These three tests are the Hannah-Gardner Language Screening Test (Hannah & Gardner, 1974), The Houston Test for Language Development, Part II (Crabtree, 1963), and the Utah Test of Language Development (Mecham, Jex, & Jones, 1967).

The Hannah-Gardner was designed to identify children who exhibit deficits which need further attention. The test was also designed as a screening device to be used when more extensive or formal test instruments cannot be validly administered. The skills assessed in the various sections are visual perception, motor development, auditory perception, and conceptual language abilities. The test was designed to

be administered by professionals associated with preschool children. Percentile scores for each section and a total percentile can be derived from the raw scores.

The purpose of the Utah test was to determine a child's receptive and expressive language abilities in the recognition and naming of colors, money, objects, and body parts; auditory memory span for digits and sentences; copying geometric shapes; and drawing a recognizable picture. A language-age-equivalent can be obtained from the computed raw score.

The Houston was designed to provide a language scale derived from the assessment of receptive and expressive language abilities for children between three and six years of age. Test sections include tasks designed to evaluate vocabulary, gesture, body parts, grammatical usage, drawing, and speech and melody patterns. A language age is obtained from the raw score. In addition, the score sheet included with the test provides the examiner with a review of the child's performance on the test.

Appendix A contains a more complete description of the three tests that were analyzed. Appendix B presents a tabular comparison of the common tasks on the three tests.

The purpose of the present study was to administer and evaluate three global language tests commonly used with preschool children as to their construction, administration, and the diagnostic information obtained from the test results. The study was conducted in an attempt to provide speech and language clinicians with a critique of the strengths and weaknesses of these three tests. With this review

examiners may be able to determine which of these three procedures best fulfills their needs.

METHOD

Subjects

Eight children, seven males and one female, selected from a single preschool class of 16 children served as subjects. All subjects were of middle socioeconomic status, spoke English as their native language, and exhibited no obvious neurological, physical, or mental handicaps. No attempt was made to control for age or sex in subject selection. Subjects ranged in age from 3-8 years to 5-1 years with a mean age of 4-4 years.

Four of the children were rated by their two teachers as the most advanced (high) in their use and understanding of language while the remaining four were rated as the least advanced (low) in their use and understanding of language. These ratings, following a short training session by the experimenter, were made on the basis of receptive and expressive language abilities. These included the child's ability to follow simple instructions; demonstration and understanding of color, shape, time, and number concepts; demonstration of the use of different objects and vocabulary words; general talkativeness; sentence structure and length; and correct use of verb tense, prepositions, and pluralization. Although the teacher ratings were done independently, agreement was 100%.

Pretest

A pretest was administered to each of twelve children, six of whom were rated by their teachers as most advanced in their receptive and expressive language skills and six rated as the least advanced in their receptive and expressive language. Subjects ranged in age from 3-8 years to 5-3 years with a mean age of 4-4 years.

The pretest, comprised of a short elicited language sample and the Peabody Picture Vocabulary Test (PPVT), was administered to each of the 12 children in order to validate the rated differences by the teachers between the two groups of subjects.

The language sample was elicited using "I wonder" pictures found in the Peabody Language Development Kit, Level 2, W-1, W-2, W-3, W-4, W-11, and W-12 (Dunn & Smith, 1967) and recorded on a cassette tape recorder. The mean length of 50 utterances was computed for each sample by counting morphemes within the utterance according to the guidelines presented by Brown (1973).

On the basis of the pretest results and teacher ratings, four children were selected from each group (high versus low) of six children for final testing. The similarity of language skills within groups was the major criterion in the final selection of subjects. The most advanced group ranged in age from 4-3 to 5-1 years with a mean age of 4-5 years while the low language group ranged from 3-8 to 4-6 years with a mean age of 4-1 years. Pretest scores for the subjects are presented in Table 1. The mean chronological ages of the two groups were not statistically significant at the .05 level of confidence.

Table 1
Pretest Means and Differences Between
the Two Groups of Subjects

Source	High-Level Children	Low-Level Children	Difference
<u>Peabody Picture Vocabulary Test</u>			
Raw Score	66 (61 - 69)	38 (19 - 49)	28
Mental Age	8-1 (8-7 - 7-1)	3-10 (2-5 - 4-11)	4-3
Verbal IQ	134 (127 - 138)	93 (67 - 108)	41
<u>Mean Length of Utterance</u>	5.4	3.9	1.5

Experimental Setting

All testing was conducted in a small, quiet room that was part of the children's classroom facility. The room contained several chairs and a small table.

Procedures

The examiner was experienced in administering these three tests.

Each of the three language tests was then administered to each of the eight children. Three children were scheduled for testing each week. All three language tests were administered to a given child during that week. The testing of the three children was arranged with the teachers according to their schedule of morning activities and the child's attendance during that week. Testing was conducted three days a week for a period of three weeks. All tests were administered in the morning with sessions varying in length from 20 to 35 minutes.

Tests were randomly assigned to each child and the order in which each child was tested during the week varied with the availability of the child during the morning.

Each child was taken from the main activity area to the experimental room by the examiner. The child was seated at a small table beside the examiner. The stimulus materials were placed in front of each child in the manner requested by the test manual. When experimental materials were not used for an item they were stored in the test container out of the child's reach. Responses were recorded immediately by the examiner on the test form. The language test was administered in its entirety by the examiner and the child was returned to the group following completion of the test.

All tests were administered by the examiner following the instruction manuals carefully. Responses were recorded on score sheets accompanying each test as directed by the manuals. Final scores were computed as recommended by the authors of each assessment procedure at the termination of each week's testing.

Test Comparison Procedures

As each test was administered, the examiner made notes of subjective judgments about the positive and negative features of the test. Specific notes were also kept of relevant comments about particular test items. A general outline was initially made of the comparison criteria on which the examiner would evaluate each test instrument and these were expanded as the testing and evaluation progressed. Each of the assessment instruments was evaluated on the following criteria: (1) test instructions, (2) test items, (3) stimulus materials, (4) response

criteria and scoring, (5) data analysis and diagnostic information, (6) normative data, and (7) reliability and validity.

In addition to these anecdotal evaluations of the three tests, a statistical comparison of the scores derived from the tests was conducted.

In order to compare the test results of the eight subjects, a common scoring procedure was required. Since the authors of two of the tests, the Utah and Houston, converted their raw scores into language ages, this was selected for the analysis. The Utah expressed the language age in terms of years and months, while the language age for the Houston was expressed in yearly intervals. In order to establish a year and month age for the Houston, the raw score points between yearly levels were divided. Since the Hannah-Gardner computed percentile scores the examiner selected the 50th percentile of the corresponding age on the scale of total scores. The chronological ages increased in five month intervals. These were divided to correspond to the percentile scores.

An analysis of variance was used to compare the language age means of the two groups on the three tests.

RESULTS AND DISCUSSION

Statistical Analysis

The mean language age for the high and low groups when the three tests were combined was 62.75 and 38.33 months respectively. The difference between these means was statistically significant ($F = 30.39$, $df = 1, 6$, $p < .001$). Thus the mean language ages of the high and low groups differed by two years although their chronological ages were not

significantly different. The mean language ages for the high and low groups combined on the Utah, Houston, and Hannah-Gardner were 53.50, 52.89, and 45.25 months respectively and within these means there was a significant difference ($F = 2.68$, $df = 1, 2$, $p < .10$). Application of a least significant difference procedure (Fryer, 1966) at the .10 level revealed that the means for the Utah and Houston were not significantly different but the means for these two tests were significantly different from the mean on the Hannah-Gardner. Thus the Hannah-Gardner produced a significantly lower language age than the other two tests.

Test Instructions

The test questions for the Utah were presented uniformly throughout the test. Each test item included: a short description of the test item, the materials required for the item, the procedure specifying the stimulus to be presented to the child, and the score or response criterion required for the examiner to record a correct response. In addition, the stimulus to be presented to the child was printed in bold type. Due to the organization of specific instructions for each item and the test questions, the Utah was the easiest of the three procedures to administer.

The specific instructions for the test items on the Houston were presented in essentially the same manner as the Utah. The examiner was provided with the procedure and scoring guidelines for each item. However, the Houston made no distinction between instructions to the examiner, the stimulus, and examples of correct and incorrect responses. If the stimulus was identified by bold type or underlining the presentation of test items would be greatly simplified.

Test questions and score sheets for the Hannah-Gardner were contained in a single booklet separate from the general test instructions. The test questions were presented in bold type with specific instructions to the examiner in italics.

Test Items

The Utah was composed of 51 test items divided into age level sections. The Houston consisted of 18 sections with a total of 81 test items. The Hannah-Gardner required responses to 70 items within four major sections.

The test items on the Hannah-Gardner and Houston were arranged according to sections. The tasks relevant to each section were fully assessed before proceeding to the next section. Due to the developmental sequencing of items on the Utah, rather than the sectional arrangement on the Hannah-Gardner and Houston, the children were often required to perform essentially the same tasks repeatedly.

The four sections of the Hannah-Gardner assessed visual perception, motor development, auditory perception, and conceptual development. The visual and auditory perception sections were further divided into memory, figure ground, closure, and sequencing items. The authors did not provide examiners with a standard definition for these terms or a rationale for their inclusion in the test. The conceptual development section evaluated the child's knowledge of numbers, prepositions, colors, and judgment. This particular arrangement of sections provided the children with a variety of activities which aided in maintaining their attention.

The 18 sections of the Houston assessed self-identity, vocabulary, body parts, auditory judgments, counting, repetition of speech and melody patterns, and the drawing of geometric shapes and a picture. Of the 18 sections, four pertained to the analysis of a spontaneous language sample and four to the child's drawing. The Houston was the only one of the three tests administered that attempted to elicit a language sample from the child. In addition, the examiner transcribed the remarks made by the child while drawing the shapes and the picture. This spontaneous sampling of the child's speech was one of the strengths of this test and should be helpful to clinicians who might not otherwise elicit a language sample. This analysis involved only superficial observations of the child's oral language, however. Crabtree did include with the Houston a three page description of the purpose of the test sections and the implications of high and low scores on the sections. These descriptions and statement of implications appeared to be based on some unfounded assumptions concerning language rather than the results of testing and research. One test item on the Houston was particularly successful in eliciting responses from the child. This item assessed the child's receptive knowledge of four prepositions. The examiner covered her eyes while the child placed a miniature farm animal in selected positions in relation to a small doll-sized chair.

Due to the developmental sequencing of the Utah, there was no sectional division of items based on the language skills evaluated. This particular arrangement was not effective in maintaining the children's attention.

Stimulus Materials

Neither the Utah nor Houston relied as heavily on stimulus materials during test administration as the Hannah-Gardner. Of the 70 items on the Hannah-Gardner, 57 or 81% required some type of tangible stimulus materials such as blocks, pictures, puzzles, and picture cards in addition to the test question. In administering the Utah, 57% of the test items required at least a pencil and paper for the child to respond while 68% of the items through the 5 and 6 year level required materials. Fifty percent of the 18 test sections in the Houston were dependent upon stimulus materials which accompanied the test instrument. The Houston, therefore, needed the fewest extraneous stimulus materials.

Of the three tests, the children all appeared to enjoy the test items on the Hannah-Gardner the most. On several of the items, however, the child seemed to be manipulating the materials for his own pleasure rather than attending to the examiner. For example, ten blocks were placed in front of the child and he was asked to hand the examiner the number of blocks she requested. The child was intent on building a tower with the blocks or lining them up in a particular way rather than counting the blocks.

The stimulus materials required for the Houston did not interfere with the child's responses. There were no idle materials to distract his attention or manipulate inappropriately. Three of the last four items on the Houston did not appear to hold the children's attention. These items involved imitating the examiner's drawing of shapes, drawing a picture, and telling about the picture. Perhaps these quiet activities could be integrated with earlier activities involving some physical

movement. The small toys accompanying the Houston, and especially the doll family, were adequate for eliciting the language sample. The examiner did find it difficult to accurately transcribe by hand the speech of highly verbal children on this item. A tape recorder could be used with this type of child.

The stimulus pictures included with the Hannah-Gardner were cartoon-like figures. The children did not appear to be distracted by these pictures and they appeared adequate for eliciting the desired responses. There is an advantage to using this type of drawing in that they should not become out-dated in a short time.

The stimulus pictures used by the Utah for assessing receptive and expressive vocabulary were bound in a small five and one-half by eight and one-half inch spiral booklet. The pictures for each item were found on a single page with as many as 13 pictures on one page. All pictures, with the exception of the plates evaluating color recognition, were black and white line drawings. Due to their small size and detail, the examiner felt that for administration of the test, larger pictures were needed to increase the clarity of the action or object. This observation was also made by Irwin, Moore, and Rampp (1972).

The only stimulus pictures included in the Houston assessed expressive vocabulary. This section required the naming of 38 picture cards: 20 objects, nine action, three descriptive words, and six colors. The action and attribute cards were well represented and did not appear out-dated despite the 1963 copyright date.

Response Criteria and Scoring

Both the Houston and Utah provided the examiner with guidelines for judging the child's responses. The Houston was more detailed in defining correct and incorrect responses than either of the other two tests. This was especially important on the items concerning the language sample and the picture drawing. Eighty-four percent of the items on the Utah merely required the examiner to make an objective comparison of the child's responses with the number of correct responses required for recording a correct response on that test item.

Following the administration of the Hannah-Gardner to the eight subjects, the examiner felt this test lacked adequate guidelines for judging appropriate responses. The scoring of many of the responses was subjective, inconsistent, and varied between subjects due to the indefinite guidelines provided by the Hannah-Gardner. One of the items without scoring guidelines required the child to take a red ball out of a cup, bounce it once, and give it to the examiner. Only one child was able to complete the item exactly as instructed. It was extremely difficult for the children to catch the small ball after the first bounce. Since this particular question was placed under the auditory perception category testing memory, the examiner usually scored the child's response incorrect because he didn't perform according to the test item. Thus it was difficult to test the children's auditory perceptual memory when the item required the children to perform a difficult motor skill. The examiner was uncertain how to score the item when this apparently ambiguous response occurred.

The Utah and Hannah-Gardner were scored according to the correctness or incorrectness of the response. Fifty percent of the Houston was

scored in this manner with the remaining 50% involved in the analysis of the language sample, drawing, and counting.

A language age was derived from the raw scores on the Utah and Houston tests. The language-age-equivalent for the Utah was expressed in years and months and allowed the comparison for the age ranges from 9 months to 16 years of age. The language ages for the Houston were expressed in years only. Crabtree stated that although language develops in an orderly sequence, it does not develop in precise intervals. This was Crabtree's rationale for presenting language ages only in broad yearly intervals.

The raw scores for each section of the Hannah-Gardner yielded a percentile and total percentile score determined from the sum of the raw scores. No language-age-equivalent was established. The examiner noted several inconsistencies when percentile scores were computed. On the percentile scales in all four sections at least one and as many as five of the six age groups had the same raw scores in two or three different percentiles. For example, on the motor development scale for the age level 5-6 to 5-11 years a raw score of 12 could be converted to either a 10th or 20th percentile score while a score of 13 for the same age could be converted to either a 50th, 75th, or 90th percentile score. One raw score point separated the 10th from the 90th percentile. This duplication of percentiles was true in both the middle socioeconomic and lower socioeconomic level scores. Additionally, one raw score point often separated two or more percentile intervals. These difficulties could be due in part to the small subject population used for the standardization of the test as only 15 subjects were included for each

age level. Another factor involved could be the number of test items which composed each of the four test sections. For example, the visual perception section contained only 12 items which provided a relatively small sample of behavior on which to compute percentiles.

The Houston was not completely scored during test administration. The spontaneous language sample and utterances transcribed while the child was drawing were analyzed following the test session. The examiner found it very important to be thoroughly familiar with the items related to the sample before attempting to score them. Age levels for the 18 items were computed from the child's responses and recorded on the score sheet.

Data Analysis and Diagnostic Information

The Utah did not include any interpretation of the results derived from administration of the test.

The Houston interpreted a language age within a year of the child's chronological age as within normal limits. A score two or more years below the established norms for the child's chronological age indicated a language deficit.

The 20th percentile on the Hannah-Gardner was the cut-off score for indicating normal language performance on this test. The 20th percentile is approximately the first standard deviation below the mean for this test. The area located between the 20th and 10th percentiles was designated as the grey area indicating possible difficulty. Any score below the 10th percentile was evidence of a definite language deficit. Any indication in the test responses of difficulty within a specific section should be noted by the examiner. A child receiving a low score

on the Hannah-Gardner should be referred to a qualified language clinician for further assessment.

The scores of the eight subjects evaluated for the present study revealed that the children of the high group achieved higher scores on all three tests than the children in the low group. These children were all considered by their teachers as essentially normally developing children. The low group consistently scored below the 10th percentile on the Hannah-Gardner while the range of scores for the high group on this test ranged from the 10th to the 75th percentile. There was therefore a greater variation of percentile scores for the children rated as high in their language abilities on the Hannah-Gardner than the low group. The major reason for the percentile scores of the children in the low group appeared less variable was because the scores of the low group were bounded by the lower end of the percentile range. The Utah and Houston seemed to present a more meaningful picture of each child's language performance by providing the language-age-equivalents rather than percentile scores alone. There was no overlap of scores between groups on the Utah or Houston. The low group all scored below their chronological age while the high group scored at or above their chronological age.

None of the three tests analyzed their results in such a way that an intervention program would be easily formulated. The Houston supplied the most complete picture of the child's performance on the tests by recording the appropriate age level for each test item on the score sheet. Crabtree did not indicate, however, which test items appeared first in the sequence of language acquisition or which items were most

essential to the normal development of language. The sample of oral language should be particularly helpful to the clinician although the analysis of the sample was by no means comprehensive.

The Utah provided no summary of test performance as to the language skills assessed beyond the language-age-equivalent. Any further diagnostic information could only be accumulated by sorting through the test items. Of the three tests, the Utah provided the least amount of information on the child's specific strengths and weaknesses.

The sectional arrangement of the Hannah-Gardner provided the examiner with an organized review of the child's responses. This arrangement allowed the examiner to determine whether there was an indication of difficulty within a specific area. The percentile was the only other evidence of a language deficit.

The Hannah-Gardner was the one test of the three designated as a screening test to identify children in need of further assessment. However, due to the time required for administration, this test would not be suitable for the typical screening situation.

The Utah would probably function more efficiently as a screening device rather than a diagnostic test since there is no interpretation of test results or summary of the results supplied with the test. Again, the examiner should keep in mind the time required for administration of this test.

Irwin et al. (1972) noted that the Houston was a relatively popular test, particularly with clinicians unsophisticated with the language development of young children. The Houston contained several test items not included on the two other tests. The language sample, verbalizations

while drawing, and telling about their picture were three of the unique tasks. For the clinician who has had little training in language assessment and particularly in the elicitation and analysis of language samples, this test would provide a base for the clinician to build on. The examiner felt the Houston was more than a screening test yet should be supported by further assessment in areas indicating a difficulty before formulating an intervention program.

Normative Data

The normative data for the Houston were collected by administering the test to 102 children between 2-6 and 6-5 years of age in metropolitan Houston. Children from bilingual homes, multiple births, institutionalized children and those with observable mental or physical defects were not tested. The norms for the items at each age level were determined by a significant difference in the percent of children passing the item from one age to the next. Further, over 50% of the children at that age level were required to have passed the item. Also the subjective judgment of the author helped to determine the norms.

Norms for the Utah were computed on 273 normal white children. All children resided in Utah and were selected as a representative sample of children from this state in respect to residence, socioeconomic status, age, and sex. Approximately 20 children were chosen from each age level from 1-6 to 12-5 years with 30 children selected for ages 12-6 to 14-5 years of age. According to the authors, additional normative data were collected from throughout the U.S. However, this data has not been published as of the third printing of the test in 1973. Language-age-equivalent scores were derived from a graph on which raw scores and

chronological ages were plotted. The authors stated that standard score and percentile equivalents were not computed due to the small sample size.

Normative data for the Hannah-Gardner were collected from a total group of 180 children from middle and lower socioeconomic levels individually. Ages ranged from 3 to 5-6 years in six month intervals, with approximately 14 girls and 16 boys in each group of age levels. However, since data were gathered for the middle and lower socioeconomic levels individually, only 15 children were actually included in each of the age levels. All children lived in the San Fernando Valley area of Los Angeles. Data were collected by students at California State University, Northridge under the supervision of the authors. The authors stated these data should be used as area norms or the normative data should be established for other localities, socioeconomic, and ethnic groups.

Reliability and Validity

To determine reliability an attempt is usually made to ascertain whether different examiners are able to obtain the same results administering the assessment procedure to the same child. This is called interexaminer reliability. Sometimes the test is administered at different times by the same examiner to the same child. This is referred to as intraexaminer reliability or test-retest reliability. Sometimes the reliability of a test is determined by comparing it with other tests or by comparing some items of a test with other items in the same test. Validity is a measure of whether the test actually tests what it purports to test.

The Houston reports no reliability or validity information for the normative data. It did, however, claim validity in the selection of stimulus items based on a reliable difference of the percentage of children passing from one age to the next.

Reliability for the Utah was established by administering the Utah to 117 subjects. There was a .967 correlation coefficient between the scores. The split half correlation coefficient done on odd versus even items on the Utah was .937. No examiner reliability was reported. Validity was obtained by correlating the language-age-equivalents of the items on the Utah against the age equivalents on the original tests from which the items were derived. The Product-moment correlation for this comparison was .983.

Reliability involved testing 29 randomly selected preschool subjects on two occasions at least two weeks but not more than three weeks apart to assess test-retest reliability for the Hannah-Gardner. A Product-moment statistical procedure indicated a correlation of .95 between the two administrations. Intra-tester reliability was .93. Since the normative data were collected by students, a report of interexaminer reliability should also have been included. Validity for the Hannah-Gardner was obtained by establishing the correlation between the Hannah-Gardner and the Illinois Test of Psycholinguistic Abilities (Kirk and McCarthy, 1969) on a preliminary study of 12 children with moderate to severe language deficits. The correlations between these two tests on auditory scores (.55) and linguistic comprehension-grammatical closure (.66) were significant at the .05 level and the total scores (.87), visual scores (.74), and conceptual scores (.70) were significant at the

.01 level. Another study using three groups of 10 subjects each was conducted. The subjects were divided into normal subjects, subjects with diagnosed language delay, and subjects with a diagnosed articulation deficit. The test authors indicated that the Hannah-Gardner distinguished among the three groups although the table that reported these data was very confusing.

SUMMARY AND CONCLUSIONS

Due to the well-organized presentation of test items and the use of few extraneous stimulus materials beyond crayons and paper, the Utah was the easiest of the three procedures to administer. In addition, responses were well defined so there were few subjective judgments that were required of the examiner. The Utah was, however, the least effective of the three tests in maintaining the children's attention during administration. The entire test was scored during the test presentation and the language-age-equivalent was easily computed following completion of the test. This test gave the examiner the most complete language age in years and months. The Utah did not provide any interpretation of test results or a summary of the child's performance on the test. This made it difficult for the examiner to determine the areas where further assessment or training should be concentrated.

The Houston provided the most comprehensive review of the child's test performance and the greatest amount of information on his strengths and weaknesses. The Houston also included the elicitation of a sample of the child's oral language. Although this was valuable information, the examiner was required to score these sections following the test

session. The response guidelines were well presented and provided for consistent scoring between children. Underlining the specific test question in the test manual would simplify the stimulus presentation. The language age obtained from the raw score was expressed in yearly age levels.

The Hannah-Gardner was the most effective of the three tests in maintaining the children's attention. Test questions were adequately presented but response guidelines were needed in greater detail for objective, consistent scoring. The stimulus materials included with the test required more extensive manipulation by the examiner than those for the Utah and Houston. All responses were scored by the examiner following presentation of each item and percentile scores computed at the completion of the test. Percentile scores were derived for each section in addition to a total percentile. The child's test performance was assessed by examining the responses in the test sections and the percentile scores.

REFERENCES

- Bereiter, C. & Englemann, S. Teaching disadvantaged children in the preschool. Toronto: Prentice-Hall, Inc., 1966.
- Berry, M. Language disorders of children. New York: Appleton-Century-Crofts, 1969.
- Bricker, W. & Bricker, D. An early language training strategy.
In R. L. Schiefelbusch & L. L. Lloyd (Eds.), Language perspectives acquisition, retardation, and intervention. Baltimore: University Park Press, 1974.
- Brown, R. A first language: the early stages. Cambridge: Harvard University Press, 1973.
- Bzoch, K. R. & League, R. Assessing language skills in infancy. Gainesville, Florida: Tree of Life Press, 1971.
- Crabtree, M. The Houston test for language development, part II. Houston: The Houston Test Company, 1963.
- Dickson, S. (Ed.) Communication disorders: remedial principles and practices. Glenview, Illinois: Scott, Foresman, and Company, 1974.
- Doll, E. A. The Vineland social maturity scale. Minneapolis: American Guidance Service, 1946.
- Dunn, L. J. Peabody picture vocabulary test. Minneapolis: American Guidance Service, 1959.
- Dunn, L. J. & Smith, J. O. Peabody language development kit, level 2. Circle Pines, Minnesota: American Guidance Service, Inc., 1967.
- Emerick, L. L. & Hatten, J. T. Diagnosis and evaluation in speech pathology. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1974.
- Friedlander, B. Receptive language development in infancy: issues and problems. Merrill-Palmer Quarterly of Behavior and Development, 1970, 16, 7-51.
- Fryer, H. D. Concepts and methods of experimental statistics. Boston: Allyn and Bacon, 1966.
- Gesell, A. L. & Amatruda, D. S. Developmental diagnosis. New York: Hoeber, 1941.
- Hannah, E. P. & Gardner, J. O. Hannah-Gardner preschool language screening test. Northridge, California: Joyce Motion Picture Co., 1974.

- Irwin, J. V., Moore, J. M., & Rampp, D. L. Nonmedical diagnosis and evaluation. In J. V. Irwin & M. Marge (Eds.), Principles of Childhood language disabilities. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1972.
- Kirk, S. & McCarthy, J. J. The Illinois test of psycholinguistic abilities. Urbana: The University of Illinois Press, 1969.
- Lavatelli, C. S. (Eds.) Language training in early childhood education. Urbana: University of Illinois Press, 1971.
- Mecham, M. J. Verbal language development scale. Minneapolis: American Guidance Service, 1959.
- Mecham, M. J., Jex, J., & Jones, J. D. Utah test of language development. Salt Lake City, Utah: Communication Research Associates, Inc., 1967.
- Schiefelbusch, R. L. & Lloyd, L. L. Language perspectives acquisition, retardation, and intervention. Baltimore: University Park Press, 1974.
- Terman, L. M. & Merrill, M. A. Measuring intelligence. Boston: Houghton-Mifflin, 1937.

APPENDIX A

Description of the Three Global Language Tests

Mecham, M. J., Jex, J. L., & Jones, J. D. Utah test of language development. Salt Lake City, Utah: Communication Research Associates, Inc., 1967. (Cost: \$20.00)

The Utah measures both receptive and expressive language by utilizing a developmental approach with test items arranged by age levels. The test may be administered to normal and handicapped individuals from 1 to 15 years of age. The purpose of the Utah as stated by the authors was to determine what the child's language abilities are in a formal testing situation. All test questions were collected from such standardized tests as The Vineland Social Maturity Scale (Doll, 1946), the Peabody Picture Vocabulary Test (Dunn, 1959), Developmental Diagnosis (Gesell & Amatruda, 1941), Verbal Language Development Scale (Mecham, 1959), and Measuring Intelligence (Terman & Merrill, 1937). The Utah is not a timed test and sessions vary from approximately 25 to 40 minutes.

Responses are recorded as plus or minus by the examiner on a score sheet that accompanies the test. Scoring requires eight consecutive correct responses before continuing the test upward until eight consecutive errors are recorded. The highest consecutive plus determines the basal score while the raw score is computed by counting the correct responses above the raw score and adding them to the basal score. A language-age-equivalent is obtained from the raw score.

Responses elicited by the stimulus tasks include pointing to pictures; naming pictures, colors, money, objects, and body parts;

repeating digits and sentences modeled by the examiner; copying geometric shapes; drawing a recognizable picture; carrying out a series of three simple commands; and the child stating his full name.

Stimulus materials included with the Utah are an instruction manual; a booklet containing 13 plates of stimulus pictures for the assessment of receptive and expressive vocabulary and colors; and a doll, gun, hammer, cup, and marble. Materials not included but required for test administration are crayons and paper for coloring, a pencil for copying shapes and for use in the commands, and a box for the three commands.

Crabtree, M. The Houston test for language development, part II.
Houston, Texas: The Houston Test Company, 1963. (Cost: \$24.00)

The purpose of the Houston as stated by Crabtree was to provide a language scale for age levels 3 to 6 years derived from the assessment of receptive and expressive language abilities of the child. The test, based on a developmental assessment of language, provides normative data for test items at each age level. The Houston requires approximately 30 minutes to administer.

A portion of the Houston elicits a spontaneous language sample from the child which is transcribed and analyzed by the examiner. Stimulus tasks are divided into 18 sections assessing self-identity, auditory judgments, gesture, prepositions, counting, vocabulary, repetition of speech and melody patterns, and communicative behavior and verbalizations while drawing.

Analysis of the language sample evaluates the (a) communicative behavior or the stage of self-communication, (b) temporal content or his ability to express time, (c) syntactical complexity or the growth of the child's use of phrases, incomplete sentences, complete sentences, and paragraphs, and (d) sentence length or the child's ability to express himself in thought units comparable to his age group in length. Verbalizations recorded while the child was drawing were analyzed according to self-communication. A score sheet, included on the last page of the work sheet, provides an overview of the test results by supplying the examiner with the (a) basal age or the lowest age at which all items were passed, (b) upper age or the highest age at which any item was passed, (c) raw score or the sum of the scores recorded under each age

level, (d) language age or the age derived from the raw score, and (e) test patterns under which each test item is placed according to the appropriate age level on the basis of responses to the test items.

Responses elicited from the test items include locating body parts on the child; performing various actions requested by the examiner; counting; repeating a single word, sentences, and melody patterns modeled by the examiner; imitating geometric shapes drawn by the examiner; and drawing a recognizable picture and telling about it.

Stimulus materials included in the purchase of the test are an instruction manual; vocabulary cards; and small objects such as farm animals, a table and two chairs, three small trucks, a boat, and a doll family. Materials not included are several crayons and paper for the drawing of geometric designs and a picture.

Hannah, E. P. & Gardner, J. O. Hannah-Gardner preschool language screening test. Northridge, California: Joyce Motion Picture Co., 1974. (Cost: \$43.50)

The Hannah-Gardner was designed as a screening test for basic psycholinguistic development to be administered when more extensive and formal procedures cannot be validly administered. The Hannah-Gardner can be administered by professionals in any field associated with preschool children with referrals made to qualified persons in the field of language on the basis of the percentile scores. This is not a timed test and requires approximately 30 minutes to administer. The authors recommend allowing 10 to 15 seconds for the child to respond to a stimulus item although motor tasks probably require more time.

The skills assessed in the various sections are auditory perception, visual perception, motor development, and conceptual language abilities. The auditory and visual sections are grouped into test items assessing memory span, figure ground, closure, and sequencing.

The record form that accompanies the test is designed to remain in the test question booklet during administration. Test questions are scored as 1 or 0 on the record form. Raw scores and percentile scores are computed at the end of each of the four sections in addition to a total score and percentile based on the results of all four sections. Normative data are provided for middle and lower socioeconomic levels separately. The 20th percentile is the cut-off score for the lower limits of normal behavior. If the total score is below the 10th percentile, it is indicative of a definite language deficit.

The Hannah-Gardner includes the following materials: test instruction manual picture file and Copy Me file, a booklet of test questions

and 10 record sheets, 1 white styrofoam and 1 red rubber ball, 1 bean, 1 small and 1 large keg, 10 wooden blocks of various colors, 1 plastic spoon, 1 yellow pencil, 1 small plastic cup, 1 matchbox with matches (without sulphur tips), 1 piece of felt with button and button hole, 1 orange circle puzzle, 1 orange square puzzle, and 12 sets of picture cards in plastic spiral booklet. All materials needed for test administration accompany the test.

APPENDIX B

Table 2

Comparison of the Common Test Items on the Hannah-Gardner, Utah, and Houston

Hannah-Gardner	Utah	Houston
Copying geometric shapes	Copying geometric shapes	Imitating geometric shapes
Counting - receptive	Counting - receptive graphic object counting	Counting - expressive serial and object counting
Colors - receptive	Colors - receptive and expressive	Colors - expressive
Auditory memory - receptive and expressive 2 sets single syllable words and 2 commands	Auditory memory - expressive digits and sentences	Auditory memory - expressive repetition of 1 word (3 syllables) to 14 words (15 syllables)
Prepositions - receptive		Prepositions - receptive
	Vocabulary - receptive and expressive objects	Vocabulary - expressive objects, action, attributes
	Body parts - receptive	Body parts - receptive

A COMPARISON OF THREE PRESCHOOL LANGUAGE TESTS

by

MEGAN LUETHA KING

B.S., 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 administer and evaluate three global language tests currently used for assessing preschool children's language abilities. The tests were the Hannah-Gardner Language Screening Test, The Houston Test for Language Development, Part II by Crabtree, and the Utah Test of Language Development by Mecham, Jex, and Jones. This research was initiated in an attempt to provide speech and language clinicians with a critical review of these tests.

Eight children selected from a preschool class served as subjects. The subjects were selected on the basis of the child's receptive and expressive language abilities as rated by their teachers and pretest results. The children were divided into two groups, one group of four children rated as most advanced in their language abilities and one group of four children rated as least advanced in their language abilities. All three tests were administered to each child.

The test results were converted into language-age-equivalents to facilitate statistical comparison. This comparison revealed significant differences among the tests in language ages.

The author evaluated and commented on the construction, administration, and diagnostic information obtained from the tests.