University of Massachusetts Amherst

ScholarWorks@UMass Amherst

1960

A speech intelligibility test for young deaf children.

Bill G. Blevins
University of Massachusetts Amherst

A SPEECH INTELLIGIBILITY TEST
FOR YOUNG DEAF CHILDREN

BLEVINS - 1960

A SPEECH INTELLIGIBILITY TEST
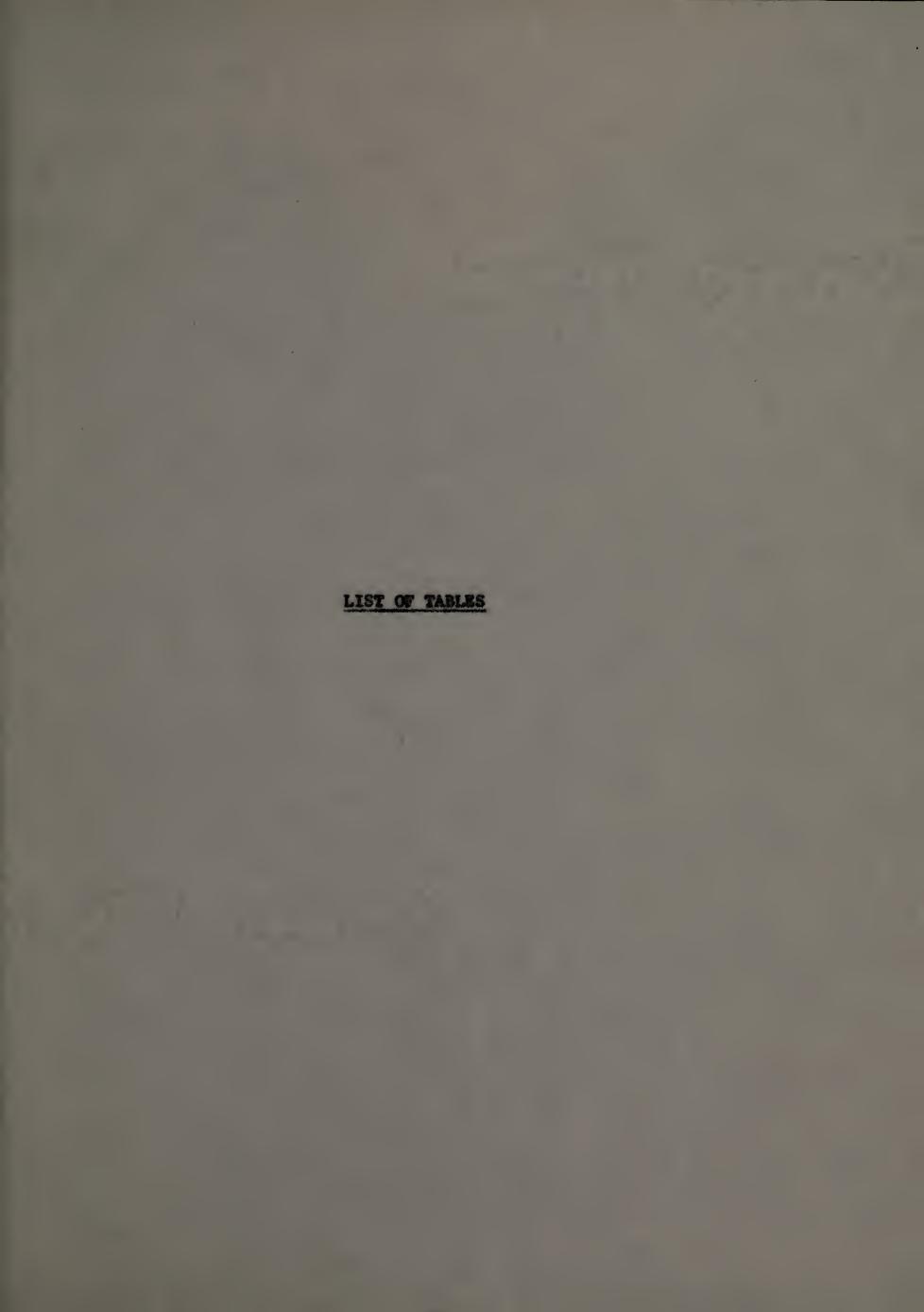FOR YOUNG DEAF CHILDREN

by
Billy G. Blevins

A problem presented in partial fulfillment
of the requirements for the
Master of Education Degree
School of Education
University of Massachusetts
1960

TABLE OF CONTENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF TABLES

CHAPTER I

INTRODUCTION

CHAPTER I

INTRODUCTION

This paper describes an attempt to find a reliable means for characterizing the speech of young deaf children in terms of the essential qualities that render it a functional tool of communication. The problem was to determine with what degree of consistency it is possible for a group of listeners to assess the speech of young deaf children for its voice quality, inflection, rhythm, phrasing, fluency, precision of articulation, and intelligibility.

It is important in gauging the success of teaching methods to be able to evaluate the speech of deaf children. Also, school officials are concerned with the value of different types of hearing aids employed in this instruction. At present, we are not aware of a satisfactory way of determining the speech intelligibility of young deaf children. It is the primary objective of this paper to develop a test that can be used in evaluating the speech methods and equipment employed in the education of deaf children.

Very little work has been done along the lines of this problem. Bjuggren[1] has been working on a test in Sweden that is quite similar to the one described in this study. As yet, only a preliminary report has been published. Few other studies have been made. There are several methods available to test the speech of older deaf children who have acquired some reading skill and sufficient language to enable them to take a test of a different nature.

---

[1]Gunnar Bjuggren, "A Method To Test The Intelligibility of The Speech of Pre-School Children With Severe Hearing Impairment. A Preliminary Report." Stockholm: Acta-Oto-Laryngologica - Supplementum 110. 1954. pp. 83-88.

The test being reported in this paper involved the use of pictures to elicit samples of the child's spontaneous speech. The samples were tape-recorded and later appraised by a group of listeners who were qualified to rate the speech according to the seven categories mentioned previously. It was assumed that a group of listeners would be much more reliable than one or two individuals and that they could agree on the rating of a child's speech.

At the Clarke School for the Deaf, where this study was conducted, the facilities needed for the development of a test of this nature were available. Schools without a research department or a person trained to conduct such tests should consider the time involved and the number of listeners needed before undertaking to use this test. It would be possible, for schools without recording equipment, to use this test by having the listeners present at the testing session.

CHAPTER II

REVIEW OF LITERATURE

CHAPTER II
REVIEW OF LITERATURE

In the past, relatively few studies have been made concerning the testing of the speech of young deaf children. While there have been several works done on testing older deaf pupils, the test suitable for young children hasn't been found. The tests designed for older children involve reading and young deaf children don't read well enough for these tests to be useful in testing them.

Gunnar Bjuggren[1] conducted a study to find an intelligibility test that does not depend upon the ability to read. Only a preliminary report has been published on this study. In the report, the author used sixteen subjects--all deaf or hard of hearing young children. The author got spontaneous speech from the children by having them talk about pictures that were shown to them. Two listeners took down the speech as it was spoken, or used tape recordings when necessary. They attempted to determine the number of words and sentences spoken and the amount of speech understood and thus give a per cent intelligibility score using the formula:

$$\text{Intelligibility} = \frac{\text{Sentences (or words) apprehensible}}{\text{Sentences (or words) intelligible}} \times 100\%$$

Preliminary results showed:

1) The general trend shown by the results was that the better the vocabulary used, the more intelligible the speech became.

2) There was an almost direct relationship between the vocabulary and the length of the sentence.

---

[1]Gunnar Bjuggren, "A Method To Test The Intelligibility of The Speech of Pre-School Children With Severe Hearing Impairment. A Preliminary Report," Stockholm: Acta-Oto-Laryngologica - Supplementum 110. 1954. pp. 83-88.

3) The results indicated that the degree of hearing loss had an effect
on speech intelligibility. Those with severe impairment fell
under the 30 per cent level of intelligibility.

4) The greater the facility for using language, the greater the
chances of understanding the speech. That is, the longer the
sentences, the better the vocabulary and the better the intelligi-
bility rating.

5) The results indicate that several factors influence the intelligi-
bility of speech.

Hudgins[2] described a method of testing speech intelligibility that

he thinks is simple, easy to administer and to score, and at the same

time has proved successful and efficient. This method involves using

speakers (those to be tested), test materials (phonetically balanced

familiar word lists (PBF), and a jury of listeners or judges to score

the speech. Each speaker read a 50-word list of PBF words that were

recorded. They were taught the unfamiliar words before the test. Using

a recorder eliminates any undue tension in the test situation that might

be created by having the listeners present at the time of testing.

Recordings also allow the material to be available for further leisure

use.

The author states that the listeners or judges should be a com-

promise between skilled experienced listeners and strangers to the speech

of the deaf. College juniors were selected for this job and trained to

the problems involved in interpreting the speech of the deaf.

All the children in the school, except the primary classes, were

tested in this manner. The test seemed to be worthwhile and valuable.

It was found that there was a high correlation between the results when

[2]C. V. Hudgins, "A Method of Appraising the Speech of the Deaf."
The Volta Review, Vol. 51 (Dec. 1949) pp. 597-601.

word lists were used and the results when sentences were used as test materials. This method is currently in use at the Clarke School in the routine testing program.

Farman and Phillips[3] constructed a test as a means of formally evaluating the speech of the pupils in the Oregon State School for the Deaf. Their primary objectives were to determine:

1) Which students have intelligible or unintelligible speech.

2) What factors, and to what degree, lead to speech intelligibility or unintelligibility.

3) What specific pathological defects each individual might possess that affect his or her speech.

4) Where weaknesses in the speech teaching program might be.

5) An individual's ability to produce the various units of speech, i.e., to compare his ability to produce an isolated sound and a word; a word and a sentence.

The test was divided into the following five parts in order to thoroughly evaluate and diagnose each student's speech:

1) A test of isolated sounds (including all of the voiced and selected unvoiced sounds).

2) PB words - a list of twenty phonetically balanced words selected from a total of 200 PB words.

3) Sponded words - a list of twenty words selected from a total of 200 spondee words.

4) Unassorted words - a list of twenty unassorted words selected from a total of 100 words devised by the test authors and assumed to be familiar to the students.

5) Sentences - a list of twenty sentences selected from a total of 200 sentences devised by the test authors.

_____

[3]Jay J. Farman and Betty Phillips, "The Farman-Phillips Speech Intelligibility Diagnostic Test." The Volta Review, Vol. 56 (Apr. 1954) pp. 168-170.

For the first part, two lists of isolated sounds were read by the pupil. Ratings of one to three; representing poor, fair, and good were assigned for each effort. The other parts of the test were scored in a similar manner. The other parts; the PB words, the spondees, the unassorted words and the sentences, were rated on the basis of six attributes which influence speech, namely; pitch, loudness, vocal quality, articulation, pronunciation, and rhythm. A rating scale of one to five, one being the lowest and five the highest rating, was used. Space was provided for comments concerning the speech by the judges. To find the intelligibility score, the rank scores were converted into percentages for all categories and averaged. Three auditors were used and the mean was taken as the rank score.

The test seems to be a fairly adequate method of analyzing the speech of deaf boys and girls. The test has diagnostic value, and its usefulness is increased, to some degree, by the remarks recorded on the score sheet by the auditors. The test was used for students in the fourth to the tenth grades.

Templin[4] constructed a short non-diagnostic test which would be a satisfactory measure of achievement in speech sound articulation. The test applies when the chief purpose of testing is to screen acceptable from unacceptable speech.

Several speech samples were analyzed to determine a discrimination value for each sound. The fifty sound elements which were found to be

[4]Mildred C. Templin, "A Non-Diagnostic Articulation Test." The Journal of Speech Disorders, Vol. 21 (Dec. 1947) pp. 392-396.

most discriminating were selected as the items to be used in this non-diagnostic test.

The fifty sounds were presented in both word and sentence tests in order to compare the results of the same sounds under two conditions. Word tests could be administered more quickly but sentences were included since it was believed that a sentence test might be more interesting to some of the children. The word test presented the fifty sound elements in forty words and the sentence test presented them in nineteen sentences. The score is the number of sounds that are articulated correctly.

Three groups of twenty-two, fifty-seven, and twenty respectively, were given the tests. Ages of the subjects varied from two years to eight years.

In the administration of the word test the children repeated the word after the examiner or responded to a picture - whichever method seemed best adapted to the particular child. The author felt that the same results would be obtained with either method. Pictures were used more frequently with the younger children. In the administration of the sentence test, the children repeated the sentences after the examiner.

The usefulness of a test is defined in terms of its validity and reliability. The reliability of the non-diagnostic test is determined by (1) the correlation of the word and sentence tests given at the same test session, (2) the test-retest correlations of the tests, and (3) a comparison of the means on the word and sentence tests.

The coefficients of reliability as computed from the raw test

scores were very high, being above .93 in all cases. The test-retest coefficients were above .97 for the total group. The reliability of the word and sentence tests was about equal. The use of the sentences is questionable with the two-year old group due to memory span factors. In no instance were the differences in mean scores between any of the measures at a single age or for a total group statistically significant.

In another work, Templin[5] sought to find out if in eliciting speech sounds for the test material, the test words or phrases should be elicited spontaneously or repeated in imitation of a pattern provided by the examiner. The author wanted to find out:

1) If there is any difference in measured articulation when a sound is tested in a word spontaneously uttered or in a word repeated after the examiner.

2) If there is any difference in measured articulation when the same sound is tested in different words.

One hundred children, presumably with normal hearing, ages two years to six years were included in the study. Three tests were administered at six-month intervals. The picture articulation test constructed for this study consisted of seventy-three words; each illustrated by a picture, and measured 113 sound elements in the initial, medial, and final positions.

Three measures of the child's articulation were obtained at each test session. (1) The first measure, designated as the Picture Test, is his articulation of specific sounds in words uttered spontaneously

_____

[5]Mildred C. Templin, "Spontaneous Versus Imitated Verbalization In Testing Articulation In Preschool Children." The Journal of Speech Disorders, Vol. (Sept. 1947) pp. 293-300.

in response to the pictures. (2) The second measure, Aural Test A, is his response when he has the word pattern of the examiner to imitate and the picture is still before him. (3) The third measure, Aural Test B, made at the end of the test session, is his repetition of each of the words after the examiner, with the picture no longer before him.

Instructions for the test were, "I am going to show you some pictures and I want you to tell me what the pictures are. After you tell me the name of the picture, I'm going to say it again. Then you say it right after me so I am sure to hear it just right." For Aural Test B, the children were asked again to repeat the words so that the examiner would be certain she had heard exactly what was said. One examiner did all the testing.

The results indicated that there was very little difference between the spontaneous and imitated vocalizations, with or without the pictures. Similar results for specific sounds were obtained when the same sound was presented in different words.

The methods employed in the two studies reported on above are not applicable to the young deaf child. While the deaf child may name a picture without difficulty, if he has been taught the name of the object, the task of imitating and repeating the word or phrase after the examiner is practically impossible, at least with speech that could be judged as belonging to the child. Thus, it would be practically worthless as a sample to be used for evaluating purposes.

Perrin[6] made a study of the rating of defective speech by trained and untrained observers. She wanted to determine if an opinion of speech by a lay person was valid.

Two groups of students served as raters. An untrained group which had had no courses in speech therapy and a trained group composed of graduate majors in speech therapy and speech correction.

Disc recordings of defective speech were analyzed. The results showed (1) that trained and untrained judges do not differ significantly in their evaluations of functional articulation defects, (2) both groups showed a significant amount of agreement within their respective groups in their rankings, (3) there were many inconsistencies in rankings shown by members of both groups, and (4) the correlation coefficient between the number of sounds misarticulated and the judges' rankings was significant at the four per cent level for the untrained judges and significant at better than the one per cent level for the trained judges.

The present study utilized some of those methods and findings described in the above studies that are best adapted to the use of young deaf children. The material of the study consisted of speech samples of deaf children elicited spontaneously by the use of pictures.

It was the purpose of this study to determine the degree of reliability with which a group of listeners could evaluate these samples.

---

[6] Elinor H. Perrin, "The Rating Of Defective Speech By Trained and Untrained Observers." The Journal of Speech and Hearing Disorders, Vol. 19 (March 1954) pp. 48-52.

CHAPTER III

DESIGN OF THE STUDY

CHAPTER III

DESIGN OF THE STUDY

A simple speech intelligibility test that would be appropriate for young deaf children was sought. Since young deaf children have very little language and their reading ability has not yet been developed, it was decided that spontaneous speech evoked by presenting simple pictures to the children would provide the best speech samples. Satisfactory speech samples can be obtained from older deaf children by having them read selected word-lists and sentence materials.

Subjects--Pupils Tested -- Two types of subjects were involved in the study. The pupils to be tested for speech intelligibility were fifty pupils at the Clarke School for the Deaf. The pupils ranged in age from six years, nine months to eleven years, seven months. There were twenty-four boys and twenty-six girls in the group. This group consisted of the youngest class in the Middle School and all of the pupils in the Lower School except those in the preparatory classes. The children were asked to talk about a set of pictures and their speech was recorded for later evaluation by the group of listeners. Generally, the children in the preparatory classes do not have enough language to be given a test of this type.

Listeners To Appraise The Speech -- After the speech samples were obtained, the task of determining whether the speech could be evaluated and scored with a fair amount of consistency was approached. It was decided that a panel of judges or listeners would be used for this purpose. A group of six listeners was used to rate the speech. This group was composed of two men and four woman who were familiar with the speech of the deaf. The background of the listeners ranged in experience

from very skilled to beginning teachers of the deaf. The group was a compromise between skilled workers with the deaf and people unfamiliar with their speech. A group with similar backgrounds to the one used in this study could be found in most schools for the deaf.

Method of Testing Pupils -- Pictures were used to elicit spontaneous speech samples from the children. A test consisted of ten pictures selected at random from a group of thirty-five pictures. The pupil was told to talk about each picture while the speech was recorded on magnetic tape.

Experience with picture tests indicates that they are not ideal for all pupils--some are stimulated by pictures and speak easily while others have difficulty and are inclined to merely name objects in the pictures. Older pupils seem willing to talk, while younger ones are not so voluble. There is also a wide range of individual differences with respect to spontaneity.

The routine for recording the speech samples was as follows: As the recorder started, the pupil's full name, the date, and an identifying test number for the particular group of pictures to be used was announced. Then the pupil's name and the number of the picture to be talked about was given. The pupil then began to talk about the picture. Each picture was spaced on the record to allow an adequate interval for the listeners to score the sample. The child's teacher or the teacher-in-charge of the department conducted the test. When the child hesitated, she silently pointed to objects or relationships in the pictures in an effort to elicit speech. A trained person was in the test room to handle

the recording equipment and to supervise the testing. This procedure
was followed for the ten pictures.

Method Of Appraising The Speech -- The tape recorded speech samples
were played back to the listeners by high-fidelity equipment through
pairs of dynamic headphones. A rating sheet, shown in Figure 1, was
devised in order to aid the listeners in evaluating the speech samples.
The rating sheet contained seven rating categories to be scored on a
scale ranging from normal to very poor. Space was provided for the
pupil's name, the picture test number, the date, and the name of the
listener who scored the test.

The listeners were given a practice period to discuss the categories
and the rating scales and to compare their own ratings of practice
samples of speech with those of the group. The seven categories were:
voice quality, pitch-inflection, rhythm, phrasing, fluency, precision of
articulation and intelligibility.

Explanation of the Categories -- Voice quality was defined as the
pleasantness of the voice to our ears and its similarity or comparison
to "normal" voices. The rating scale for voice quality was a five-point
scale with the ratings of (1) normal, (2) good, (3) fair, (4) poor, and
(5) very poor. This takes into account such things as nasality, high or
low pitched voices, and "breathy" voices.

Inflection was defined as changes of pitch or the rise and fall of
the voice indicating expressive meanings given to a word or phrase. The
rating scale for this category was a five-point scale rating (1) natural,

FIGURE I

THE RATING SHEET

Speech Intelligibility Picture Test

Name_____   Scorer_____

Test No._____   Date_____

Categories

1. **Voice quality**
   1. Natural
   2. Good
   3. Fair
   4. Poor
   5. Very poor

1	2	3	4	5

2. **Inflection**
   1. Natural
   2. Good
   3. Fair
   4. Poor
   5. None

1	2	3	4	5

3. **Rhythm**
   1. Normal
   2. Near Normal
   3. Abnormal
   4. Non-Rhythmical

1	2	3	4

4. **Phrasing**
   1. Natural
   2. Good
   3. Poor
   4. Single Word

1	2	3	4

5. **Fluency**
   1. Normal
   2. Good
   3. Fair
   4. Very slow

1	2	3	4

6. **Precision of Articulation**
   1. Very high
   2. High
   3. Medium
   4. Low
   5. Very low

1	2	3	4	5

7. **Intelligibility**
   1. Normal
   2. High
   3. Fair
   4. Low
   5. Jargon

1	2	3	4	5

(2) good, (3) fair, (4) poor, and (5) none. Natural inflection indicated the speech was similar to that of a normal person, while "none" indicated utterances of speech that were a complete monotone.

Rhythm was defined as the way the syllables were grouped in unit groups with accents properly placed. This was rated on a four-point scale rating (1) normal, (2) near normal, (3) abnormal, and (4) non-rhythmical. In this case, abnormal would mean grouping syllables in an abnormal manner, such as "MoTHER went (pause) home." A non-rhythmical rating would indicate that speech was uttered in a monotonous manner as "Mother-went-to-the-store."

Phrasing was defined as the way the speaker controlled or used his breath. That is, did the speaker say one word per breath, two words per breath, or did he utter the number or words in a breath that a "normal" speaker would? The rating scale for phrasing rated the speaker as (1) normal, (2) good (meaning good use of breath), (3) poor, and (4) single word (meaning that the speaker took a breath before each word).

Fluency was defined as the rate of utterance. The scale rated (1) normal rate of speech, (2) good, (3) fair, and (4) very slow.

Precision of articulation was defined as the ability to put the speech components together to make syllables, words, or phrases. Or more simply, the quality of the enunciation of the speech. This category takes into consideration the control and use of the articulatory organs and the control of pressures. The rating scale was a five-point scale, rating the speakers as (1) very high, (2) high, (3) medium, (4) low, or

(5) very low.

Intelligibility was defined as the general ease with which the listener could understand the speech. The child was rated on his ability to make himself understood. The rating scale for general intelligibility ranged from normal to "jargon" on a five-point scale. The scale was (1) normal, (2) high, (3) fair, (4) low, and (5) jargon (completely unintelligible).

A Typical Listening Session -- After an adequate explanation of the job to be done and upon the completion of a practice session to prepare them for the job, the listeners rated the speech samples of the children.

The group assembled in a classroom after the school day and listened to the speech of approximately eight children per hour. Each listening session was approximately 1½ hours.

The tape recorded speech samples were played back through the group hearing aid equipment available in the classroom. The listeners used individual headphones instead of a loud speaker since the headphones reduced outside noises and helped the listener concentrate on the speech. As the recorded speech was replayed, the corresponding picture was placed in view of the listeners. The picture helped make the speech more intelligible since tape-recording the speech had entirely eliminated the lipreading factor. Complete unfamiliarity with the topic made the speech samples of the poorer speakers extremely difficult to interpret. Thus, knowing the topic helped make up for the absence of facial

expressions and other clues.

Method of Scoring the Test -- To determine an over-all score for each child, the listeners ratings were totaled for the seven categories and the average score for the six listeners was used as the child's over-all speech score. The range for the test was twenty-five. The best possible score was seven and the poorest score possible was thirty-two. Hence, the smaller the child's score the higher his intelligibility. A careful study of the scores for each category would make this score more meaningful, however.

The Tests For Reliability -- In order to determine the reliability of the test, the speech samples of each of the fifty children were rated twice. The second ratings were made after a waiting period of approximately six weeks. This made it almost impossible to remember a rating from the first time until the second. The total time involved to rate the speech of fifty children once was approximately seven hours.

Coefficients of correlation between the various scores obtained from Test I and Test II were determined. These correlations were found for the pupils' scores on each rating category and for the pupils' over-all test scores. The scores given by the group of listeners and the scores given by individual listeners were also correlated. A high correlation between the two tests should indicate a high degree of reliability among the listeners in rating the speech.

An attempt was made to determine the relative importance of each

rating category on the test rating sheet. The ratings of each category were correlated with the rating of the intelligibility category. These inter-category correlations should indicate the relative importance of each category in determining intelligibility.

CHAPTER IV

PRESENTATION AND ANALYSIS OF DATA

CHAPTER IV

PRESENTATION AND ANALYSIS OF DATA

According to Best[1] a test is said to be valid if it measures
what it is supposed to measure, or if it proves to be useful in
accomplishing its desired purpose. A test is considered reliable if
it measures accurately and consistently, yielding comparable results
when administered a number of times. The degree of reliability may be
established by correlating the results when the same individuals take
duplicate or equivalent forms of the test. The degree of reliability
may also be established by correlating the scores on two or more
successive administrations of the same test, putting the scores on the
first administration of the test against scores made by the same students
on a repeat performance. The latter method was used in this study.

In order to develop this test, it was necessary to determine the
reliability of a group of listeners in evaluating the speech samples
of young deaf children. The tape-recorded speech samples were rated
by a group of six listeners who were familiar with the speech of the
deaf. After a time lapse of approximately six weeks, this same group
of listeners rated the same speech samples a second time, though not in
the same order as the first. The results indicate that this method of
testing and scoring the speech of young deaf children is a reliable one.

Comparison of the Over-all Speech Scores for Test I and Test II --
The over-all speech scores assigned to each pupil by the listeners on
Test I were correlated with those assigned on Test II. The scores of

---

[1]John W. Best, Research In Education (Englewood Cliffs, N.J.:
Prentice-Hall, Inc., 1959). pp. 242-243.

all fifty pupils were used. The coefficients of correlation between the scores on these tests was .89. Table 1 shows the over-all speech scores given on the first and second test.

TABLE 1

OVER-ALL SPEECH SCORES OBTAINED BY PUPILS ON THE FIRST AND SECOND TEST.

Pupil	Test 1	Test 2	Pupil	Test 1	Test 2	Pupil	Test 1	Test 2
J A	19	17	J F	25	23	RRP	27	25
S A	25	25	S F	24	21	S P	21	21
G B	24	22	L F	26	23	R P	24	22
GJB	23	26	M G	23	25	J Q	25	23
P B	25	24	D G	26	25	S R	27	27
J B	14	13	N H	28	27	R S	30	30
A B	24	22	P H	26	26	N S	17	17
E B	24	24	J H	23	20	D S	20	20
S C	26	26	L H	25	25	RMS	21	23
D C	27	26	C H	15	15	J S	27	24
J C	15	16	K L	19	18	H T	29	28
E D	19	19	D L	27	24	J T	18	19
J D	19	18	P L	14	15	S T	17	16
EAD	20	19	T L	21	21	D V	27	26
R E	20	19	B L	22	21	C V	25	22
W E	13	13	M M	22	24	C W	25	23
F F	16	15	G M	28	24			

The listeners showed a tendency to rate the samples more critically on the first than on the second test.

Table 2 shows the differences between the over-all speech scores for Test I and Test II.

TABLE 2

SUMMARY OF THE DIFFERENCES BETWEEN THE OVER-ALL SPEECH INTELLI-
GIBILITY SCORES FOR TEST ONE AND TWO.

Difference	Number of cases	Percent of cases
0	14	28%
1	17	34%
2	11	22%
3	7	14%
4	1	2%

There was no difference in the two scores for fourteen of the
fifty children or 28% of the cases. There was a difference of one point
between the two scores for seventeen children or 34%. There was a
difference of two for eleven children of 22%. In summary, there was a
difference of two or less in 84% of the cases. The remaining 16% was
divided into a difference of three, seven times or 14%, and a difference
of four once or 2%.

Ratings of the Speech Samples on the Basis of Individual Categories
By the Six Listeners -- The listeners' ratings on Test I and Test II were
correlated for each rating category in order to determine the ability of
the listeners to repeat a judgment. A correlation for each category was
obtained from scores representing the sum of the ratings of each of six
listeners. The sums of the individual ratings were considered to be as
meaningful as the mean rating or other derived figure.

Tables 3A through 3G show a summary of the ratings or scores made by six listeners for each rating category.

TABLE 3A

SUMMARY RATINGS OF VOICE QUALITY BY THE SIX LISTENERS.

Pupil	Test 1	Test 2	Pupil	Test 1	Test 2	Pupil	Test 1	Test 2
J A	17	18	J F	21	18	RRP	23	21
S A	20	22	S.F	22	21	S P	19	19
G B	18	22	L F	23	22	R P	18	21
GJB	23	25	M G	16	21	J Q	25	24
P B	20	20	D G	17	23	S R	22	21
J B	13	15	N H	22	23	R S	20	25
A B	20	26	P H	20	22	N S	16	19
E B	22	24	J H	22	21	D S	19	18
S C	25	23	L H	25	25	RMS	21	21
D C	24	22	C H	14	14	J S	18	22
J C	12	16	K L	16	17	H T	24	25
E D	13	15	D L	22	22	J T	17	18
J D	18	17	P L	12	15	S T	16	16
EAD	19	18	T L	22	19	D V	24	25
R E	19	19	B L	19	18	C V	22	25
W E	13	13	M M	20	23	C W	24	26
F F	14	13	G M	23	25			

TABLE 3B

SUMMARY RATINGS OF INFLECTION BY THE SIX LISTENERS

J A	20	18		J F	24	22		RRP	26	27
S A	23	26		S F	25	20		S P	17	17
G B	23	24		L F	28	21		R P	25	23
GJB	25	27		M G	22	23		J Q	25	23
P B	26	24		D G	27	24		S R	27	28
J B	13	13		N H	29	28		R S	30	28
A B	23	25		P H	27	27		N S	19	16
E B	26	26		J H	19	19		D S	19	18
S C	27	27		L H	22	24		RMS	22	24
D C	25	25		C H	14	13		J S	27	26
J C	14	15		K L	15	16		H T	29	29
E D	21	20		D L	26	25		J T	19	20
J D	18	19		P L	13	14		S T	26	18
EAD	19	18		T L	20	23		D V	26	26
R E	19	18		B L	18	20		C V	26	18
W E	12	12		M M	21	24		C W	23	24
F F	14	14		G M	27	25				

TABLE 3C

SUMMARY RATINGS OF RHYTHM BY THE SIX LISTENERS.

Pupil	Test 1	2	Pupil	Test 1	2	Pupil	Test 1	2
J A	15	14	J F	21	19	RRP	22	20
S A	22	20	S F	20	16	S P	16	15
G B	23	18	L F	23	19	R P	20	16
GJB	18	21	M G	19	21	J Q	19	21
P B	23	19	D G	22	19	S R	21	22
J B	12	11	N H	24	22	R S	24	24
A B	19	16	P H	23	22	N S	12	12
E B	24	.18	J H	20	13	D S	17	15
S C	22	22	L H	20	19	RMS	17	19
D C	23	21	C H	11	12	J S	23	20
J C	14	11	K L	13	17	H T	24	20
E D	17	13	D L	23	20	J T	13	14
J D	17	12	P L	11	12	S T	14	12
EAD	17	15	T L	17	17	D V	21	21
R E	16	15	B L	18	16	C V	22	16
W E	10	10	M M	16	18	C W	22	17
F F	14	12	G M	22	20			

TABLE 3D

SUMMARY RATINGS OF PHRASING BY THE SIX LISTENERS.

Pupil	Test 1	Test 2	Pupil	Test 1	Test 2	Pupil	Test 1	Test 2
J A	13	14	J F	18	18	RRP	22	18
S A	19	20	S F	19	16	S P	16	16
G B	20	15	L F	18	16	R P	18	17
GJB	15	19	M G	19	19	J Q	18	15
P B	19	19	D G	21	18	S R	21	20
J B	12	10	N H	21	19	R S	24	23
A B	19	13	P H	22	18	N S	12	12
E B	17	17	J H	17	11	D S	16	16
S C	18	21	L H	18	18	RMS	13	16
D C	20	20	C H	12	12	J S	23	17
J C	11	11	K L	15	13	H T	23	22
E D	18	17	D L	22	18	J T	13	14
J D	13	12	P L	11	12	S T	14	13
EAD	14	15	T L	15	15	D V	22	20
R E	14	13	B L	17	16	C V	19	17
W E	11	10	M M	17	17	C W	18	16
F F	12	12	G M	23	17			

TABLE 3E

SUMMARY RATINGS OF FLUENCY BY THE SIX LISTENERS.

Pupil	Test 1	Test 2	Pupil	Test 1	Test 2	Pupil	Test 1	Test 2
J A	16	14	J F	18	17	RRP	19	15
S A	19	21	S F	16	15	S P	20	18
G B	18	17	L F	21	18	R P	19	17
GJB	12	14	M G	18	18	J Q	17	16
P B	19	18	D G	20	17	S R	21	22
J B	12	9	N H	23	22	R S	24	24
A B	18	14	P H	15	14	N S	12	12
E B	16	17	J H	17	14	D S	15	16
S C	16	16	L H	21	20	RMS	13	14
D C	19	20	C H	12	12	J S	20	16
J C	12	11	K L	15	13	H T	23	23
E D	19	19	D L	21	16	J T	10	10
J D	15	12	P L	11	13	S T	16	14
EAD	15	14	T L	13	14	D V	19	17
R E	14	14	B L	17	16	D V	19	17
W E	11	10	M M	15	17	C W	18	16
F F	12	14	G M	17	14			

TABLE 3F

SUMMARY RATINGS OF PRECISION OF ARTICULATION BY THE SIX LISTENERS.

Pupil	Test 1	Test 2	Pupil	Test 1	Test 2	Pupil	Test 1	Test 2
J A	18	13	J F	24	23	RRP	28	24
S A	24	23	S F	21	20	S P	20	20
G B	21	18	L F	21	21	R P	22	20
GJB	24	26	M G	24	24	J Q	23	20
P B	22	23	D G	25	24	S R	26	24
J B	14	12	N H	27	24	R S	30	29
A B	21	20	P H	26	26	N S	16	16
E B	21	23	J H	21	19	D S	17	18
S C	25	23	L H	23	23	RMS	22	21
D C	25	23	C H	13	12	J S	25	21
J C	16	18	K L	20	18	H T	26	26
E D	16	14	D L	24	21	J T	18	20
J D	17	17	P L	13	12	S T	16	16
EAD	20	16	T L	21	19	D V	26	24
R E	19	17	B L	22	22	C V	22	20
W E	13	12	M M	22	23	C W	23	21
F F	14	13	G M	29	22			

TABLE 3G

SUMMARY RATINGS OF INTELLIGIBILITY BY THE SIX LISTENERS.

Pupil	Test 1	Test 2	Pupil	Test 1	Test 2	Pupil	Test 1	Test 2
J A	16	13	J F	24	22	RRP	24	22
S A	23	23	S F	20	17	S P	19	18
G B	20	15	L F	19	20	R P	20	20
GJB	19	24	M G	21	24	J Q	20	19
P B	21	22	D G	25	24	S R	24	23
J B	9	10	N H	24	23	R S	30	28
A B	19	18	P H	24	24	N S	13	12
E B	19	21	J H	20	19	D S	17	16
S C	20	23	L H	21	20	RMS	19	21
D C	23	22	C H	12	12	J S	24	20
J C	9	11	K L	18	15	H T	26	24
E D	12	13	D L	22	20	J T	17	18
J D	15	17	P L	12	11	S T	12	11
EAD	17	15	T L	19	17	D V	24	21
R E	17	17	B L	20	20	C V	20	21
W E	10	12	M M	19	22	C W	20	19
F F	15	12	G M	25	22			

Ratings for Test I and Test II are included in the tables. Pupils are listed in alphabetical order. A single rating or score for a given pupil is the total of the ratings of six listeners for each category.

The scores of the first test were paired with the scores on the second test. The Pearson Product-Moment method[2] for finding correlation was used. Table 4 shows the correlations between listeners' scores on Test I and Test II for each rating category.

TABLE 4

SUMMARY OF CORRELATIONS BETWEEN LISTENERS' SCORES FOR TEST 1 AND TEST 2 FOR EACH SPEECH RATING CATEGORY.

Category	Correlation	Category	Correlation
Voice quality	.77	Fluency	.86
Inflection	.89	Precision of articulation	.93
Rhythm	.82		
		Intelligibility	.91
Phrasing	.80		

The correlations ranged from a low of .77 for voice quality to a high of .93 for precision of articulation. Correlations for voice quality, rhythm, and phrasing were among the lower ones, while fluency, inflection, precision of articulation, and intelligibility had somewhat higher correlations, being above .86 in all cases. Throughout the test,

---

[2]Henry E. Garrett, Statistics In Psychology and Education (New York: Longmans, Green & Co., 1926) p. 163.

the listeners had a slight tendency to rate a little more critically on the first test than on the second.

The lower correlations found for voice quality, rhythm, and phrasing indicate that the categories themselves were harder to evaluate. This could be due to the make-up of the particular rating scale for the categories and also to the type of speech samples that were being rated. It is probably more difficult to rate a child's spontaneous speech for the rhythm and phrasing than to rate his speech efforts while reading selected speech samples, since the latter contains stereotyped phrases. Another aspect is that these categories themselves may not play as important a role in determining intelligibility as some of the others.

Ratings of the Speech Samples on the Basis of Individual Categories By Individual Listeners -- Coefficients of correlation were also determined for each of the six listeners for each rating category. Table 5 shows the correlations between Test I and Test II of individual listeners for each speech rating category. The individual coefficients of correlation were calculated by means of the "raw score method"[3] which is derived from the Pearson Product-Moment method.

---

[3] John W. Best, op. cit. p. 236

TABLE 5

SUMMARY OF THE CORRELATIONS OF INDIVIDUAL LISTENERS FOR TEST I
AND TEST II FOR EACH SPEECH RATING CATEGORY.

| Category | Listener | | | | | |
	A	B	C	D	E	F
Voice quality	.60	.59	.68*	.50	.31	.67
Inflection	.70	.61	.72	.74*	.63	.60
Rhythm	.52*	.47	.40	.42	.45	.29
Phrasing	.39	.62	.65*	.50	.59	.52
Fluency	.58	.73	.55	.56	.67	.87*
Precision of Articulation	.73*	.70	.66	.71	.56	.54
Intelligibility	.78	.88*	.74	.83	.69	.74

*Indicates the highest correlation for this category.

The six listeners had reliability coefficients of correlation for
the voice quality category ranging from a low of .31 to a high of .68.
Four of the listeners were within an eight point range of each other,
having .59, .60, .67, and .68 respectively. The other listener had a
correlation of .50.

The ratings of individual listeners showed correlations of .60, .61,
.63, .70, .72, and .74 for inflection; correlations of .29, .40, .42,
.45, .47, and .52 for rhythm; correlations of .39, .50, .59, .62, and
.65 for phrasing; correlations of .55, .56, .58, .67, .73, and .87 for

fluency; correlations of .54, .56, .66, .70, .71, and .73 for articula-
tion; and for intelligibility, correlations of .69, .74, .78, .83, and
.88.

The rhythm, phrasing, and voice quality categories appeared more
difficult than the others to rate and gave smaller correlations. The
correlations for the group of six listeners and the correlations of
the individual listeners bear this out (See Table 6.).

TABLE 6

SUMMARY OF THE CORRELATIONS OF THE RATINGS OF INDIVIDUAL LISTENERS
AND THOSE OF THE GROUP OF LISTENERS FOR EACH SPEECH CATEGORY.

Category	A	B	Listener C	D	E	F	Group of Listeners
Voice quality	.60	.59	.68	.50	.31	.67	.77
Inflection	.70	.61	.72	.74	.63	.60	.89
Rhythm	.52	.47	.40	.42	.45	.29	.82
Phrasing	.39	.62	.65	.50	.59	.52	.82
Fluency	.58	.73	.55	.56	.67	.87	.86
Precision of Articulation	.73	.70	.66	.71	.56	.54	.93
Intelligibility	.78	.88	.74	.83	.69	.74	.91

Table 5 shows that the correlations of individual listeners were
not consistently high or low. The numbers marked with an asterisk show

the highest correlation for each rating category.

This variation between the high and low scores of the listeners indicates that each listener found certain categories easier than others. However, it should be noted that all listeners were lower on certain categories than on others. For example, the low correlation (.54) for precision of articulation is higher than the highest correlation (.52) for rhythm. Thus, it can be concluded that certain rating categories were easier to rate consistently than others, and no one listener was consistently high or low for all categories.

Table 6 shows the correlations between the ratings of speech samples made by individual listeners and those made by the group of six listeners on Test I and Test II. Table 6 shows that in almost every case the individual correlations were smaller than those for the group of listeners. This was expected and at the beginning it was decided that a group rather than one or two individuals must rate the speech. The higher correlations for the group of listeners are due primarily to the fact that as a group the listeners tend to maintain a more consistent level of rating than it is possible for an individual to maintain. The group reduces the effect of the random errors of the individual.

From the data presented thus far, it appears that a group of listeners can repeat ratings of the speech of deaf children on the basis of the four categories: inflection, fluency, precision of articulation, and intelligibility with an acceptable amount of consistency. There is some question as to the value of ratings based on voice quality, rhythm, and phrasing as they now stand. Perhaps with some changes in the

rating scale and with more detailed instructions for rating these categories they can be made to yield more consistent results.

As the results thus far are considered, it appears that, with minor changes, the test can be used to evaluate the speech efforts of young deaf children.

The Relationship of the Rating Categories to Intelligibility -- Inter-category correlations were worked out in an attempt to determine the relative importance of each category to intelligibility. The importance of the intelligibility category was self-evident and so the correlation between each of the other categories and intelligibility was found. Table 7 shows the inter-category correlations.

TABLE 7

INTER-CATEGORY CORRELATIONS SHOWING THE RELATIONSHIP BETWEEN THE INTELLIGIBILITY CATEGORY AND THE REMAINING CATEGORIES.

Category	Correlation
Voice quality vs intelligibility	.69
Inflection vs intelligibility	.85
Rhythm vs intelligibility	.83
Phrasing vs intelligibility	.86
Fluency vs intelligibility	.72
Precision of articulation vs intelligibility	.99

The correlation between voice quality and intelligibility was .69.
The correlation between inflection and intelligibility was .85, for
rhythm and intelligibility .83. Phrasing and intelligibility had a
correlation of .86, fluency and intelligibility was .72, and precision
of articulation and intelligibility correlated almost perfectly, having
a correlation above .99.

The correlations for voice quality vs intelligibility and fluency vs
intelligibility are low enough to indicate that they are not very im-
portant factors in determining intelligibility. However, there are
inter-relations between these categories and others that make us hesi-
tate before eliminating them from the rating sheet. For example, if
voice quality was eliminated as a category, then the voice might become
a factor in the rating of another category such as precision of articu-
lation. If the listener had not already rated the voice quality,
then he might let a poor voice influence his rating of good articula-
tion.

The remaining five categories, with correlations of .83 to .99,
appear to be rather closely related to intelligibility. It appears
that precision of articulation is so closely related to intelligibility
that the rating of one can be predicted from the other. This does not
always hold true, however, especially when a child that is only hard-
of-hearing is considered. It is possible for him to have low precision
of articulation, yet have high intelligibility due to the naturalness of
his voice quality, rhythm, phrasing, fluency, and inflection. In
discussing this problem, Bell said, "Ordinary people who know nothing
of phonetics or elocution have difficulty in understanding slow speech

composed of perfect elementary sounds, while they have no difficulty in comprehending an imperfect gabble if only the accent and rhythm are natural."[4]

Results of Additional Groups of Listeners -- The question arises as to whether or not another group of listeners under the same test conditions would yield similar results. The answer to this question was sought by testing three other groups of listeners.

Eighteen college graduates who were engaged in a teacher training program at the Clarke School for the Deaf were divided randomly into three groups of six listeners. These students were considered to have approximately the same background for understanding the speech of deaf children as beginning teachers of the deaf.

These three groups were given identical instructions and practice periods and followed the same technique for rating the speech as the original group of listeners. They did the rating in the same room at the same time and so conditions for these three groups of listeners were identical. These three groups of listeners rated ten of the fifty pupils chosen at random. Table 8 shows the over-all speech scores given to the ten pupils by the original group of listeners and the three test groups of listeners. These scores were determined by the method mentioned earlier for finding a child's over-all speech score (p. 19).

---

[4]Alexander G. Bell, The Mechanism of Speech (New York and London: Funk & Wagnalls Co., 1906) p. 15.

TABLE 8

SUMMARY OF OVER-ALL SPEECH SCORES GIVEN TO TEN PUPILS BY THE THREE
TEST GROUPS OF LISTENERS AND THE ORIGINAL GROUP.

| Pupils | Original group | | Test groups | | |
	Test 1	Test 2	A	B	C
G B	23	26	24	26	25
E B	24	24	23	24	25
R E	20	19	22	22	21
W E	13	13	14	14	14
J H	23	20	22	25	25
T L	21	21	19	20	20
P L	14	15	16	15	16
M M	22	24	24	23	24
R S	21	23	21	25	24
J T	18	19	20	23	20

The three test groups of listeners showed a slight tendency to rate
the speech samples more critically than the original group of listeners.
However, the scores of the three test groups compare very favorably
with each other and with those of the original group, as Table 8 shows.
It is believed that if these groups had been somewhat selected for the

task and if they had been able to rate more than ten pupils, thus giving them practice, the results would have been even nearer to the original group. However, from this limited bit of information, this author is encouraged to believe that a group of listeners can be selected and trained to rate the speech with a high degree of consistency.

Careful analysis of the data presented indicates that it is possible for a test of this type to be administered to young deaf children and scored by a group of listeners with satisfactory and reliable results.

CHAPTER V

SUMMARY AND CONCLUSIONS

CHAPTER V

SUMMARY AND CONCLUSIONS

The purpose of this paper was to determine the reliability of a group of listeners in evaluating speech samples of young deaf children. At present, we are not aware of a satisfactory way of determining the speech intelligibility of young deaf children. The problem was to determine with what consistency it is possible for a group of listeners to assess the speech for its voice quality, inflection, rhythm, phrasing, fluency, precision of articulation, and intelligibility. The final objective of this study was to standardize a test that could be used with young deaf children.

Fifty pupils in the Lower and Middle Departments of the Clarke School for the Deaf were used in the study. There were twenty-four boys and twenty-six girls ranging in age from six years, nine months to eleven years, seven months.

Spontaneous speech samples were gathered from this group of fifty children and were appraised and scored by a group of six listeners who were considered qualified for this job.

The spontaneous speech was stimulated by presenting groups of ten simple pictures to the children. The ten pictures were chosen at random from a group of thirty-five pictures. As a child talked about the picture, the teacher silently pointed out items in the picture that might stimulate speech. It was necessary to encourage some of the younger pupils. The speech samples were tape recorded to enable them to be scored at a later and more convenient time.

The listeners were teachers of the deaf who were familiar

with the speech of deaf children. Some of the listeners were beginning
teachers and one had considerable experience with the speech of the deaf.

After a training period and a practice session, the speech samples
were played back to the listeners through high-fidelity equipment and
individual headphones.

A rating sheet (Figure I, p.16) consisting of seven speech rating
categories and a rating scale for each was devised to aid in the rating
of a child's intelligibility. The seven speech rating categories were:
voice quality, inflection, rhythm, phrasing, fluency, precision of
articulation, and intelligibility.

The child's over-all speech score was considered to be the average
score given by the six listeners. The score given by a single listener
was the sum of individual ratings on each of the seven categories.

In order to determine the reliability of the test, each of the
fifty children was rated twice. The second ratings were made after a
waiting period of approximately six weeks.

The data were analyzed as follows: (1) Correlations were deter-
mined for the over-all speech scores of Test 1 and Test 2; (2) correla-
tions were determined for the scores given by each individual listener
on Test 1 and Test 2 for each rating category; (4) correlations were
determined between the scores for intelligibility versus the scores for
the remaining categories to determine the relationship between intelli-
gibility and the other categories; and (5) comparisons of scores given
by additional groups of listeners for a sample group of pupils and the
original group of listeners were made.

The coefficient of correlation between the over-all speech scores given by the listeners on the first test and on the second test was .89. There was a possible range of twenty-five points on the test. There was no difference between the two scores for 28% of the cases. There was a difference of two or less for 84% of the cases. Table 2 p.24 shows these results.

The coefficients of correlation for the ratings of the six listeners for each rating category was .77 for voice quality, .89 for inflection, .82 for rhythm, .80 for phrasing, .86 for fluency, .93 for precision of articulation and .91 for intelligibility.

The correlations between the two tests for each individual listener for each speech rating category ranged from .31 to .68 for voice quality, .60 to .74 for inflection, .29 to .52 for rhythm, .39 to .65 for phrasing, .55 to .87 for fluency, .54 to .73 for precision of articulation, and .69 to .88 for intelligibility.

The data showed that no single listener was consistently high or low in his ratings. This indicates that one listener may be superior in rating one category and not quite as good as another listener in rating another category.

From an examination of this information, we can predict that the use of a group of listeners for scoring is acceptable and will yield reliable results. The coefficients of correlation for the ratings of individual listeners leaves considerable doubt as to the reliability of individual listeners to rate the speech. However, as a group, the

listeners do an acceptable job and this method should be maintained.

An attempt to determine the relative importance of each speech rating category in relationship to intelligibility was made. This was determined by inter-category correlations. The correlation for voice quality vs intelligibility was .69. Correlations were .85 for inflection vs intelligibility, .83 for phrasing vs intelligibility, .72 for fluency vs intelligibility, and .99 for precision of articulation vs intelligibility.

The data indicates that probably all of the speech rating categories except, possibly, voice quality and fluency, are closely related to intelligibility and deserve consideration in the evaluation of speech. However, voice quality and fluency are felt to be so involved with the speech process that, even though the correlations between them and intelligibility were lower than for the others, they are important and should be considered as factors that determine the intelligibility of speech.

An additional test was made to determine if similar groups of listeners would yield similar results. Three groups composed of teachers-in-training at the Clarke School for the Deaf were used for this purpose. These test groups rated ten pupils, chosen at random, under similar conditions as the original group of listeners.
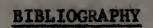
The over-all speech score for a pupil as determined by each of the three test groups of listeners were very close to each other. These scores compared very favorably to those assigned by the original group of listeners, also. The scores given to pupils by any one group were

similar enough to those of any other group, except in a very few cases, to indicate that they are reliable scores. It is believed that if the test groups of listeners had had practice equal to the original group, the scores would have been even closer together. This indicates that any group of listeners with similar backgrounds and training can be trained and used to rate the speech of young deaf children with a reliable degree of consistency.

From the information learned from this study, it can be concluded that, with some minor revisions, this test can be used to evaluate the speech of young deaf children with satisfactory and reliable results.

A change in the rating sheet is recommended for future use. The categories would be easier to rate, in certain instances, if there were broader scales with either more definite, descriptive terms to aid in judging or merely a number scale ranking from excellent or normal to very poor.

Whenever we scratch the surface of a body of information, we always find that there is a great deal more to be learned. This is especially true in the field of speech teaching, speech testing, and definitely, speech evaluation.

BIBLIOGRAPHY

BIBLIOGRAPHY

Bell, Alexander Graham. The Mechanism of Speech. New York &
London: Funk & Wagnalls Co., 1906. 129 pp.

Best, John W. Research in Education. Englewood Cliffs, N.J.:
Prentice-Hall, Inc., 1959. 320pp.

Bjuggren, Gunnar, "A Method To Test The Intelligibility of The
Speech of Pre-School Children With Severe Hearing Impairment.
A Preliminary Report." Stockholm: Acta-Oto-Laryngologica--
Supplementum 110. 1954. pp. 83-88.

Diehl, Charles F. and McDonald, Eugene T., "Effect of Voice Quality
On Communication." Journal of Speech and Hearing Disorders,
Vol. 21 (June 1956), pp. 233-237.

Diehl, Charles F. and Stinnett, Charles D., "Efficiency of Teacher
Referrals In a School Speech Testing Program." The Journal of
Speech and Hearing Disorders, Vol. 24 (Feb. 1959) pp. 34-37.

Diehl, Charles F., White, Richard, and Burk, Kenneth W., "Voice Quality
and Anxiety." Journal of Speech and Hearing Research, Vol. 2
(Sept. 1959) pp. 282-285.

Egan, James P., "Articulation Testing Methods." The Laryngoscope,
Vol. 58 (1948) pp. 955-991.

Farman, J. Jay and Phillips, Betty, "The Farman-Phillips Speech
Intelligibility Diagnostic Test." The Volta Review, Vol. 56
(Apr. 1954) pp. 168-170.

Garrett, Henry E. Statistics In Psychology and Education. New York:
Longmans, Green & Co., 1926. 317 pp.

Hudgins, C. V., "A Method of Appraising the Speech of the Deaf."
The Volta Review, Vol. 51 (Dec. 1949) pp. 597-601.

Larr, Alfred L. and Stockwell, Robert P., "A Test of Speech Intelligi-
bility." The Volta Review, Vol. 61 (Nov. 1959) pp. 403-407.

Oyer, Herbert J., "Speech Error Recognition Ability." Journal of
Speech and Hearing Disorders, 24 (Nov. 1959) pp. 391-395.

Perrin, Elinor Horwitz, "The Rating of Defective Speech By Trained And
Untrained Observers." The Journal of Speech and Hearing Disorders,
Vol. 19 (March 1954) pp. 48-52.

Sherman, Dorothy and Morrison, Sheila, "Reliability of Individual
Ratings of Severity of Defective Articulation." The Journal of
Speech and Hearing Disorders, Vol. 20 (Dec. 1955) pp. 347-359.

Templin, Mildred C., "A Non-Diagnostic Articulation Test."
The Journal of Speech Disorders, Vol. 21 (Dec. 1947) pp. 392-396.

Templin, Mildred C., "Spontaneous Versus Imitated Verbalization In
Testing Articulation In Preschool Children." The Journal of
Speech Disorders, Vol. 12 (Sept. 1947) pp. 293-300.

Tolhurst, Gilbert C., "What Are The Jobs The Therapist Will Do?"
The Bulletin of the National Association of Secondary School
Principals, Vol. 34 (Nov. 1950) pp. 73-79.

Wood, Kenneth Scott, "Measurement of Progress in the Correction of
Articulatory Speech Defects." The Journal of Speech and Hearing
Disorders, Vol. 14 (June 1949) pp. 171-174.

ACKNOWLEDGMENTS

ACKNOWLEDGMENTS

PROBLEM APPROVED BY:

Albert J. Anthony
C. V. Hudgins
(PROBLEM COMMITTEE)

DATE: MAY 1960