A COMPARISON OF THREE SPEECH SOUND
DISCRIMINATION TESTS

by

ANNE LOUISE AHLERS

B.A., Creighton University, 1974

———————————

A MASTER'S THESIS

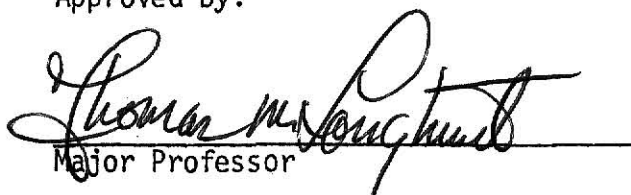submitted in partial fulfillment of the

requirements for the degree

MASTER OF ARTS

Department of Speech

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1975

Approved by:

Major Professor

ACKNOWLEDGEMENTS

I wish to express my sincere thanks and appreciation to my major professor and advisor Dr. Thomas Longhurst for his guidance and support during the preparation of this thesis and throughout my graduate training.

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

My appreciation is also extended to the teachers and the children at the Kansas State University Developmental and Infant and Child Care Centers whose assistance was greatly appreciated.

TABLE OF CONTENTS

LISTING OF TABLES

INTRODUCTION

Speech sound discrimination is the ability to distinguish between individual sounds used in speech. It is generally agreed that speech sound discrimination should be part of a diagnostic battery of tests for speech handicapped children although discussion continues concerning the relationship between discrimination abilities and other variables including articulation, reading, and spelling skills (Hall, 1938; Wepman, 1960; Christine and Christine, 1964; Marquardt and Saxman, 1972). The present study deals entirely with the question of test selection.

Three speech sound discrimination tests which were currently available for clinical usage were evaluated by administering the tests to the same subjects. The Auditory Discrimination Test (Wepman, 1958; revised, 1970), the Goldman-Fristoe-Woodcock Test of Auditory Discrimination (GFW) (Goldman, Fristoe and Woodcock, 1970), and the Test of Listening Accuracy in Children (TLAC) (Mecham, Jex and Jones, 1973) were comparatively evaluated in an attempt to determine their efficiency for differentiation of speech sound discrimination abilities in preschool children. The purpose of the present experiment was to administer the three tests and evaluate the procedures, materials, test instructions, and appropriateness of each test.

Various other discrimination tests were available. The Templin Speech Sound Discrimination Test (Templin, 1957) and the Boston University Speech Sound Discrimination Test (Pronovost and Dumbleton, 1953) are sometimes used clinically although they are not commercially available and include incomplete normative data and reliability and validity statistics. The Oliphant Auditory Discrimination Memory Test (Oliphant,

1971) is a newly developed test which measures auditory memory as well as speech sound discrimination.

None of the tests with the exception of the Wepman have been examined extensively by other authors. It was difficult to relate findings of the previous literature on the Wepman since the various studies used different techniques and subject samples and evaluated specific variables related to the test (reinforcement vs. non-reinforcement; Berlin and Dill, 1967; the significance of the number of same and different items on the test; Vellutino, DeSotto and Steger, 1972). No study has comparatively evaluated the Wepman with the TLAC or the GFW. The use of the GFW is becoming increasingly widespread although relatively little literature is available regarding it. The TLAC is a new test and little research has been conducted utilizing it beyond that reported by the authors of the test.

The Wepman assesses a child's ability to make fine distinctions between English sounds. The test consists of 40 word pairs; 30 pairs that are different and 10 pairs that are the same. The subject is required to judge whether two orally presented words are the same or different.

The GFW was designed to measure speech sound discrimination ability in a test situation relatively unconfounded by other factors. It provides a measure of speech sound discrimination under quiet conditions and during the presence of controlled background noise. Background noise which was 9 dB less intense than the signal was superimposed on the noise subtest portion of the tape. The noise environment consists of cafeteria noise and voice babble. The quiet and noise subtests

consist of 30 items each. The subject is asked to point to one among four pictures which represents the word presented through a pre-recorded audio tape.

The TLAC requires a word-picture matching response by the child who hears the stimuli in the presence of white noise on a pre-recorded audio tape. The signal to noise ratio was not reported in the test manual. Three words are presented, one of which is the correct name of one of three pictures. The other two words are acoustically similar to the names of the other two pictures. The child is asked to point to the picture that was named correctly. Eighty-six items comprise the test.

Appendix A provides a more complete description of the three tests. Appendix B presents comparisons of the tests on 13 factors gleaned from the test manual and instructions.

A brief appraisal of the Wepman, the TLAC, and the GFW indicates that the tests differ in length, familiarity of items, types of response, stimulus presentation, scoring, methods and normative data. For these reasons the present experiment was conducted in an attempt to study the tests on a single population so that the differences among the tests might be more closely scrutinized. The study was designed to contribute to a more knowledgeable selection of a speech sound discrimination test for clinical use.

METHOD

Subjects

Eighteen children constituted the sample in this study. All subjects were attending preschool classes at the Kansas State University Developmental and Infant and Child Care Centers. The children ranged in

age from 3 years 1 month to 5 years 6 months. Mean age was 4 years 2 months. There were eight 3 year-olds, eight 4 year-olds and two 5 year-old children in the sample. No children below the chronological age of three years were chosen as subjects. No attempt was made to control the sex ratio of the sample. All subjects had no known physical, mental or neurological handicaps and were of middle socio-economic status. No bilingual children were incl ded as subjects.

Pretests

The Peabody Picture Vocabulary Test (PPVT) Form A (Dunn, 1959) and an audiological screening were administered to the children prior to administration of the speech sound discrimination tests. Vocabulary intelligence quotients on the PPVT ranged from 78 to 134 with a mean of 111. A pure tone audiometric screening was administered to each subject at 25 dB ISO at 500 Hz, 1, 2, 4, and 6K Hz. All of the children responded appropriately at all test frequences in both ears.

Experimental Facility

The tests were administered in a quiet room at the children's school. A cassette tape-recorder, stopwatch and test materials were present for test administration. The examiner and the subject were seated facing each other for all test administrations. The three discrimination tests were placed on a shelf underneath the table. Three cards containing the names of the tests were placed face down on the table. Each subject selected one card and was administered the test chosen. On succeeding days the subjects selected from the remaining cards. Each child responded to the three tests and no subject was administered more than one test per session. The time required for test

administration was recorded for all subjects. Testing was conducted four mornings a week for a period of five weeks. Individual sessions ranged from 4.30 minutes to 32.14 minutes. The standard instructions and training procedures specified in the manuals of the Wepman, the TLAC, and the GFW were given to each subject prior to test administration. The tests were administered according to the instructions designated by the authors of the tests.

Test Comparison Criteria

The three tests were compared using the following criteria: type of stimulus presentation, method of stimulus presentation, practice items, stimuli, discrimination task, type of response, scoring methods, word characteristics, vocabulary training procedure, number of test items and administration time, retest availability, test score analysis, normative data and reliability and validity. The examiner made subjective notes on these criteria as the tests were administered and a more extensive evaluation was made after the test administration had been completed for all of the children. Each of the criteria will be discussed individually for each test.

In order to compare the test results of the 18 subjects, the percentage of items responded to correctly and the mean number of correct responses were computed for each of the tests. The GFW quiet subtest and the Wepman (quiet) were compared on these features as were the GFW noise subtest and the TLAC (noise).

The percentile scores were also determined for the quiet and noise subtests of the GFW from the normative data provided in the manual. It was not possible to use the normative data supplied with the Wepman or

the TLAC since it was collected on children older than those who partici-
pated in the present study.

RESULTS AND DISCUSSION

Type of Stimulus Presentation

The GFW contained a quiet subtest and a noise subtest. The entire
TLAC was administered in a background of white noise with no provision
for quiet testing while the Wepman tested speech sound discrimination in
quiet only.

Speech sound discrimination testing under quiet conditions may not
be representative of the environment in which the children would
normally be required to discriminate. Testing under such conditions
provided only a measure of discrimination ability in an unrealistic
situation, the results of which cannot be assumed to be representative
of their ability during noisy conditions. While speech sound discrim-
ination testing in quiet is not representative of many real life
situations, tests with both quiet and noise subtests allow the examiner
to compare responses under the two types of test presentation. Consid-
erably lower scores on noise tests in comparison to quiet tests would
indicate that the child had difficulty discriminating in a noisy
environment provided the tests were of equal difficulty on all other
variables.

The GFW was most complete in this aspect since it was the only test
of those examined which contained a noise and a quiet test.

Method of Stimulus Presentation

The stimuli on the TLAC and the GFW were presented from pre-recorded audio tapes while the stimuli for the Wepman were presented by the examiner orally. The TLAC and the GFW should yield consistent results regardless of the clinical setting because of the standard presentation of stimuli. This cannot be assured with the Wepman since the test items were presented orally by the examiner. Individual examiner differences in voice quality and inflections as well as the ability to present the words with the same rate and intensity on the Wepman will vary between examiners. Slight emphasis or prolongations of sounds which differentiate the word pairs may affect the children's scores (Pronovost and Dumbleton, 1953).

Practice Items

There were no practice items on the quiet subtest of the GFW and three on the noise subtest which were presented on an audio tape. There were six practice items on the TLAC which were also administered through an audio tape. The five practice items on the Wepman were presented orally by the examiner. Prior to the test administration on the GFW the children were administered a vocabulary training procedure which required pointing to one of four pictures. Thus the children were accustomed to the pointing task prior to the taped presentation of test items. This orientation to the task appeared to be an advantage in comparison with the TLAC in which initially no oral instructions by the examiner were presented. The tape on the TLAC began with the instructions to the children regarding their task and immediately presented the practice items. The instructions for the TLAC were not sufficient for

any of the children. It was necessary to stop the tape and oral instruc-
tions were presented by the examiner before the children indicated
understanding. This was not necessary with the GFW; only three of the
18 subjects indicated a lack of understanding of the task required of
them when the tape began which required the examiner to stop the tape
and explain the instructions to the children. Although most of the
children responded well to the GFW immediately, the examiner felt that
three practice items on the GFW quiet subtest would have been appropri-
ate to accustom the children to listening to the tape before the test
items were presented.

The practice items on the Wepman were presented orally by the
examiner. The time necessary for assuring that the children understood
their task was longest on the Wepman based on the time needed for having
the children respond correctly to all practice items. The GFW provided
the easiest and most clear method of training the children how to
respond to the test stimuli.

Stimuli

Visual and auditory stimuli were used on the TLAC and the GFW. The
Wepman utilized auditory stimuli only. The auditory stimuli on the TLAC
were three words per test item presented from an audio recording. The
auditory stimuli on the GFW were single words presented for each test
item from an audio recording. The only stimuli used on the Wepman were
word pairs presented orally by the examiner. The visual stimuli used in
the TLAC and the GFW were black and white line drawings which adequately
depicted the vocabulary they were intended to represent. The use of
pictures appeared to be advantageous since they served to aid in

maintaining the children's attention in comparison with the Wepman in which only auditory stimuli were provided. The use of pictures also precluded the necessity of speech or writing abilities by the subjects which may be an important consideration with relatively young or old subjects as well as handicapped subjects.

Discrimination Task

The task on the Wepman involved discriminating between two words. Three words were discriminated among on the TLAC and four were discriminated on the GFW. Only two words were presented for each test item on the Wepman allowing on the average a 50% chance of correct responses simply through guessing based on the same-different method of responding. If the subjects were required to repeat the word pairs after the examiner the possibility of chance guessing would be reduced since the subjects would need to be able to discriminate between the words in order to repeat them correctly. This method may involve other problems, however, including the possible inability of the subjects to articulate all of the consonants used in the test items. An unequal number of same and different items are included on the Wepman. There are 30 different items and 10 same items. The unequal balance of same and different items may affect scores (Vellutino, DeSotto, and Steger, 1972).

The chances of guessing correctly on the GFW and the TLAC were 25% and 33-1/3% respectively on the average. Thus the children did not have as great a chance to guess correctly as they did on the Wepman.

Type of Response

The GFW and the TLAC required the subjects to point to one of four pictures and one of three pictures respectively. On the Wepman the

children were instructed to respond "same" or "different" or indicate by nodding and shaking their heads whether the pairs of words presented by the examiner were the same or different.

The use of the pointing response was considered an asset by the examiner since the pictures appeared to interest the children and maintained their attention. Blank, (1968); Monses, (1968) and the present examiner considered the same-different method of responding the main disadvantage of the Wepman with children aged three and four years. None of the subjects indicated understanding of the task required of them on the first explanation by the examiner nor did they indicate understanding after the first practice item was presented. The examiner felt that the concepts of same and different were not known by many of the children which resulted in some invalid test scores. The concepts are abstract and were difficult to train in a short time before test administration. Knowledge of these concepts should not be a necessary prerequisite to speech sound discrimination testing.

The examiner noted that seven of the children imitated the word pairs on the Wepman after the examiner before responding "same" or "different." Many of the children repeated items such as "tub-tug" correctly by differentiating the last sound but responded "same" to the pair. Based on the repetitions of the word pair the examiner knew that the children had the ability to discriminate the words but had responded incorrectly possibly because of the confusion between the terms same and different. Consequently the children's scores were low based on the same-different responses but they would have been higher had the test been scored according to the children's repetitions of the word pairs.

The examiner scored the test record form according to the instructions in the manual based on the responses of same and different. She also scored the repetitions of the word pairs in the margins of the record form for those seven children who repeated items after the examiner. These scores were used to determine the score obtained with this type of response and were compared with the scores obtained when the test was scored according to the same-different responses.

The percent correct on the Wepman based on the same-different method of scoring ranged from 55% to 85% for the seven children who repeated the word pairs after the examiner. The mean percent correct was 68 for these children.

Table 1

Percent Correct on Two Types of
Responding to the Wepman

Subjects	Same-Different Method	Repetition of Word Pair Method	Percent Increase
1	34/40 = 85%	37/40 = 93%	8
2	26/40 = 65%	39/40 = 98%	33
3	29/40 = 73%	35/40 = 88%	15
4	31/40 = 78%	35/40 = 88%	10
5	22/40 = 55%	26/40 = 65%	10
6	31/40 = 78%	34/40 = 85%	7
7	27/40 = 68%	32/40 = 80%	12

Based on the repetitions of word pairs after the examiner, the percentage of items responded to correctly ranged from 65 to 98 with a mean of 85. The mean percent increase between the two types of

responding was 13 with a range of 7 to 33. The average number of errors was 11 and six on the standard method of responding and the repetition method respectively.

The percentage of items responded to correctly increased with all seven of the subjects who repeated the word pairs in comparison with the same-different responses. The examiner felt these scores more accurately reflected these children's discrimination abilities than the scores based on the same-different responses.

Scoring Methods

Each of the three tests was easy to score. The tests required objective scoring of responses according to the correctness or the incorrectness of the responses.

Two response columns were labeled on the Wepman test form; one column marked DIFFERENT and the other column marked SAME. Scoring on the Wepman consisted of marking a plus or a minus after the word pairs based on the responses of same and different which was adequate.

The TLAC was scored as a plus for correct responses and a minus for incorrect responses. The total score was derived by adding the number of correct responses on the test.

The examiner marked the number of the picture selected on the GFW and did not attempt to score the items as correct or incorrect during the test administration. The number of errors in each section was used to determine the percentile rank of the children.

The GFW was the only test which required additional time to score after the test was administered. The examiner compared the numbers selected by the subjects with the correct responses which were printed

on the test form in parentheses beside each test item. The additional time required was minimal and worthwhile since it provided information on the correctness or the incorrectness of the responses and also on what type of substitutions were made for the correct responses.

Word Characteristics

The words for each test item on the Wepman differed in one sound only which was important for analyzing test results. The words on the individual test items on the GFW also differed in one sound only although there were four words per test item. The individual test items on the TLAC differed in more than one sound.

The advantage of the difference in one sound only was that the examiner considered an error response directly attributable to the sound that varied. If a child responded incorrectly to a test item on the GFW or the Wepman the examiner knew immediately the sounds which the child had difficulty discriminating between. This was not possible with the TLAC which resulted in information only in whether the children had adequate discrimination ability or not. The GFW and the Wepman provided information on what types of discrimination problems or speech sound confusions the children had which would be useful information if intervention was applied.

The words on the GFW and the Wepman were equated for length in each test item. This was not a feature of the TLAC in which one, two, and three syllable words were presented as stimuli in a single test item. The fact that the words were not equated for length in the TLAC did not appear to be detrimental to adequate discrimination.

It appeared that the TLAC tested auditory memory as well as speech sound discrimination which may have reduced the effectiveness of the instrument as a speech sound discrimination test. The children were required to listen to and retain three words before they could compare the mentally stored words with the pictures and discriminate among them.

The Wepman also required auditory memory abilities as the children were required to listen to and retain a sequence of two words long enough to judge if they were the same or two different words.

Vocabulary Training

The GFW was the only test of the three which included a vocabulary training procedure. This was not necessary with the Wepman as it was with the GFW and the TLAC since the children did not need to associate a picture with a vocabulary word on the Wepman but only needed to indicate if the words were the same or different. The TLAC was inadequate in this aspect.

Another disadvantage of the TLAC was the number of vocabulary words used throughout the test. The authors of the TLAC presumed the test words were known by kindergarten, first and second grade children. It was not possible to determine if the children failed items due to poor discrimination ability, poor auditory memory or inadequate receptive vocabularies.

The GFW included a vocabulary training procedure which was essential because it trained the vocabulary to be used later in the quiet and noise subtests. Sixteen plates with four pictures per plate were trained. Two words were trained initially for each test plate. Following this section the entire set of plates was used again to train an additional 32 words.

The entire procedure was time consuming and the examiner was required to complete the training procedure with subjects who pointed to the correct picture initially. The training procedure required an additional period of time for subjects who did not know receptively the vocabulary words depicted on the first trial. In such cases a second trial was given and, if necessary, a third trial would have been appropriate. None of the subjects required three trials, however, 10 of the 18 subjects required two trials.

It was presumed by the authors of the GFW that after the training procedure the children were familiar with the vocabulary words and their only task was to discriminate among the words. The examiner felt that in some cases it was not appropriate to consider the vocabulary words learned sufficiently well after such a minimal amount of training to assure that the children's only task would be discrimination of the items subsequently on the two subtests. This was a subjective judgment based on the observation that some of the children did not respond to some items which used the vocabulary words on which they had just been trained.

Number of Test Items and Administration Time

The GFW was arranged according to a 64 item vocabulary training procedure, a 30 item quiet section and a 30 item noise section. The Wepman consisted of five practice items and 40 test items while the TLAC consisted of six practice items and 86 test items.

Total test administration time of the GFW was reported as 7-1/2 minutes and approximately 5 minutes for the training procedure. The mean time for GFW administration based on 18 subjects was 14.03 minutes.

The time required for training the vocabulary significantly increased the total test administration time. The examiner felt it was appropriate to consider the training procedure time in the total test administration time since the children were required to attend to the stimuli and respond at this time also.

The mean time required for administration of the 86 test items and the six practice items on the TLAC was 26.15 minutes based on 18 subjects. The authors reported the average test administration time as approximately 20 minutes. Fifteen of the 18 subjects performed poorer on the second half of the TLAC although the difficulty of the test items supposedly did not increase. This was considered due to the length of the test. These 15 children indicated lack of attention and poor responsiveness although the tape was stopped momentarily and the children were encouraged to continue.

An advantage of the Wepman was the length of time required for administration of the practice and test items. In the manual Wepman stated that test administration time was approximately 5 minutes. The mean test administration time was 7.32 minutes for the 18 subjects in the present study including the time required for the five practice items. The examiner considered this an advantage over the GFW and the TLAC which required considerably more time to administer and increased the difficulty of maintaining the children's attention throughout the entire test.

The tasks on the TLAC and the Wepman theoretically remained the same throughout the entire test. The GFW was divided into three sections; the training procedure (presented orally by the examiner), the quiet subtest and the noise subtest. The examiner considered the GFW

better organized than the TLAC because the tasks on the GFW varied
reducing the monotony for the children. The children's task remained
the same throughout the entire TLAC. This arrangement was not effective
in maintaining the children's attention. Although the task remained the
same on the Wepman throughout the entire test, it was not considered a
hindrance since the total test administration time was relatively short.
The children attended best to the GFW.

Retest

A retest was available for the Wepman which consisted of an entire
set of 40 new word pairs. The provision of a second test was not
included with the GFW or the TLAC. This feature would be a valuable
asset should a subject perform poorly on the initial testing and require
further testing.

Test Score Analysis

Wepman vs. GFW (Quiet)

The mean percent correct on the Wepman and the GFW quiet subtest
were very similar (81% and 79% respectively). The mean number of
correct responses was 32 out of 40 items and 24 out of 30 items on the
Wepman and the GFW quiet subtest respectively. The percent correct
scores on the GFW and the Wepman were within 5% of each other for 12 out
of the 18 subjects used in the present study.

TLAC vs. GFW (Noise)

The percentage of items responded to correctly on the TLAC ranged
from 43 to 77 with a mean of 61. The percent correct ranged from 23 to
60 on the GFW noise subtest with a mean of 42. The mean number of items
responded to correctly on the GFW noise subtest was 12 out of 30 items.

The mean number of correct responses on the TLAC was 58 out of 86 items. Although the mean number of errors on the TLAC was higher than on the GFW noise subtest the percentage of items responded to correctly remained higher on the TLAC since the TLAC contained 56 more test items. Fifteen of the 18 subjects achieved 50% correct or above on the TLAC while only five of the 18 children scored 50% or more correct on the GFW noise subtest.

Appendix C contains a tabular comparison of the percentage of items responded to correctly on the TLAC, the Wepman and both subtests of the GFW for all subjects.

GFW Analysis

The test scores on the GFW were interpreted in two ways. The simplest way required counting of the errors made in each subtest, recording the number of errors on the record form and translating the error scores into percentile scores. This method of determining the percentile score was not applicable for five of the 18 subjects because the normative data in the GFW manual for the percentile scores began at age 3 years 8 months. These five children's chronological ages were below 3 years 8 months so the error scores of the remaining 13 children were used to determine percentile rankings. Percentiles on the noise subtest of the GFW ranged from the 13th to the 80th with a mean percentile of 38. The percentiles on the quiet subtest ranged from the 18th to the 86th with a mean of 65. Of the 13 children used in this section of the test analysis, 10 were below the 50th percentile on the noise subtest while only three were below the 50th percentile on the quiet subtest. Thus 77% of the children performed below the 50th percentile on the noise subtest and 23% scored below the 50th percentile on the

quiet subtest. It was expected that the children who participated in this study would perform at least at the 50th percentile on both subtests of the GFW since they were normal children and the normative data on which the percentile scores were based were collected through administration of the GFW to normal children. The results of the percentile rankings indicated that the GFW noise subtest was too difficult for adequate discrimination by normal three, four, and five year old children. The second type of scoring on the GFW produced an error score for the voiced sounds, unvoiced sounds, plosives, continuants, and nasals. The method involved in this procedure consisted of counting the number of errors for each of the five types and recording the errors in the appropriate cell of the matrix on the record form for both subtests.

Table 2

Mean Number of Errors in
the Five Categories

	Plosives	Continuants	Nasals	Voiced	Unvoiced
Quiet	2.6	2.7	1.0	4.1	3.0
Noise	6.5	7.2	2.9	7.8	6.1

The results obtained with the subjects in the present study did not entirely support the data on the proportion of errors reported by the authors of the GFW who found nasals and voiced continuants were the most frequent types of errors on the noise subtest and nasals were the most frequent on the quiet subtest. Errors on voiced continuants did occur with the greatest frequency on the noise subtest for the 18 subjects in

the present study. The errors on the nasals were the lowest type of error in both subtests. The authors of the GFW cautioned that the use of the Error Matrix Analysis method should be limited to research purposes and should not be a routine clinical procedure since reliability was low.

Normative Data

TLAC

Tentative normative data was collected on the TLAC through administration of the test to 300 elementary school aged children selected randomly from two school districts in Utah. The districts were rural, suburban, and urban. There were 86 kindergarten, 126 first grade, and 94 second grade children tested. Approximately half of the subjects were male and half were female at each age level. Percentile scores based on the raw scores of the stratified random sample were reported in the manual. Means and standard deviations of the scores were also provided in the manual. A scale was devised which indicated whether the children performed superior, excellent, average, inferior, or very poor on the test in comparison with other children of the same general age level. Further normative data obtained over a wider age range would increase the value of the TLAC. The present data would be more precise and meaningful if levels in ages were reported rather than grade levels since the ages in kindergarten, first, and second grade can vary among a single class.

Wepman

Normative data were provided for ages 5 through 8 years on the Wepman. Wepman did not include information concerning the collection of

the normative data in the manual. The data were based on a study by Turiads, Wepman, and Morency (1972) in which 1,008 children (800 from schools in upper New York state and the remainder from a public school in Chicago) were administered a perceptual test battery including the original form of the Wepman (1958). Median discrimination scores at ages 5 through 8 years were reported which were used to develop the percentile rating scores. The standardization and interpretation tables derived from the results of the study were easily readable and had the advantage over the GFW and the TLAC of being based on the results of scores of a larger group of children. Wepman did not report how he derived the relationship between the rating scale which ranged from very good development to below the level of the threshold of adequacy and the raw scores. For example, the median test score of the normative group at age 6 was 24.6 which was logically interpreted as average discrimination ability and designated as such on the rating scale. Wepman did not report how he derived ratings between the other four categories; very good development, above average ability, below average discrimination ability, and below the threshold of adequacy and the raw scores of the children on whom the normative data were based.

GFW

The standardization sample for the GFW included 745 subjects ranging in age from 3 to 84 years. The number of subjects at each age level ranged from six to 83. Median number of subjects was 57. Mean error scores and standard deviations by age level were reported for the quiet and noise subtests. Performance curves were plotted representing the 25th, 50th and 75th percentiles at all age ranges. The tables of norms included standard score norms and percentile score norms for the

various age ranges. Percentile scores to middle of score intervals were reported but designated for research use only. The normative data on the GFW ranged from age 3 years 8 months to above 70 years which was an advantage over that provided for the TLAC or the Wepman. It allowed for a much finer age distinction than the TLAC which was divided roughly into three one year intervals and the Wepman which was divided into four one year levels. The extent of the ages to which the GFW was applicable was considered an important asset considering the emphasis currently placed on evaluation of children with possible speech or hearing deficits at an early age. The GFW provided the most detailed selection of subjects, analysis of sample subject scores and normative data in comparison with the other two tests.

Reliability and Validity

TLAC

Internal consistency reliability of the TLAC was assessed through the Kuder-Richardson analysis of variance technique. Reliability estimates were determined for three grade levels. The reliability of the kindergarten grade level was .828, first grade was .792 and second grade was .754. No test-retest reliability or validity estimates were reported.

GFW

The GFW was administered to 242 clinical children to determine standard errors of measurement for scores on the quiet and noise subtests. The internal consistency reliabilities were calculated by the split-half method and correlated by the Spearman-Brown formula. Standard errors of measurement on the internal consistency reliability

for the standardization sample were 2.07 and 2.39 for the composite sample. Test-retest reliability of the GFW was determined through administration of the test to 17 preschool handicapped children and readministration after a period of two weeks. A test-retest correlation of .87 was obtained on the quiet subtest and a correlation of .81 was obtained on the noise subtest. Content, concurrent and construct validity of the GFW were reported. The authors stated that the content of the GFW was valid since it consisted of familiar words and required subjects to make fine speech sound discriminations in a controlled test situation which closely resembled speech sound discrimination in a real life situation. Concurrent validity of the test was assessed through correlations with other measures of the same variable. GFW test scores were compared with the ratings of clinical judgers at the Bill Wilkerson Hearing and Speech Center, Nashville, Tennessee who rated 18 subjects as poor discriminators and 12 subjects as good discriminators. A point biserial correlation coefficient of .72 was obtained on test scores on the noise subtest and clinical judgments. A correlation of .68 was obtained between the test scores and clinical judgments on the quiet subtest. Product-moment correlations were .53 between error scores on both subtests in a composite clinical sample aged 4 through 12 years. Based on standardization sample, product-moment correlations of .47, .38, and .32 between error scores were obtained for subjects aged 3 to 6 years, 7 to 11 years, and 12 to 70 years and above respectively. The total product-moment correlation was .39. Three types of evidence were included which indicated construct validity of the GFW; changes in performance with age, relative levels of performance among selected

groups of subjects, and low correlations with non-measures of speech
sound discrimination. Scores on the quiet and noise subtests demon-
strated the expected changes with age; mean error scores decreased while
subjects were in their twenties, scores remained stable until approxi-
mately age 40 after which there was a gradual increase in error scores.
A second type of construct validity of the GFW was determined through
the interquartile range and median performance of the standardization
sample based on the general population in comparison to groups expected
to perform poorer on a valid test of speech sound discrimination than
the general population. A group of 12 good discriminators performed
better than children with speech and language problems and the popu-
lation in general. The authors concluded that the marked distinction
between the good and poor discriminators contributed further to the
evidence that the GFW was a valid measure of speech sound discrimination.
The final area of evidence for support of the construct validity of the
GFW was the presence of relatively low correlations between the noise
and quiet subtests and several non-measures of speech sound discrimina-
tion including the Stanford-Binet which resulted in product-moment
correlations of .60 and .52 between the quiet and noise subtest scores
and Stanford-Binet IQs. Product-moment correlations of .15 and .00 were
obtained between the Peabody Picture Vocabulary Test IQs and the test
scores on the quiet and noise subtests based on a sample of 19 educable
mentally retarded children. Correlations of -.18 and .15 were obtained
between Templin-Darley Screening scores and both subtests of the GFW
based on the same subject sample. Both subtests also obtained low
correlations between the Primary Mental Abilities Tests, the raw scores

of unpublished tests of letter recognition and auditory blending, and a test of auditory retention span for related syllables.

Wepman

The test-retest reliability coefficients of the Wepman were .91 (N = 109) and .95 (n = 279). The two forms of the test showed a reliability coefficient of .92. A Spearman rank-order correlation coefficient of .62 was obtained on the difficulty of the phoneme comparisons (n = 67). Eight studies measuring the validity of the test were reported by Wepman in the test manual. A six sub-scale Perceptual Test Battery was administered to 1,008 children aged 5 through 8 years which included the Wepman Auditory Discrimination Test. Median test score at age 5 was 23.8, age 6 - 24.6, age 7 - 26.3 and at age 8 - 27.3. Positive and significant relationships continued to exist as assessed in a longitudinal study relating auditory and visual perceptual ability at the first grade to school achievement at the fourth, fifth, and sixth grades. Of 213 fourth grade children referred to an urban remedial reading program for study, 94 showed inadequate speech sound discrimination, 114 showed adequate speech sound discrimination and five showed invalid tests. The differences in speech sound discrimination and reading were significant (.01 level) for 80 children in the first grade and 76 children in the second grade. The difference in IQ was not significant at the first grade level but significant at the .05 level in the second grade. Mean improvement differences between speech sound discrimination scores were 3.4 (S.E. 0.142) in a longitudinal study of 172 children in the first and third grades. A t-test of the difference was 8.34, significant at the .01 level. Auditory perceptual ability and

school achievement as assessed through the Metropolitan Achievement Test in the first grade showed significant correlations at the .01 level between each subtest of the MAT and Wepman test scores ranging between .235 and .348 (n = 177). First grade children with articulatory defects and children with normal articulation were compared on four factors; speech sound discrimination, articulation, intelligence, and reading achievement. Significant differences were found between the groups on speech sound discrimination and articulation at the .01 level but not between intelligence and reading achievement.

Suggestions for Further Research

An important disadvantage of each of the three tests was that the children were only required to discriminate words presented by others. Schiefelbusch (1958) suggested that children may show excellent discrimination for the errors of others but have difficulty recognizing their own errors. Another disadvantage of no active participation by the children was the inability of the examiner to determine if they were attending to the tasks.

A more complete speech sound discrimination test should assess subjects' discrimination abilities for speech as they hear words from others, as they produce them themselves, and as they evaluate them silently (Schiefelbusch, 1958). Sanders (1972) provided a simple yet complete measure based on principles similar to these which were suggested by Aungst and Frick (1964). Her evaluation consisted of four sections. In the first section the child matched two extraneously produced sounds presented by the examiner by responding "same" or "different" to the stimuli. In the next step the child matched an

external sound presented by the examiner to one produced by himself through imitation. The third section assessed the subject's ability to match an external sound to an internal criterion. The subject indicated whether the examiner's production of single words was correct or incorrect. The final section required the child to match an internally produced sound with an internal criterion. The child produced the sound and determined if his production was accurate or inaccurate. Except for the Phoneme Perception Task (Locke, 1970) no formal tests or normative data which included the last three criteria are available.

SUMMARY AND CONCLUSIONS

The Wepman was considered the best speech sound discrimination test of the three evaluated for screening purposes provided the subjects knew the concepts same and different. The test could also be useful with subjects who did not understand these concepts if they were instructed to repeat the items after the examiner. The Wepman was selected as the best test for a screening instrument primarily due to the time required for test administration. Another advantage of the test was that it required no visual, speech or reading abilities on the part of the subjects.

The examiner should be thoroughly familiar with the word pairs and method of presentation prior to test administration in an attempt to assure that the stimuli will be presented identically with the same rate, inflection, and intensity.

The Wepman maintained the children's attention better than the TLAC but not as well as the GFW particularly with those seven children who

repeated the word pairs after the examiner although they had not been told to do so. The main disadvantage of the Wepman was the same-different method of responding.

The GFW was regarded as more suitable as a diagnostic instrument to be used with children who performed poorly on a screening test and further evaluation was desired. The GFW required approximately twice as long ᴛᴏ administer as the Wepman. The longer time needed for administration of the GFW would be profitable with children suspected of having speech sound discrimination problems. The more complete analysis of the children's abilities obtained from the GFW would justify the length of time needed for test administration.

The GFW was most effective in maintaining the children's attention probably because the tasks on the test changed reducing the monotony for the children. It also supplied the most complete information on the children's speech sound discrimination abilities since it assessed performance in a quiet and noise background. It was the only test to provide a method to analyze the children's errors beyond converting the test scores to percentiles and a rating.

The TLAC was the easiest test to administer since the instructions, the practice items, and the test items were presented from an audio tape. The length of the test, the inability to differentiate which item caused the discrimination difficulty for failed items, the extent of the auditory memory required and the number of vocabulary words used out-weighed the advantages of good pictures and the tape-recorded presentation of the test.

The TLAC was least effective in maintaining the children's attention. They often attended to approximately half of the test items

and had difficulty concentrating on the last section of the test which reduced their overall scores and consequently their percentile rankings. This resulted in scores not indicative of their actual speech sound discrimination ability. The children's task on the TLAC remained the same for all 86 test items which was ineffective in maintaining their attention. The examiner felt it would be more profitable to indicate the letter (A,B,C) of the item the children selected since this would provide information on the types of errors should a detailed analysis of the scores be undertaken.

REFERENCES

Anderson, P. The relationship of normal and defective articulation of the consonant[s] in various phonetic contexts to auditory discrimination between normal and defective[s] productions among children from Kindergarten through 4th grade. M.A. thesis, State University of Iowa, 1949.

Aungst, L.F. & Frick, J.V. Auditory discrimination ability and consistency of /r/. Journal of Speech and Hearing Disorders, 1964, 29, 76-85.

Barnes, H.G. Diagnosis of speech needs and abilities of students in a required course in speech training at the State University of Iowa. Doctors thesis, State University of Iowa, 1932.

Berlin, C.I. & Dill, A.C. The effects of feedback and positive reinforcement on the Wepman Auditory Discrimination Test, scores of lower-class Negro and white children. Journal of Speech and Hearing Research, 1967, 10, 384-389.

Blank, M. Cognitive processes in auditory discrimination in normal and retarded readers. Child Development, 1968, 39, 1091-1101.

Christine, D. & Christine, C. The relationship of auditory discrimination to articulatory defects and reading retardation. The Elementary School Journal, 1964, 65, 97-100.

Dunn, L.M. Peabody Picture Vocabulary Test. Minneapolis: American Guidance Service, Inc., 1959.

Ewing, A.W.G. Aphasia in Children. Oxford University Press, London: Humphrey Milford, 1930.

Goldman, R., Fristoe, M., & Woodcock, R. Test of Auditory Discrimination. Circle Pines, Minnesota: American Guidance Service, Inc., 1970.

Gruber, L. & Steer, M.D. Auditory perceptual abilities of institutionalized mentally retarded children. American Journal of Mental Deficiency, 1965, 70, 287-290.

Hall, M.E. Auditory factors in functional articulatory speech defects. Journal of Experimental Education, 1938, 7, 110-132.

Locke, J. Phoneme Perception Task. Champaign, Illinois: Children's Research Center, University of Illinois, 1970.

Marquardt, T.P. & Saxman, J.H. Language comprehension and auditory discrimination in articulation deficient kindergarten children. Journal of Speech and Hearing Research, 1972, 15, 383-389.

Mecham, M.J. & Jex, J.L. Picture Speech Discrimination Test.
    Brigham Young University Press, 1962.

Mecham, M.J. The development and application of procedures for
    measuring speech improvement in mentally deficient children.
    American Journal of Mental Deficiency, 1955, 60, 301-306.

Mecham, M.J., Jex, J.L., & Jones, J.D. Test of Listening Accuracy in
    Children. Salt Lake City, Utah: Communication Research Associates,
    Inc., 1973.

Morency, A.S., Wepman, J.M., & Weiner, P.S. Studies in speech: develop-
    mental articulation inaccuracy. The Elementary School Journal,
    1967, 67, 329-337.

Monses, E.K. Temporal sequence and expressive language disorders.
    Exceptional Children, 1968, 35, 141-147.

Oliphant, G.G. Oliphant Auditory Discrimination Memory Test. Educators
    Publishing Service, Inc., 75 Moulton Street, Cambridge, Massachu-
    setts, 02138.

Prins, D. Relations among specific articulatory deviations and responses
    to a clinical measure of sound discrimination ability. Journal of
    Speech and Hearing Disorders, 1963, 28, 383-387.

Pronovost, W. & Dumbleton, C. A picture-type speech sound discrimination
    test. Journal of Speech and Hearing Disorders, 1953, 18, 258-266.

Sanders, L.J. Procedure guides for evaluation of speech and language
    disorders in children. Danville, Illinois: The Interstate
    Printers & Publishers, Inc., 1972.

Schiefelbusch, R.L. & Lindsey, M.J. A new test of sound discrimination.
    Journal of Speech and Hearing Disorders, 1958, 23, 153-159.

Schlanger, B.B. & Galanowsky, G.I. Auditory discrimination tasks
    performed by mentally retarded and normal children. Journal of
    Speech and Hearing Research, 1966, 9, 434-440.

Sherman, D. & Geith, A. Speech sound discrimination and articulation
    skill. Journal of Speech and Hearing Research, 1967, 10, 277-280.

Templin, M. A study of sound discrimination ability of elementary
    school pupils. Journal of Speech Disorders, 1943, 8, 129-132.

Templin, M. Sound Discrimination Test. Minneapolis: University of
    Minnesota Press, 1957.

Thompson, B.B. A longitudinal study of auditory discrimination.
    The Journal of Educational Research, 1963, 56, 376-378.

Travis, L.E. & Rasmus, B. The speech sound discrimination ability of cases with functional disorders of articulation. Quarterly Journal of Speech, 1931, 17, 217-226.

Turiads, D., Wepman, J.M., & Morency, A. A perceptual test battery: development and standardization. The Elementary School Journal, 1972, 72, 351-361.

Vellutino, F.R., DeSotto, L., & Steger, J.A. Categorical judgment and the Wepman test of auditory discrimination. Journal of Speech and Hearing Disorders, 1972, 37, 252-257.

Wepman, J.M. Auditory discrimination, speech, and reading. The Elementary School Journal, 1960, 60, 325-333.

Wepman, J.M. Auditory Discrimination Test. Chicago, Illinois: Joseph M. Wepman, Ph.D., 1958.

Wepman, J.M. Auditory Discrimination Test. Chicago, Illinois: Joseph M. Wepman, Ph.D., 1973.

Winitz, H. & Bellerose, B. Sound discrimination as a function of pretraining conditions. Journal of Speech and Hearing Research, 1962, 5, 340-348.

Winitz, H. & Lawrence, M. Children's articulation and sound learning ability. Journal of Speech and Hearing Research, 1961, 4, 259-268.

Zedler, E.Y. Effect of phonic training on speech sound discrimination and spelling performance. Journal of Speech and Hearing Disorders, 1956, 21, 245-249.

APPENDIX A

Description of the Three Speech Sound Discrimination Tests

Goldman, R., Fristoe, M., & Woodcock, R. Test of Auditory Discrimi-
nation. Circle Pines, Minnesota: American Guidance Service, Inc.,
1970. (Cost: $22.00)

The Test of Auditory Discrimination was designed to provide
measures of speech sound discrimination ability relatively unconfounded
by other factors. The test consists of three sections; a training
procedure, a quiet subtest, and a noise subtest.

The training procedure was designed as an attempt to familiarize
the subject with each of the word-picture associations to be used in the
following subtests. The second section of the test provides a measure
of speech sound discrimination in the absence of background noise. The
stimuli used are single words which are depicted on a single plate for
each test item containing four pictures, one of which corresponds to the
word presented. The stimuli are presented through a prepared audio tape
designated to be played at a comfortable loudness level for the subject.
There are a total of 30 plates containing four pictures each in this
section.

The noise subtest provides a measure of speech sound discrimination
in the presence of background noise. Prior to test items there are
three stimuli presented having an increasing level of background noise
in an attempt to help the subject adapt to the new and more difficult
listening conditions. Instructions and test stimuli are also presented
through an audio tape. Thirty plates containing four pictures each
comprise this section of the test also. The four words on the 30 plates

in each section differ in a single sound only. The tape may be presented through earphones or sound field provided the testing conditions are quiet.

The record form of the test is divided into the training procedure, the quiet subtest, and the noise subtest. Blanks are provided for the examiner to record the number (1,2,3,4) of the subject's selection. The test words are printed on the test form followed by a blank for the subject's responses and the number of the correct response in parentheses. Errors are scored after test administration is completed by marking a slash through the number printed on the Correct Response column.

Wepman, J.M. Auditory Discrimination Test. Chicago, Illinois:
Joseph M. Wepman, Ph.D., 1973. (Cost: $8.00)

The Auditory Discrimination Test was designed to identify subjects
who are unable to recognize fine differences between sounds in English
speech. The task involved in the test requires the subject to listen to
pairs of words read by the examiner and indicate whether the words heard
were the same or different. The selection of words used in the test
were chosen from Thorndike-Lorge Teacher's Word Book of 30,000 Words and
matched on familiarity and frequency of occurrence. Four vowel pair
discriminations, 26 consonant pair discriminations (13 initial and
13 final consonants) and ten false choices (identical word pairs)
comprise the test. Words paired are matched for phonetic category,
equated for length and differ in a single sound only. A short training
procedure is scheduled prior to test administration.

The score sheet lists the 40 word pairs used as test stimuli.
Beside each word pair are two columns designated SAME and DIFFERENT.
The examiner scores responses by marking a plus or minus in the
"different" column for contrasting word pairs depending on the subject's
responses. Identical word pairs are scored with a plus or minus in the
"same" column according to the subject's responses.

The total numerical score earned by the subject can be compared to
a five point scale ranging from "very good development" to "below the
threshold of adequacy."

Mecham, M.J., Jex, J.L., & Jones, J.D. Test of Listening Accuracy in
    Children. Salt Lake City, Utah: Communication Research Associates,
    Inc., 1973. (Cost: $20.00)

The Test of Listening Accuracy in Children was derived from a

previously published test by the same authors: the Picture Sound

Discrimination Test. Test items on the TLAC were selected from 132

words used as the auditory stimuli in a listening test that was given to

40 mentally defective children. Items which elicited more than 85%

correct responses and those which elicited less than 15% correct

responses were omitted and the 86 items which remained comprise the TLAC.

The test stimuli are presented from a prepared audio-recording;

earphones are not to be used. The tape begins with the instructions

that a man will say three words but only one of the words corresponds to

one of three pictures on a test plate. The other two pictures are named

incorrectly; the subject is instructed to point to the picture which is

named correctly. Six practice items presented from the audio-tape are

administered before the test begins. The acoustic stimuli were recorded

in a background of white noise to increase the sensitivity of the test

and insure a wider dispersion of scores. No provisions for quiet

testing are included.

The score sheet for the TLAC consists of the three stimulus words

per test item preceded by a blank for scoring the subject's response and

the correct letter (A,B,C) of the test item in parenthesis. Instructions

for scoring are simply a plus for correct selections and a minus for

incorrect responses. The total score can be used to determine a percen-

tile ranking according to three age ranges; kindergarten, first, and

second grades. The percentile can be compared to a rating including supe-

rior, excellent, normal, inferior, and very poor discrimination ability.

APPENDIX B

Table 3

Comparison of Characteristics on the
GFW, TLAC and Wepman

	Wepman	TLAC	GFW
1. Type of stimuli presentation:	quiet test	noise test	quiet and noise subtests
2. Method of stimuli presentation:	oral	audio-recording (male)	audio-recording (female)
3. Stimuli:	auditory	auditory and visual	auditory and visual
4. Word characteristics:	phonetically balanced	not phonetically balanced	phonetically balanced
5. Words equated for length:	yes	no	yes
6. Word length:	1 and 2 syllable words	1, 2, and 3 syllable words	1 syllable words
7. Discrimination task:	between 2 words	among 3 words	among 4 words
8. Number of practice items:	5	6	3 (only on noise subtest)
9. Number of test items:	40	86	60
10. Response:	oral	pointing	pointing
11. Vocabulary training:	not available	not available	64 item task
12. Retest:	40 different word pairs	not available	not available
13. Normative data:	5-8 years	kindergarten, first and second grades	3 years, 8 months to above 70 years

APPENDIX C

Table 4

Percentage of Items Responded to Correctly
on the GFW, TLAC and Wepman

Subjects	TLAC (noise)	GFW (noise)	GFW (quiet)	Wepman (quiet)
1	63/86=73%	16/30-53%	28/30=93%	38/40=95%
2	39/86=45%	10/30=33%	21/30=70%	22/40=55%
3	45/86=52%	7/30=23%	11/30=37%	25/40=63%
4	47/86=55%	10/30=33%	24/30=79%	33/40=83%
5	63/86=73%	15/30=50%	29/30=97%	39/40=98%
6	52/86=60%	9/30=30%	22/30=73%	25/40=63%
7	62/86=72%	18/30=60%	28/30=93%	39/40=98%
8	47/86=55%	10/30=33%	22/30=73%	31/40=78%
9	61/86=71%	14/30=47%	26/30=87%	35/40=88%
10	47/86=55%	13/30=43%	23/30=77%	31/40=78%
11	57/86=66%	15/30=50%	21/30=70%	34/40=85%
12	62/86=72%	18/30=60%	27/30=90%	38/40=95%
13	37/86=43%	11/30=37%	23/30=77%	35/40=88%
14	40/86=47%	10/30=33%	21/30=70%	26/40=65%
15	51/86=59%	11/30=37%	22/30=73%	27/40=68%
16	60/86=70%	12/30=40%	28/30=93%	39/40=98%
17	66/86=77%	14/30=47%	23/30=77%	29/40=73%
18	46/86=53%	13/30=43%	27/30=90%	37/40=93%

APPENDIX D

A Review of Related Research

A significant amount of research has been conducted utilizing Wepman's Auditory Discrimination Test (1958; revised 1970). Wepman (1960) has conducted numerous experiments relating speech sound discrimination to other variables. His data indicated a positive relation between low reading achievement, functional articulatory defects, and poor speech sound discrimination in a study conducted with second grade children. He tested a group of first and second grade children with the Auditory Discrimination Test and concluded that there was a significant relationship between poor articulation and poor speech sound discrimination and a less definite but significant relationship between poor reading ability and poor speech sound discrimination. He also hypothesized that there was a relationship between speech sound discrimination and intelligence.

Wepman has stated that adequate speech sound discrimination is not fully achieved until eight years of age although the degree of competence varies with individuals and some children acquire the skill much earlier. He indicated that remedial work regarding sound discrimination should not begin until after eight years of age or third grade.

The results of Schiefelbusch and Lindsey's 1958 study further supported the evidence relating poor speech sound discrimination and poor articulation. First and second grade students with normal intelligibility and a matched group with articulatory defects had significant differences in sound discrimination abilities. The maturational

component of sound discrimination ability was also reported. The authors found the second grade normal speakers to have better speech sound discrimination than the first grade normals, however, the second grade speech defective group did not exhibit significantly higher discrimination scores than the first grade speech defective children.

Prins (1963) conducted a study with children with functional articulatory defects and a control group of normal speaking first grade students utilizing the Wepman Auditory Discrimination Test. He concluded that speech sound discrimination is a function of articulation and the total learning process. An interesting finding of the study was that the children who confused place of articulation during speech sound discrimination also experienced difficulty in discriminating minimal word pairs in which one sound was altered in place of articulation.

Christine and Christine (1964) conducted an experiment with a group of retarded children, a group of children with articulatory defects and a control group to investigate the relationship between oral language and reading. The study tested the hypothesis suggested by numerous researchers that poor speech sound discrimination is causative of poor articulation and reading retardation in children. The Auditory Discrimination Test was administered to all subjects. The data collected suggested the conclusion that poor speech sound discrimination is one etiological factor of functional articulation problems and reading retardation.

The relationship between language comprehension and speech sound discrimination was investigated by Marquardt and Saxman (1972) in kindergarten children with numerous misarticulations and kindergarten children

with normal articulation. The Auditory Discrimination Test was administered to the children. Results of the study showed that the articulatory defective group performed deficiently on the discrimination test in comparison with the articulatory proficient group.

Sherman and Geith (1967) reversed the traditional method of assessing the relationship between speech sound discrimination and articulation ability. They administered the Templin Sound Discrimination Test (1957) to 529 children. The 18 children with the lowest scores on the discrimination test and the 18 children with the highest scores were then administered the Templin-Darley Test of Articulation. Significantly higher scores were achieved on the articulation test by the children superior in speech sound discrimination skill.

Travis and Rasmus (1931) analyzed speech sound discrimination in a group of university freshmen and elementary school students. Their results indicated that the students in both age groups with articulation errors tended to perform poorer in their ability to discriminate sounds.

Ewing (1930) tested children with congenital aphasia and concluded that the high frequency losses which many of the children exhibited prevented accurate discrimination of speech sounds at normal conversational intensity. He considered the loss responsible for their failure to develop speech.

Anderson (1949) found that subjects had a greater number of discrimination errors on the /s/ sound in contexts in which they misarticulated /s/ than in contexts in which they did not. Based on such evidence Winitz and Bellerose (1962) suggested that specific types of sound discrimination may be a function of specific articulation

errors. Rather than poor speech sound discrimination subsequently affecting articulation proficiency, they hypothesized that the learning of an incorrect phonetic response may affect discrimination between the correct sound and the incorrectly learned sound.

The Templin modification (1943) of the Travis-Rasmus Speech Sound Discrimination Test (1931) was used by Zedler (1956) in a study designed to evaluate the effect of phonic training on speech sound discrimination and written spelling. The results warranted the conclusions that written spelling and speech sound discrimination abilities are significantly related and speech sound discrimination increased significantly with phonic training.

Schlanger and Galanowsky (1966) administered the Templin Speech Sound Discrimination Test (1957), the Boston University Speech Sound Discrimination Test (1953), the Wepman Auditory Discrimination Test (1958), and the Nonsense Syllable Sound Discrimination Test to normal and mentally defective children matched for mental age. The normal children scored significantly better on all tests. The discrimination tests were significantly correlated with articulation in the retarded group, however, in the normal group only the Nonsense Syllable Test had significant, yet low correlations with articulation.

Mecham (1955) using his Picture Speech Sound Discrimination Test assessed the speech sound discrimination ability of mentally defective children. Following speech therapy the subjects improved in articulation, speech sound discrimination, auditory memory span, and average sentence length. The improvement was independent of IQ, however, the improvement in speech sound discrimination was not independent of chronological age.

Winitz and Bellerose (1962) concluded that speech sound discrimination tests may measure the learned equivalence and distinctiveness of speech sounds as well as developmental skills.

Thompson's (1963) data agreed with that of Wepman (1960). Based on results of her sample of second grade children she concluded that speech sound discrimination and intelligence are highly correlated with good reading abilities. She also agreed essentially with Wepman on the maturation of speech sound discrimination skills. Twenty-four percent of her second grade sample had inadequate speech sound discrimination. The percentage was much higher during the first grade.

In an experiment with mentally retarded children Gruber and Steer (1965) utilized the Templin Speech Sound Discrimination Test (1957). The results of the study indicated that there were no differences in speech sound discrimination abilities between retardates with adequate articulation and those with a low articulation index. No significant sex differences were found in the sample regarding speech sound discrimination abilities.

Winitz and Lawrence (1961) investigated the relationship between articulation and sound learning ability in 12 kindergarten children with good articulation and 12 children with poor articulation. Each child was asked to imitate three non-English sounds. No differences were found between the groups in sound learning ability.

No relationship between articulation proficiency and speech sound discrimination was found by Barnes (1932) in a group of university freshmen contradicting the data in the Travis and Rasmus study. He found no differences between the speech defective group and the superior

speakers on the Travis-Rasmus Speech Sound Discrimination Test. A study conducted by Hall (1938) utilized the same speech sound discrimination test. She found no significant differences between functional speech defectives and normal speakers in their ability to discriminate between pairs of speech sounds. No relationship was found between the ability to discriminate and recognize the sounds in a complex auditory pattern and articulation proficiency.

A COMPARISON OF THREE SPEECH SOUND
DISCRIMINATION TESTS

by

ANNE LOUISE AHLERS

B.A., Creighton University, 1974

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

Three speech sound discrimination tests were compared by adminis-
tering each of them to eighteen 3, 4, and 5 year-old preschool children.
The tests evaluated were the Test of Listening Accuracy in Children by
Mecham, Jex, and Jones, the Goldman-Fristoe-Woodcock Test of Auditory
Discrimination, and the Auditory Discrimination Test by Wepman.

The percent correct and the mean number of correct responses on
each test were determined. The results of the TLAC (noise) and the
GFW (noise) were compared while the GFW (quiet) test results were
compared with those of the Wepman (quiet). The GFW was analyzed
according to two additional procedures described in the test manual.
The tests were also evaluated on the type and method of stimulus presen-
tation, practice items, vocabulary training, response type, retest
availability, scoring methods, discrimination task, word characteristics,
number of test items, administration time, reliability, validity, and
normative data. Advantages and disadvantages of each test were
discussed. The results of the comparisons indicated marked differences
between the tests.