1974

An Analysis of the Value of a Sentence Articulation Inventory

Carol Ann Jones
Eastern Illinois University

PAPER CERTIFICATE #2

TO: Graduate Degree Candidates who have written formal theses.

SUBJECT: Permission to reproduce theses.

The University Library is receiving a number of requests from other institutions asking permission to reproduce dissertations for inclusion in their library holdings. Although no copyright laws are involved, we feel that professional courtesy demands that permission be obtained from the author before we allow theses to be copied.

Please sign one of the following statements:

Booth Library of Eastern Illinois University has my permission to lend my thesis to a reputable college or university for the purpose of copying it for inclusion in that institution's library or research holdings.

_____ _____

Date Author

I respectfully request Booth Library of Eastern Illinois University not allow my thesis be reproduced because I wish to give my written permission first.

Dec. 13, 1974

Date

pdm

AN ANALYSIS OF THE VALUE OF

A SENTENCE articulation INVENTORY

(TITLE)

BY

CAROL ANN JONES

THESIS

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF

MASTER OF SCIENCE

IN THE GRADUATE SCHOOL, EASTERN ILLINOIS UNIVERSITY
CHARLESTON, ILLINOIS

1974
YEAR

I HEREBY RECOMMEND THIS THESIS BE ACCEPTED AS FULFILLING
THIS PART OF THE GRADUATE DEGREE CITED ABOVE

13 Dec 74
DATE

ADVISER

13 Dec 74
DATE

DEPARTMENT HEAD

AN ANALYSIS OF THE VALUE OF A SENTENCE ARTICULATION INVENTORY

BY

CAROL ANN JONES

ACKNOWLEDGEMENTS

I want to express my appreciation, first of all, to my parents and family for their continuous love and encouragement.

I also want to thank Dr. Miner for his advice in the planning and writing of this study. Many, many thanks go to Mr. Jimmie Page for his patience and understanding and for his assistance in locating subjects to be used in my research.

To Lynette, who was willing to type her first thesis, a big thank-you.

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION

Listeners form many of their ideas about a person based on the adequacy of his speech or by his ability to communicate. Since, according to Anderson (1953) and Powers (1971), the various types of articulatory disorders account for about 75 to 80 per cent of the total speech defects found among the public school children, defective articulation is likely to be that feature of speech which is distracting to the listener and causes him to react negatively. One of the major goals of therapy is to reduce as quickly and as efficiently as possible these deviations which are distracting to the listener. Several features of defective articulation affect the listener's reactions. One which might reasonably be supposed to relate to listener reaction is the number of speech sounds misarticulated. Perrin (1954) found a high correlation between number of articulation errors and listener judgments of severity of articulatory defect. Ordinarily, the more sounds a person has trouble with the more severe his general speech handicap is considered to be. However, we cannot overlook the fact that some consonant sounds occur much more frequently in the language than do others. A person who cannot produce a frequently-occurring speech sound is more socially disabled than if he could not produce a speech sound of infrequent occurrence. Wood (1949) utilized this

theory when he developed a method for quantifying social adequacy of connected speech by frequency weighting. Wright (1954) sought a more refined method for obtaining quantative measures of defective articulation than counting number of errors. Recognizing that articulation errors vary in degree of defectiveness as well as in type, he employed a seven-point scale describing the type of error. To summarize, defective articulation may be characterized by various types of errors, such as omissions, substitutions, or distortions. The defective sound may occur frequently, or it may occur infrequently. The number of defective sounds may be many, or it may be only a few. The most valid clinical tool for analyzing articulatory behavior, then, must take into consideration as many of these characteristics as possible.

When assessing the articulation skills of a child, the speech pathologist has a variety of single-word articulation tests from which to choose. However, the validity of these tests to predict the impact on their listeners of the articulatory characteristics of children's connected speech through scores based on small samples of highly structured speech is questionable.

Research by Daniloff and Moll (1968), Faircloth and Faircloth (1969), and Amerman, Daniloff, and Moll (1970) on the effects of coarticulation has shown that the production of speech sounds in single words is more intelligible than the production of the same words in connected speech.

Since conversation is the most demanding level of speech production, sounds should also be evaluated at this level.

Goldman and Fristoe (1969) have attempted to assess conversational speech in their Sounds-in-Sentences Subtest. One difficulty encountered here is the problem of eliciting spontaneous speech while controlling the content. The examiner must be able to evoke the sounds he wishes to examine.

Sentence articulation tests have been developed to provide a more systematic means of assessing speech sound production at a complex level which is similar to conversational speech. One of these tests was developed by Templin and Darley (1969) and another by Fisher and Logemann (1971). The selection of stimulus words which appear on both these tests was based on their familiarity to children. Familiarity was determined by the appearance of these words on basic reading and oral lists of children. However, research which has been done since these tests were developed suggests another factor which might be important in the selection of test stimuli. Griffith and Miner (1973), using the first 1000 words of the Thorndike-Lorge 10,000 most frequently occuring words, found that phonetic contexts rank order themselves according to frequency of occurrence. Two identical studies were done by Dorn (1973) and Schneider (1973). They analyzed the Thorndike-Lorge third and fourth grade list of the most frequently occurring words for /r/ and /s/, /l/, respectively. Since some phonetic contexts occur more frequently than others, it would seem important to consider phonetic context in the selection of stimulus items on articulation tests.

The purpose of this study was to determine the relationship between scores obtained from a sentence articulation inventory (SAI),

which controls for sentence length, frequency of occurrence of words, and phonetic context, to scores obtained from single-word articulation tests and listener judgments. Stated as a research hypothesis: Scores obtained from a sentence articulation inventory are more representative of a person's conversational speech and correlate more highly with listener judgments of connected speech than scores obtained from single-word articulation tests. The following questions were posed at the outset of this investigation:

1. Can observers reliably rate the severity of misarticulations in the connected speech of subjects?
2. What is the relationship between the scores obtained from the single-word articulation tests and the ratings of listeners made from tapes of the subjects' connected speech?
3. What is the relationship between the scores obtained from the sentence articulation inventory, the ratings of listeners made from tapes of the subjects' connected speech, and the scores obtained from single-word articulation tests?
4. What is the relationship between the scores obtained from the sentence articulation inventory and the ratings of listeners made from tapes of the subjects' connected speech?
5. What is the relationship between scores obtained from the sentence articulation inventory and the scores obtained from the single-word articulation tests?

CHAPTER II

REVIEW OF THE LITERATURE

Traditionally, the speech of persons with defective articulation has been analyzed through the use of single-word articulation tests which claim to test the production of speech sounds in the initial, medial and final positions in words and in blends. Stetson (1957), however, pointed out that there are no bases in the physiology of connected speech for such terms as initial, medial and final sounds. His investigations revealed that connected speech consists of series of syllables and that consonant sounds act to release or arrest these syllables. This may mean that single-word articulation tests, which test the production of speech sounds in the initial, medial, and final positions in words and blends, are not valid indications of the person's connected speech since sounds do not exist in these positions in connected speech. It also seems to indicate that an analysis of connected speech might provide a more accurate evaluation of the adequacy of the person's speech for communicative purposes.

The validity of single-word articulation tests for indicating the production of sounds in connected speech was also challenged by Templin (1947). She recognized that a person may be able satisfactorily to produce given sound elements in single words but be unable to maintain this degree of articulatory competence in rapidly moving speech. Other weaknesses in single-word articulation tests have been recognized

and attempts have been made to improve them. Some of these will be discussed.

Counting the errors made on an articulation test, which is the most common method of measuring the ability of articulatory behaviors, is another possible weakness of the single-word articulation test. Counting sound errors gives a rough numerical indication of defectiveness, however, the sounds in the English language do not occur with the same frequency. Inability to produce sounds which appear very seldom cannot be considered as severe a problem as inability to produce more frequently used sounds. The degree to which a sound-error is distracting may be related to the frequency of occurrence of that sound in the English language. This factor was taken into consideration by Wood (1949) when he attempted to quantify the social adequacy of connected speech by devising an articulation index in which each consonant was weighted by prorating the values of Travis's (1949) table of the frequency of occurrence of consonant sounds in the speech of American children into the initial, medial, and final positions in words. The numerical value of the sounds correctly produced were added together to obtain a quantitative description of the child's ability to articulate sounds correctly. Research has now been done in the area of phonetic context by Griffith and Miner (1971), Dorn (1973), and Schneider (1973). Using the Thorndike-Lorge Word lists of the most frequently occurring words, they found that particular sounds do indeed occur more frequently than other sounds. The phonetic contexts with which these sounds occur have also been determined.

Wood's method of quantifying the social adequacy of connected speech by weighting the various consonant sounds according to their position in the word was a step in the right direction. However, Henrickson (1948) critized Wood's equal prorating by showing that consonants did not appear equally in all positions. Another error in the construction of Wood's Articulation Index is that he included only consonants. If a score is to accurately represent speech adequacy, it must include all speech sounds. Barker (1960) did this by devising a method to compute an Articulation Score based on the relative frequency of all speech sounds. She found the Articulation Score to be related to social adequacy of speech.

Another attempt was made by Wright (1954) to develop a method of analyzing defective articulation. He recognized that speech sound-errors probably vary in their relative degrees of distractability to the listener. The misarticulations of one speaker may not be as distracting to the listener as the misarticulations of another speaker. Wright devised a method for scaling the magnitude of speech sound-errors by assigning numbers according to the type of error; one, representing a correctly articulated sound; two through five, progressive amounts of distortion of the sound; six, a substitution; and seven, an omission. The rationale for this scale is based in part on the finding of Roe and Milisen (1942) that, in general, as articulation skills develop, sounds are likely to be first omitted, then distorted, and finally produced correctly. From this order of development, Wright assumed that, in general, listeners will be distracted more by omissions than by substitutions, and more by substitutions than distortions.

Alcorn (1971), in his study to determine the comparisons of articulation severity ratings of /s/ and /r/ by lower-, middle-, and upper-socioeconomic groups, found that a frontal lisp was more accepted by the observer than a lateral lisp, and a lateral lisp more than a distorted /r/. According to this then, there appears to be degrees of acceptability of articulation errors within a class itself. In this study by Alcorn, for example, all errors were in the class of distortions, but one distortion was more acceptable than another distortion.

Within the past ten years an interest has developed in coarticulation and the effects of coarticulation on the production of adjacent speech sounds. Speech articulators have been found not to function individually and independently. Research by Daniloff and Moll (1968) showed that coarticulation of lip protrusion extends over as many as four consonants in a sequence preceding the rounded vowel /u/. In addition, syllable and word boundaries do not appear to affect the starting of protrusion. Amerman, Daniloff, and Moll (1970), in a study of lip and jaw coarticulation for /æ/, found jaw lowering to occur two phonemes preceding the vowel and, in ninety percent of the cases, found jaw movement to occur following the phoneme. Since coarticulation has been shown to affect the production of sounds occurring three or four sounds prior to or following a particular sound, it is possible that in conversational speech, these sounds which are affecting the production of a particular sound may not occur in the word containing that sound but in words preceding or following it. Therefore, determining the adequacy of connected speech from what is heard in isolated word utterances may

not be a valid interpretation of scores derived from single-word articulation tests.

Faircloth and Faircloth (1960), in their analysis of the articulatory behavior of a speech-defective child in connected speech and in isolated-word responses, found large differences between words produced in connected speech and the same words produced in isolation. Also, the responses of isolated-words were consistently judged to be more intelligible than the same words when produced in connected speech. The results from this study strongly suggest that an analysis of connected speech is a more appropriate means of describing a person's habitual articulatory behavior than single-word articulation testing.

Although many attempts have been made to improve the usefulness of single-word articulation tests, the preceding arguments would seem to support the need for testing articulatory behaviors in connected speech. Some attempts have been made to develop sentence articulation tests. Templin and Darley developed a sentence form of their articulation test, using the test words which appeared in their picture articulation test. All but four of the stimulus words appear on the first grade level of the Rinsland Basic Vocabulary. This list includes both oral and written words of children. The four stimulus words that do not appear on the Rinsland list at this level were felt by the author to be familiar to a majority of young children.

The Fisher-Logemann sentence form consists of fifteen sentences. Within these fifteen sentences were a total of 168 words. After eliminating

proper nouns and repetitions, there were 118 words. Of these 118 words, 82 per cent were included in the Gates list of 2500 most common words in the reading vocabulary of Grade 2. Seventy-eight per cent were included in the first 1000 words of the Thorndike list. The remaining words were included in the first 12000 words of the Thorndike lists.

Goldman and Fristoe felt the formal methods used previously for testing connected speech had been limited to direct imitation of sentences and reading of prepared material, and that neither represented the type of spontaneous production of speech sounds that occurs in actual conversation. They devised a method of eliciting conversational-type speech which was also content-controlled. Their method consisted of two narrative stories which were read aloud by an examiner. The pictures which were used to illustrate the stories were to be used as memory aids as the subject retold the story. By using the pictures, some control was placed over context of the story. Although this test has the potential for assessing most of the consonant sounds, it has been limited to those which are most likely to be defective.

Some weaknesses in the construction of these three tests, the Templin-Darley sentence articulation test, the Fisher-Logemann sentence articulation test, and the Goldman-Fristoe Sounds-in-Sentences Subtest, are: phonetic context was not considered in the selection of the test stimuli and all speech sounds are not tested.

In summary of the literature reviewed:

1. Single-word articulation tests are designed
   to test the production of speech sounds in the

initial, medial, and final positions in words and blends for which there is no bases in the physiology of connected speech. Rather, consonant sounds act to release or arrest syllables which comprise connected speech.

2. Various consonant sounds and their phonetic contexts with which they exist have been found to occur more frequently than others. If an articulation test score is to be representative of conversational speech, the frequency of occurrence of consonant sounds and the contexts in which they occur must be taken into consideration.

3. Types of speech-sound errors, omissions, substitutions, and distortions, vary in their relative degrees of distractibility to the listener. There are also degrees of acceptability of articulation errors within a class itself. These degrees of distractibility/acceptability are important features in determining communicative abilities.

4. Coarticulation is a major factor influencing the production of speech sounds in connected speech, the effects of which may extend up to four syllables preceding or following a particular sound. Coarticulation is not restricted by word boundaries. This, then, is another phenomenon of speech which causes connected speech to be more difficult than single-word utterances.

CHAPTER III

CONSTRUCTION OF SENTENCES, SELECTION OF SUBJECTS, PROCEDURES, AND EQUIPMENT

Construction of the Sentence Articulation Inventory

Two sets of sentences containing two of the most frequently misarticulated phonemes, /r/ and /s/, were developed to determine if a sentence articulation inventory would be a more valid tool for assessing the adequacy of a person's conversational speech than single-word articulation tests. These two sounds were chosen since they are two of the most frequently misarticulated sounds and two of the four most frequently occurring sounds (Travis, 1949). These sounds, /r/ and /s/, were tested in all phonetic contexts with which that sound has been found to occur in words contained in the Thorndike-Lorge list for grades one and two. Initial and final positions in both accented and unaccented syllables in words and in blends were tested. Information for phonetic context was based on data gathered from research done by Griffith and Miner (1971), Dorn (1973), and Schneider (1973). All test words were taken from the Thorndike-Lorge list of the 1000 most frequently occurring words, this being the level for grades one and two. Sentence length was also a factor controlled as retention could bias responses in longer sentences. To summarize, the sentence articulation inventory contains two sets of sentences to test /r/ and /s/. Three factors which were

controlled were: sentence length, word frequency, and phonetic context. A child's sentence articulation score was derived by counting the number of correct responses. A screening form of the sentence articulation inventory was used to obtain speech samples to be judged by listeners. The screening sentences contain some of the contexts which occur for the /r/ and /s/ phonemes. These contexts range from high to low frequency of occurrence. These contexts are shown in Table 1.

Selection of Subjects

Ten subjects were chosen randomly from a group of children having been identified by a speech pathologist as having articulation errors for /r/ or /s/. Five children were used who had articulation errors for /r/, and five were used who had articulation errors for /s/. Other variables which were controlled were age, intelligence, and hearing acuity.

1. Age: All subjects were between the ages of ten and eleven. Their birthdate was determined by referring to each child's permanent school record.
2. Intelligence: In an attempt to get an "average" intellectual range of subjects, students which

Table 1.--Frequency of Occurrence of Phonetic Contexts for /r/ and /s/ in Screening Form of the SAI

Hi	Moderate	Lo
I [rɛ]	I [ro]	F [ɜɪr]
F [ɪr]	F [ar]	F [rs]
F [rt]	I [fr]	
F [ɪs]	I [sʌ]	F [aus]
I [st]	I [str]	I [sæ]
F [st]		I [sm]
		F [nst]
		F [ks]

have been or were presently enrolled in EMH or learning disability classes were excluded from this study.

3. Hearing: All subjects were to have normal hearing. A child was considered to have normal hearing if he had passed a hearing screening test administered at 20 dB for the frequencies 250, 500, 1000, 2000, and 4000 Hz.
4. Prior Therapy: All subjects were to have had speech therapy prior to their participation in this study.

Testing Procedures

Each subject was given nine single-word articulation tests and the sentence articulation inventory. These nine single-word articulation tests included the following: Templin-Darley Test of Articulation, Goldman-Fristoe Test of Articulation, Fisher-Logemann Test of Articulation Competence, Developmental Test of Articulation, McDonald Screening Deep Test of Articulation, McDonald Deep Test of Articulation, Bryngleson-Glaspy Test of Articulation, Predictive Screening Test of Articulation, and the Laradon Articulation Scale. A sample of the child's connected speech was also obtained by recording his speech while reading the screening form of the sentence articulation inventory. The presentation of each of the above was randomized for each subject to allow for any improvement or changes which might occur in the child's articulatory behavior as a result of the testing.

All the tests and the recording were completed within the same day for each subject.

Recordings

All recordings were made in quiet conditions on a Rheem Caliphone 74-Tc Solid State tape recorder, serial number 301601691, with a Shure Microphone. Scotch Magnetic tape, silicone lubricated 1.5 mil. acetate backing was used. Tape speed was set at 7.5 inches per second. Although studies by Morrison (1955) and Sherman and Morrison (1955) have found ten seconds to be sufficiently long for listeners to reliably rate severity of articulation defectiveness, the length of the speech sample to be judged did not create as much concern as the content of the speech sample. Since severity is perceived in part as a function of frequency of occurrence of the misarticulated sound (Morrison, 1955), the speech sample to be judged was made by recording each subject's speech as he read the screening sentences which contained the phoneme for his misarticulated sound, /r/ or /s/. The recordings were made with the subject approximately two to three feet from the microphone. The child was facing the microphone and was speaking directly into it. Between each speech sample which the listeners were to judge was a space of seven seconds to allow the listeners to make decisions and mark the scale.

Listeners

Research has been done to examine the reliability of observers to scale articulation severity. Two of these studies will be reviewed

as a justification for the use of observers in this study to scale the connected speech of the subjects. Perrin (1954), in a study done to investigate whether there was any difference in the ratings of severity given by trained and untrained judges to functional articulation defects, found that trained and untrained observers do not differ significantly in their evaluations of the severity of functional articulation defects. Both groups agreed significantly in their rankings within their respective groups. However, the untrained observers were slightly more in agreement with themselves than were the trained observers. DeMuth (1960), in another study, found mothers, teachers, and speech pathologists able to reliably scale articulation severity. In this investigation, both trained and untrained listeners were used to rate each child's connected speech. The untrained judges were students enrolled in the Speech Pathology and Audiology 2800 class. This was the introductory class to speech pathology and audiology. The trained judges were advanced students in speech pathology with a minimum of 100 hours of clinical practicum, trained and experienced in the evaluation of articulation defects. The reliability of judges was set at .95. As many listeners as was necessary were used to obtain this level of agreement. It was estimated that a maximum of twenty minutes would be required for the listeners to scale the recordings. Before recording their evaluations of each subject's speech, these instructions were read to the judges:

> You are asked to judge a group of sentences which are read by children. You are to judge each group of sentences read by each child in relation to a seven-point scale of articulation defectiveness. Articulation

defectiveness is defined as the degree to which the misarticulations interfere with the communication process. Make your judgments solely on articulation defectiveness, not on the basis of reading difficulty.

The scale is one of equal intervals from one to seven, with one representing the lowest degree of articulation defectiveness, and seven representing the highest degree. The interval four is the middle between one and seven in degree with the other numbers following at equal distances along the scale. Do not attempt to place segments between any two of these seven points, but only at these points.

After hearing the groups of sentences read by each child, you will record immediately the number of the scale position you think the sample should have. You will record your scale number on your answer sheet. Notice that you will start at the top of the column and work toward the bottom.

Before you record any judgments, listen to the entire set of grouped sentences to acquaint yourself with the experimental task and to acquaint yourself with the range of samples with respect to the degree of articulation defectiveness, which you are asked to judge.

After you have acquainted yourself with the task and the range, make a judgment on every sample. If you are somewhat doubtful, make a guess as to the most suitable scale position. Are there any questions? Ready...

Scaling Method

Each listener rated the severity of articulatory defectiveness on an equal-appearing interval scale, containing seven degrees of severity; number one representing the least defective articulation progressing to number seven representing the most defective articulation. Each child's speech was rated immediately following its presentation. Justification for using equal-appearing intervals was based on data from a study done by Morrison (1955) in which she was attempting to find a method for obtaining measures of articulatory defectiveness for short segemets of continuous speech. She obtained

reliable scale values of severity when using the equal-appearing interval method. Similar results were found by Sherman and Morrison (1955).

Examiner Reliability

The examiner was the only person involved in the evaluation of the subjects' responses as correct or incorrect on the articulation tests. Therefore, the reliability of the examiner to make these evaluations had to be established. Interexaminer reliability was tested. This was tested before any of the testing for research information was initiated. Two children, with articulation errors of /r/ and /s/ and who were not used in the research testing, were administered the sentence articulation inventory. The subjects' responses were scored by the examiner and two other advanced students in speech pathology. A percentage of agreement in the scoring of the responses as right versus wrong was determined (Winitz, 1969). Interexaminer reliability was to be accepted if a percentage of agreement of 90 per cent was obtained. Examiners were found to agree on 93 per cent of their judgments for /r/ and on 95 per cent of their judgments for /s/.

Construct Validity

The construct validity, or the meaning of the test, was also examined. The theory underlying the sentence articulation inventory was that the scores obtained by an individual on the test would be an indication of how that individual's speech would be judged by his listeners. To establish construct validity, a comparison was made

between the sentence articulation inventory scores and the listener judgments. If a high correlation was found, .90, the sentence articulation inventory would be considered to be a valid test.

Analysis of Data

As stated earlier, the purpose of this paper was to determine the relationship between scores obtained from a sentence articulation inventory to scores obtained from single-word articulation tests and listener judgments. To determine the association of all measures used in this investigation, the scores obtained from the listener judgments, from the nine articulation tests stated earlier, and from the sentence articulation inventory were placed in rank order. The Kendall Coefficient of Concordance, W, which is applicable when data are in rank form and there are more than two sets of such data, was used (Downie & Heath, 1970). This tells the overall degree of association between the ranks. If there were no association whatever between the rater groups, and a rank-order coefficient of correlation were computed between the ranks, it would be near zero. On the other hand, if there is agreement, the coefficient would be significantly different from zero. The coefficient of concordance, W, expresses the average agreement, on a scale of .00 to 1.00, between ranks. The desired Kendall W was set at .80. If a coefficient of concordance of .80 or greater was obtained, then the ranks from the scores obtained from the various tests and the listener judgments would be considered to have a high overall degree of relationship. On the other hand, if the coefficient of concordance was less than .80, the overall degree of

relationship was considered to be low. If this was the case, it was the interest of this study to determine the relationship of any pairs of tests and especially the test(s) which most closely related to the listener judgments. The Kendall Tau correlation coefficient would be used in this comparison. For this study sixty-six, $\frac{(n)\ (n-1)}{2}$, Kendall tau's would be determined for each of the two defective groups, /r/ and /s/. In determining the relationship of any two measures, Kendall Tau was set at .80.

CHAPTER IV

RESULTS AND DISCUSSION

Articulation scores were obtained for five /r/ and five /s/ articulation defective children, and a sample of each child's speech was recorded. Two groups of listeners, trained and untrained, were chosen to scale the children's speech on a seven-point scale of articulation defectiveness. All test scores and listener evaluations were placed in rank-order for means of comparison. In order to meaningfully interpret these ranks in answer to the questions posed at the outset of this investigation, statistical measures were applied. In this chapter, statistical analyses are reported and discussed.

1. Can observers reliably rate the severity of misarticulations in the connected speech of subjects?

Two groups of observers, one trained, the other untrained, were chosen to evaluate the connected speech of children according to a seven-point scale of articulation defectiveness. The children's connected speech was obtained by having each subject read the screening form of the Sentence Articulation Inventory appropriate for his defective sound. Length of the connected speech to be evaluated was not limited as it was felt control for occurrence of the defective sound was more important. The group of trained judges consisted of

graduate students in speech pathology. The group of untrained judges, on the other hand, was composed of college students who had had no prior experience in evaluating articulation defectiveness. By using ten trained judges, the reliability of their evaluations was found to be .96. Reliability was determined by using the intra-class correlation coefficient for averages. However, it was necessary to use more untrained than trained judges to obtain the level of reliability desired at the outset of this investigation (.95). With forty-three untrained judges, the reliability was found to be .98. According to this study, observers can reliably rate the severity of articulation defectiveness in connected speech. It should be noted, however, that the number of trained judges was smaller than the number of untrained judges. There were ten judges in the trained group as compared to forty-three in the untrained group. This would seem to indicate that the trained listeners were more consistent in their judgments than the untrained judges. These findings are not consistent with those obtained in other studies. Alcorn (1971) found the evaluations of defective speech to be more reliable when using mothers and teachers as judges than when using speech pathology majors.

2. What is the relationship between the scores obtained from the single-word articulation tests and the ratings of listeners made from tapes of the subjects' connected speech?

To determine the relationship between scores obtained from the single-word articulation tests and the ratings of listeners made from

tapes of the subjects' connected speech, all raw scores were placed in rank order within its defective group. For example, the raw scores for those subjects having a defective /r/ were placed in one group. The same procedure was followed with those having a defective /s/. The rank-order for each raw score was then determined within each respective group. In order to interpret these ranks, Kendall's Tau-Correlation between ranks (Downie-Heath, 1959) was used. This statistic provides information about the degree to which two tests vary with each other in terms of rank-order, i.e., do they rank order individuals in a similar manner? A single-word articulation test which correlates highly with the listeners' judgments is one which rank-orders the subjects tested in a similar manner.

For the purpose of this study, using Kendall's Tau-Correlation between ranks, a correlation coefficient of .80 was designated as the level an articulation test and the listeners' judgments would be considered to have a substantial relationship. Nine articulation tests were used. These were the Templin-Darley Test of Articulation, the Goldman-Fristoe Test of Articulation, the Fisher-Logemann Test of Articulation, the Developmental Test of Articulation, the McDonald Screening Deep Test of Articulation, the McDonald Deep Test of Articulation, the Bryngleson-Glaspey Test of Articulation, the Predictive Screening Test of Articulation, and the Laradon Articulation Scale. Each child in the defective /r/ and defective /s/ groups was assigned two values relating to listener judgments. One value was obtained by using the median scale score for the group of trained

listeners. The other value was obtained by using the median scale score for the group of untrained listeners. An intercorrelation matrix was prepared for subjects with misarticulations for /r/, illustrating the coefficients for any single-word articulation test and listener judgments. This is shown in Table 2. An intercorrelation matrix was also prepared for subjects with misarticulations for /s/ which is shown in Table 3.

Examination of the matrix for /r/ defective children showed that only two of the nine single-word articulation tests used in this study had a substantial relationship with the trained listeners' judgments. These two tests were the Developmental Test of Articulation and the McDonald Deep Test of Articulation, both with a .90 Kendall Tau-Correlation coefficient. Three of the single-word articulation tests reached or exceeded the designated correlation with the judgments of untrained listeners. These were the McDonald Screening Deep Test of Articulation, the Bryngleson-Glaspey Test of Articulation, and the Predictive Screening Test of Articulation. The correlations of these tests to the judgments of untrained listeners were .80, .90, and .80, respectively.

For the /s/ defective children, only one single-word articulation test was found to have a significant correlation coefficient with the trained listeners' judgments. This was the McDonald Deep Test of

Table 2.--Intercorrelation Matrix Summarizing Kendall's Tau-Correlation Between Ranks for the Defective /r/ Group

	Templin-Darley Test of Articulation	Goldman-Fristoe Test of Articulation	Fisher-Logemann Test of Articulation	Developmental Test of Articulation	McDonald Screening Deep Test of Articulation	McDonald Deep Test of Articulation	Predictive Screening Test of Articulation	Bryngleson-Glaspey Test of Articulation	Laradon Articulation Scale
Trained Listener Judgments	.50	.60	.60	.90	.40	.90	.70	.40	.70
Untrained Listener Judgments	.60	.70	.70	.60	.80	.50	.80	.90	.70

____= Correlation reached the designated magnitude of 0.80

Table 3.--Intercorrelation Matrix Summarizing Kendall's Tau-Correlation Between Ranks for the Defective /s/ Group

	Templin-Darley Test of Articulation	Goldman-Fristoe Test of Articulation	Fisher-Logemann Test of Articulation	Developmental Test of Articulation	McDonald Screening Deep Test of Articulation	McDonald Deep Test of Articulation	Predictive Screening Test of Articulation	Bryngleson-Glaspey Test of Articulation	Laradon Articulation Scale
Trained Listener Judgments	.10	.10	.60	.00	.30	<u>.80</u>	.30	.10	.30
Untrained Listener Judgments	.30	.30	<u>.80</u>	.40	.30	<u>1.0</u>	.40	.30	.50

<u>___</u> = Correlation reached the designated magnitude of 0.80

Articulation. Two tests were found to have a substantial relationship with the judgments of the untrained listeners. The Fisher-Logemann Test of Articulation was found to have a .80 correlation coefficient, and the McDonald Deep Test of Articulation was found to have a correlation coefficient of 1.0. It is interesting to note that the ranks assigned by the untrained listeners had a significant correlation with more single-word articulation tests than did those by the trained listeners. However, since the trained listeners' judgments showed a significant agreement with only one single-word articulation test and the untrained listeners' judgments showed significant agreement with only two single-word articulation tests, this difference is one that could be expected to occur by chance.

Overall, there seems to be a low agreement among the ranks obtained from the single-word articulation tests and the listeners' judgments. It is concluded that the results of articulation tests are not comparable to the manner in which listeners evaluate the defectiveness of a subject's speech.

3. What is the relationship between the scores obtained from the sentence articulation inventory (SAI), the ratings of listeners made from tapes of the subjects' connected speech, and the scores obtained from single-word articulation tests?

In order to determine the relationship between the scores obtained from the SAI, the ratings of listeners made from tapes of the subjects' connected speech, and the scores obtained from the single-word articulation tests, all raw scores were again placed in rank order within its defective group. To meaningfully interpret these ranks, Kendall's

Coefficient of Concordance, (W), which is appropriate when the relationship among three or more sets of ranks is to be determined, was used. This statistical measure tells the overall degree of association between the ranks. The size of the coefficient of concordance indicates the magnitude of agreement. If there were no association whatever between the rater groups, and a rank-order coefficient of concordance were computed between the ranks, it would be near zero. If there was agreement between the ranks, the correlation coefficient would be significantly different from zero. Perfect agreement is indicated by a W = 1.0 and lack of agreement by a W = .00.

For this investigation, the desired Kendall W was set at .80. The Kendall Coefficient of Concordance, W, for the defective /r/ group was found to be .82. This would indicate that the agreement between the scores obtained from the SAI, the ratings of listeners made from tapes of the subjects' connected speech and the scores obtained from single-word articulation tests is high.

This was not found to be the case for the defective /s/ group. The Kendall Coefficient of Concordance was found to be .63. From this, it can be said that there was not a high agreement between the scores obtained from the SAI, the scores obtained from the single-word articulation tests, and the ratings made from the connected speech of the five subjects in the defective /s/ group. One reason for this lack of agreement (.63) in the defective /s/ group as opposed to substantial agreement (.82) in the defective /r/ group might lie in the construction of the SAI for the /s/ phoneme.

Another reason may be response variability of the listeners. Some feature present in the connected speech for the defective /s/ group made the evaluation too difficult. Isolation of this feature could prove to be a factor important in the diagnosis or correction of a defective /s/.

As was stated earlier, if the Kendall W was less than .80 for either the defective /r/ or /s/ group, the relationship of any two tests would be determined. This relationship would be determined by using the Kendall Tau-Correlation between ranks. Since the Kendall W for the defective /s/ group did not meet the designated magnitude of .80, the relationship of all test pairs was determined. The desired correlation coefficient was set at .80. A total of sixty-six, $\frac{n(n-1)}{2}$, Kendall taus were found. A matrix summarizing the inter-correlation of ranks for the defective /s/ group is shown in Table 4.

Scores from the SAI were not found to have a substantial agreement with any of the single-word articulation tests or the two groups of judges. The Templin-Darley Test of Articulation was found to have a .90 coefficient of agreement with five single-word articulation tests. These were the Goldman-Fristoe Test of Articulation, the Developmental Test of Articulation, the McDonald Screening Deep Test of Articulation, the Bryngleson-Glaspey Test of Articulation, and the Predictive Screening Test of Articulation. A high agreement, .90, was also found between the Goldman-Fristoe Test of Articulation, the Developmental Test of Articulation, the McDonald Screening Deep Test

Table 4.--Intercorrelation Matrix Summarizing Kendall's Tau-Correlation Between Ranks for /s/

	Templin-Darley Test of Articulation	Goldman-Fristoe Test of Articulation	Fisher-Logemann Test of Articulation	Developmental Test of Articulation	McDonald Screening Deep Test of Articulation	McDonald Deep Test of Articulation	Bryngleson-Glaspey Test of Articulation	Predictive Screening Test of Articulation	Laradon Articulation Scale	Trained Listener Judgments	Untrained Listener Judgments
Sentence Articulation Inventory	.30	.30	.30	.30	.20	.50	.10	.30	.60	.50	.50
Templin-Darley Test of Articulation		.90	.50	.90	.90	.40	.90	.90	.60	.10	.30
Goldman-Fristoe Test of Articulation			.50	.90	.90	.40	.90	.90	.60	.10	.30
Fisher-Logemann Test of Articulation				.50	.50	.70	.50	.60	.60	.60	.80
Developmental Test of Articulation					.70	.20	.70	.80	.50	.00	.40
McDonald Screening Deep Test of Articulation						.30	.90	.90	.60	.30	.30
McDonald Deep Test of Articulation							.30	.40	.50	.80	1.0
Bryngleson-Glaspey Test of Articulation								.90	.60	.10	.30
Predictive Screening Test of Articulation									.70	.30	.40
Laradon Articulation Scale										.30	.50
Trained Listener Judgments											.80

____= Correlation reached the designated magnitude of 0.80

of Articulation, the Bryngleson-Glaspey Test of Articulation, and the Predictive Screening Test of Articulation. Scores from the Fisher-Logemann Test of Articulation Competence did not have a significant agreement with any of the single-word articulation tests, but they did with those from the group of untrained judges. The Fisher-Logemann Test of Articulation Competence and the group of naive judges had a .80 correlation coefficient. The Developmental Test of Articulation was found to have a .80 correlation coefficient with the Predictive Screening Test of Articulation in addition to its .90 correlation with the Templin-Darley Test of Articulation and the Goldman-Fristoe Test of Articulation which was mentioned previously. The McDonald Screening Deep Test of Articulation was found to have a .90 correlation coefficient with the Bryngleson-Glaspey Test of Articulation and the Predictive Screening Test of Articulation. It also had an agreement of .90 with the Templin-Darley Test of Articulation and the Goldman-Fristoe Test of Articulation. The McDonald Deep Test of Articulation did not have a substantial agreement with any of the single-word articulation tests, but it did have with the two groups of judges. It had a .80 agreement with the group of trained judges and a perfect agreement, 1.0, with the group of untrained judges. A .90 correlation coefficient was found between the ranks on the Bryngleson-Glaspey Test of Articulation and the Predictive Screening Test of Articulation as well as the Templin-Darley, the Goldman-Fristoe, and the McDonald Screening Deep Test of Articulation which was mentioned above. The Predictive Screening Test of Articulation was not found to have a significant amount of agreement with any

tests other than those mentioned previously. These were the Templin-Darley, the Goldman-Fristoe, the McDonald Screening Deep Test, and the Bryngleson-Glaspey Test of Articulation, all four having a correlation coefficient of .90, and the Developmental Test of Articulation having a .80 correlation coefficient. The Laradon Articulation Scale, like the SAI, had no significant correlation between ranks with any of the single-word articulation tests for either of the two groups of judges. The correlation between the ranks given by the trained and untrained judges was .80.

To summarize, sixty-six Kendall taus were computed showing the relationship of any test pairs for the defective /s/ group. Seventeen test pairs were found to have a correlation between ranks of .80 or higher. It is interesting to note that the trained judges had a significant correlation with only one of the single-word articulation tests, the McDonald Deep Test of Articulation, while the untrained judges had a significant correlation with two of the single-word articulation tests, the Fisher-Logemann Test of Articulation Competence and the McDonald Deep Test of Articulation. Six single-word articulation tests were included in the test pairs having significant agreement. These were the Templin-Darley Test of Articulation, the Goldman-Fristoe Test of Articulation, the Developmental Test of Articulation, the McDonald Screening Deep Test of Articulation, the Bryngleson-Glaspey Test of Articulation, and the Predictive Screening Test of Articulation. Therefore, if the speech clinician must choose a single-word articulation test to determine the defectiveness of a

child's misarticulations, one of these six tests would be the most valid tool.

4. What is the relationship between the scores obtained from the sentence articulation inventory and the ratings of listeners made from tapes of the subjects' connected speech?

The Kendall Tau-Correlation between ranks (Downie-Heath, 1959) was used to determine the relationship between the scores obtained from the SAI and the ratings of the listeners made from tapes of the subjects' connected speech. This was done for both the defective /r/ and the defective /s/ groups. All raw scores were placed in rank order within its defective group. The Kendall tau tells the relationship between these ranks.

A correlation coefficient of .80 was specified as the level at which the SAI and the ratings of the listeners made from the tapes of the subjects' connected speech would be regarded as having significant agreement.

Both the trained group and the untrained group of listeners were found to have a correlation coefficient of .80 with the SAI. This can be interpreted to mean that the SAI and the trained judges and the SAI and the untrained judges rank the subjects in the defective /r/ group in a similar manner. Consider these two statements:

a) How listeners evaluate a person's connected speech is considered to be the ultimate measure of his communicative ability.

b) The SAI has a high correlation with the judgments of trained and untrained listeners.

Therefore, the SAI for the /r/ phoneme would indicate a person's articulation ability in connected speech. The SAI for the /r/ phoneme would also represent the manner in which listeners would evaluate his speech.

For the defective /s/ group, the SAI and the evaluations of trained listeners were found to have a correlation coefficient of .50. The same was true with the SAI and the group of untrained listeners. A correlation coefficient of .50 was again determined. This was not a high enough correlation to be considered significant. Neither of these pairs, the SAI for the /s/ phoneme and the trained judges and the SAI for /s/ and the untrained judges, ranked the subjects in a similar manner.

Even though this task was able to be done with the defective /r/ group, it was not able to be done with the defective /s/ group. This may be due to a possible difference in the degrees of severity represented within each defective group. For example, if there was a greater degree of severity present in one defective group, the task of scaling articulation defectiveness would probably have been easier than if the range of severity were small. If the samples in the defective group were within a small range, the misarticulations would be more similar and, consequently, more difficult to scale. If this were true of the defective /s/ group, the speech samples would have been more difficult to scale. In this study, one child in the defective /s/ group had errors in the class of substitutions while the other four had errors in the class of distortions. The difference might also

have been in the listeners. Alcorn (1971) found in his study that there were degrees of acceptability of articulation errors within a class of errors. According to this present study, it seems that there might be a difference in the acceptability of these errors between various listeners. A third possibility which might account for this low correlation is the construction of the SAI for the /s/ phoneme. A Kendall tau for the two groups of listeners showing the correlation between ranks was found to be .80. This would mean that both the trained and the untrained listeners agreed significantly in the manner in which they ranked the subjects. Yet these groups did not agree with the SAI. Since the two groups were able to agree in their ranks of the subjects, but were unable to agree with the SAI, the reason might lie in the construction of the SAI.

5. What is the relationship between scores obtained from the sentence articulation inventory and the scores obtained from the single-word articulation tests?

In determining the relationship between scores obtained from the SAI and the scores obtained from the single-word articulation tests, the subjects were divided into two groups based on their defective phoneme. Their raw scores were then placed in rank order within each group. To interpret these ranks, Kendall's Tau-Correlation between ranks (Downie-Heath, 1959) was used. This tells the degree to which two tests differ in terms of rank-order.

In this investigation, a correlation coefficient of .80 was specified as the level the SAI and a single-word articulation

test would be considered to have a significant relationship. An intercorrelation matrix containing the correlation coefficients for the SAI and the single-word articulation tests for subjects with misarticulations for /r/ and for /s/ is shown in Table 5.

A study of the matrix for the defective /r/ group showed that the SAI had a substantial relationship with all of the single-word articulation tests used in this study, with the exception of the Bryngleson-Glaspey Test of Articulation. The SAI and the Predictive Screening Test of Articulation had perfect agreement, shown by a Kendall tau of 1.0.

Investigation of the matrix for the defective /s/ group revealed no significant coefficients between the ranks for the SAI and the single-word articulation tests. None of the single-word articulation tests ranked the subjects in the /s/ group in the same manner as the SAI. The highest correlation was found to be with the McDonald Deep Test of Articulation, having a .50 coefficient, and the Laradon Articulation Scale, with a .60 coefficient. The reason for this lack of agreement would seem most likely to be due to the difference in the construction of the tests. The single-word articulation tests contained and tested all sounds whereas the SAI tested only one phoneme but in many contexts.

Table 5.--Intercorrelation Matrix Summarizing Kendall's Tau-Correlation Between Ranks for SAI and Nine Single-Word Articulation Tests

	Templin-Darley Test of Articulation	Goldman-Fristoe Test of Articulation	Fisher-Logemann Test of Articulation	Developmental Test of Articulation	McDonald Screening Deep Test of Articulation	McDonald Deep Test of Articulation	Bryngleson-Glaspey Test of Articulation	Predictive Screening Test of Articulation	Laradon Articulation Scale
Sentence Articulation Inventory /r/	.80	.90	.90	.80	.80	.80	.70	1.0	.90
Sentence Articulation Inventory for the /s/ Phonem	.30	.30	.30	.30	.20	.50	.10	.30	.60

____ = Correlation reached the designated magnitude of 0.80

CHAPTER V

SUMMARY AND CONCLUSIONS

The purpose of this study was to determine the relationship between scores obtained from a sentence articulation inventory (SAI) to scores obtained from various single-word articulation tests and listener judgments of the subjects connected speech. The SAI, which was developed for the /r/ and /s/ phonemes, controlled for sentence length, frequency of occurrence of words, and phonetic context.

Ten subjects, five having misartioulations for /r/ and five having misarticulations for /s/, were used in this study. All were between the ages of seven and eleven, had passed a hearing screening test administered at 20 dB for the frequencies of 250, 500, 1000, 2000, and 4000 Hz, and had "average" intelligence. The subjects were considered to have an "average" intelligence if they were not presently or had not, in the past, been enrolled in EMH or learning disability classes.

All subjects were administered the SAI for his defective sound and nine single-word articulation tests. These tests included the following: the Templin-Darley Test of Articulation, the Goldman-Fristoe Test of Articulation, the Fisher-Logemann Test of Articulation Competence, the Developmental Test of Articulation, the McDonald Screening Deep Test of Articulation, the McDonald Deep Test of Articulation, the Bryngleson-Glaspey Test of Articulation, the Predictive

Screening Test of Articulation, and the Laradon Articulation Scale. A sample of each subject's connected speech was also obtained by recording his speech while reading the screening portion of the SAI. The presentation of all tests and the obtaining of the speech sample was randomized for each child to allow for any improvement which might occur in the child's articulatory behavior due to the testing. All testing and the recording of the child's connected speech were completed within the same day for each subject.

Two groups of listeners were used as judges to evaluate the connected speech of the ten subjects. One group was composed of trained listeners who were graduate students in speech pathology, trained and experienced in diagnosing defective speech. The second group was classified as untrained listeners. This group consisted of college students who had had no experience in evaluating the defectiveness of speech. These two groups evaluated the defectiveness of the subjects' connected speech on a scale of equal intervals. The scale contained seven intervals, one representing the lowest degree of articulation defectiveness and seven representing the highest degree. The interval four was the middle between one and seven in degree with the other numbers following at equal distances along the scale. Instructions were given to the judges explaining the evaluating procedure. Each judge recorded their evaluations on an individual score sheet.

Raw scores obtained from the SAI, the single-word articulation tests, and the listener judgments were placed into two groups on the basis of the defective phoneme. These raw scores were then placed in

rank order within their respective group. In order to interpret these ranks in answer to the questions posed at the outset of this investigation, statistical measures were used.

At the outset of this investigation, the research hypothesis was stated as thus: Scores obtained from a sentence articulation inventory are more representative of a person's conversational speech and correlate more highly with listener judgments of connected speech than scores obtained from single-word articulation tests.

Correlation coefficients representing the degree with which the SAI and the single-word articulation tests were able to predict the manner in which the listeners would evaluate the subjects' connected speech were computed by using Kendall's Tau-Correlation between ranks. Results for the defective /r/ group support the research hypothesis. Scores obtained from the SAI were more representative of the subjects' conversational speech and correlated more highly with listener judgments of connected speech than scores obtained from single-word articulation tests. Ranks from the SAI and both groups of judges had a correlation coefficient of .80, which was considered to be a substantial agreement. Two single-word articulation tests, the Developmental Articulation Test and the McDonald Deep Test of Articulation, had a correlation coefficient of .90 with the group of trained listeners. The McDonald Screening Deep Test of Articulation, the Bryngleson-Glaspey Test of Articulation, and the Predictive Screening Test of Articulation had correlation coefficients of .80, .90, and .80, respectively, with the group of untrained listeners. Four single-word

articulation tests did not have significant coefficients with either the trained or untrained listeners.

Although three of the single-word articulation tests, the Developmental Articulation Test, the McDonald Deep Test of Articulation, and the Bryngleson-Glaspey Test of Articulation, had a higher correlation with listener judgments than the SAI, the correlation was significant with only one of the two groups of listeners. The SAI had a significant agreement of ranks with both the trained and untrained listeners. According to this, the SAI for the /r/ phoneme is representative of a person's conversational speech and predicts the manner in which listeners will evaluate his articulatory defectiveness.

Correlation coefficients obtained for the defective /s/ group show somewhat different results. Coefficients showing the relationship between ranks for the SAI and the trained listeners did not reach the magnitude designated as that considered to be significant, .80. Instead, a .50 correlation coefficient was obtained for both comparisons. Even though this was not considered to be a substantial relationship, it was higher than most of the other correlations between single-word articulation tests and listener judgments. Those coefficients which were below this ranged from .00 to .40. Only two of the single-word articulation tests had a higher correlation coefficient. These were the Fisher-Logemann Test of Articulation Competence and the McDonald Deep Test of Articulation. The first had a .80 correlation with the untrained listeners. The McDonald

Deep Test of Articulation had a .80 correlation with the trained listeners and a 1.0 correlation with the untrained listeners. Again, even though the SAI for the /s/ phoneme did not obtain substantial agreement with the trained or untrained listeners, scores from this did correlate more highly with listener judgments of connected speech than scores obtained from most of the single-word articulation tests.

The Kendall Tau-Correlation between ranks was also used to show the relationship between the SAI and the various single-word articulation tests. The SAI for the /r/ phoneme was found to have a .70 correlation with the Bryngleson-Glaspey Test of Articulation, a .80 correlation with the Templin-Darley Test of Articulation, the Developmental Articulation Test, the McDonald Screening Deep Test of Articulation, and the McDonald Deep Test of Articulation, a .90 correlation with the Goldman-Fristoe Test of Articulation, the Fisher-Logemann Test of Articulation Competence, and the Laradon Articulation Scale, and a 1.0 correlation with the Predictive Screening Test of Articulation. All coefficients were interpreted as showing substantial agreement with the exception of the .70 coefficient for the Bryngleson-Glaspey Test of Articulation. The SAI for the /s/ phoneme, on the other hand, did not show a significant relationship with any of the single-word articulation tests. The tests and their correlation coefficients were as follows: Templin-Darley Test of Articulation, .30; Goldman-Fristoe Test of Articulation, .30; Fisher-Logemann Test of Articulation Competence, .30; Developmental Articulation Test, .30; McDonald Screening Deep

Test of Articulation, .20; McDonald Deep Test of Articulation, .50; Bryngleson-Glaspey Test of Articulation, .10; Predictive Screening Test of Articulation, .30; and the Laradon Articulation Scale, .60.

Kendall's Coefficient of Concordance, showing the relationship among the ranks from the SAI, the single-word articulation tests, and the listener judgments for all subjects within their defective groups, showed the relationship for the defective /r/ group to be substantial. However, the Kendall Coefficient of Concordance, W, was not found to be significant for the defective /s/ group.

It is felt that the SAI for the /r/ phoneme is the better of the two in terms of its value as a predictor of the manner in which listeners will evaluate the speaker's articulatory behavior. The SAI for the /r/ phoneme also obtains similar results with most of the single-word articulation tests used in this study.

From this study, it was also determined that observers can reliably rate the severity of misarticulations in the connected speech of subjects. The group of trained judges was able to rate the defectiveness of the subjects' connected speech with a .96 level of reliability. The group of untrained listeners was able to do this with a .98 level of reliability. By using two groups of listeners, one trained, the other naive, it was shown that the trained listeners were better able to perform the task than were the untrained listeners. A smaller number of trained listeners, ten, was able to obtain a significant level of reliability than that required for the untrained listeners, forty-three.

In conclusion of this study, the following statements can be made:

1. Scores obtained from a SAI, particularly the one developed for the /r/ phoneme, are more representative of a person's conversational speech and correlate more highly with listener judgments of connected speech than scores obtained from many single-word articulation tests.

2. Scores obtained from the SAI for the /r/ phoneme correlate significantly with scores obtained from the following single-word articulation tests: Templin-Darley Test of Articulation, Goldman-Fristoe Test of Articulation, Fisher-Logemann Test of Articulation Competence, Developmental Articulation Test, McDonald Screening Deep Test of Articulation, McDonald Deep Test of Articulation, Predictive Screening Test of Articulation, and Laradon Articulation Scale.

3. Scores obtained from the SAI for the /s/ phoneme do not have a substantial relationship with scores obtained from the single-word articulation tests used in this study.

4. The relationship between the scores obtained from the SAI for the /r/ phoneme, the ratings of listeners made from tapes of the subjects' connected speech, and the scores obtained from single-word articulation tests was found to be significant.

5. The relationship between the scores obtained from the SAI for the /s/ phoneme, the ratings of listeners made from tapes of the subjects' connected speech, and the scores obtained from single-word articulation tests was not found to be significant.

6. Listeners can reliably rate the severity of the misarticulations in the connected speech of subjects.

Implications for further study:

1. Replication of this study using the /s/ phoneme.

2. Replication of this study using a larger sample size.

3. Modification of this study by using different error phonemes and sentence articulation inventories developed for those phonemes.

SELECTED BIBLIOGRAPHY

Alcorn, Steven F., "Comparisons of Articulation Severity Ratings of /s/ and /r/ by Lower-, Middle-, and Upper-Socio Economic Groups." Master's Thesis, Department of Speech Pathology and Audiology. Eastern Illinois University, 1971.

Amerman, Daniloff, and Moll. "Lip and Jaw coarticulation for the phoneme /æ/." J. Speech Hearing Res., 13, 1970, 147-69.

Anderson, V. A. Improving the Child's Speech. New York: Oxford University Press, 1953.

Barker, Janet O'Neill. "A numerical measure of articulation." J. Speech Hearing Dis., 25, 1960, 79-88.

Bryngleson, F. and Glaspey, E. Speech in the Classroom. Chicago: Scott, Foresman and Co., 1957.

Curry, R., Kennedy, L., Wagner, L., and Wilke, W. "A phonographic scale for the measurement of defective articulation." J. Speech Hearing Dis., 8, 1943, 123-26.

Daniloff, R. and Moll, K. "Coarticulation and Lip Rounding." J. Speech Hearing Res., 11, 1968, 707-21.

Dorn, John E. "The /r/ phoneme and phonetic context." Unpublished paper Eastern Illinois University, February 26, 1973.

Downie, N. M. and Heath, D. W., Basic Statistical Methods, third edition, New York: Harper & Row, 1970.

Edmonston, W. Laradon Articulation Scale. Beverly Hills, California: Western Psychological Services, 1963.

Fairbanks, G. Voice and Articulation Drillbook, 2nd edition. New York: Harper and Row, Publishers, 1940.

Faircloth, M. A., and Faircloth, S. R. "An analysis of the articulatory behavior of a speech-defective child in connected speech and in isolated-word responses." J. Speech Hearing Dis., 35, 1970, 51-61.

Fisher, H. and Logemann, J. The Fisher-Logemann Test of Articulation Competence. Boston: Houghton-Mifflin Co., 1971.

Fleming, Katherine Jane. "Guidelines for choosing appropriate phonetic contexts for speech-sound recognition and procedure practice." J. Speech Hearing Dis., 36, 1971, 356-67.

Garret, Edgar, and Buros, Oscar. Mental Measurement Yearbook. New Jersey: The Gryphon Press, 2, 1972, 951.

Griffith, J. R. and Miner, L. E. "A phonetic context approach to articulation therapy." Paper presented at the Illinois Speech and Hearing Association Convention, Arlington Heights, Illinois, 1973.

Griffith, J. R. and Miner, L. E. "Building articulation skills: phonetic context and Zipf's Law." Illinois Speech Hearing J., 6, 1973.

Griffith, J. R. and Miner, L. E. Building Basic Articulation Skills. Chicago: Bell and Howell, 1973.

Goldman, R. and Fristoe, M. Goldman-Fristoe Test of Articulation. Circle Pines, Minn.: American Guidance Service, 1969.

Guildford, J. P. Psychometric Methods. New York: McGraw-Hill Pub. Co., 1954.

Hardy, J. C. "A phonetic study of misarticulation of /r/." Master's Thesis, University of Iowa, 1957.

Hejna, R. Articulation Testing Cards. Ann Arbor: Speech Materials, 1955.

Henrickson, E. H. "An analysis of Wood's Articulation Index." J. Speech Hearing Dis., 13, 1948, 233-35.

Jordan, E. P. "Articulation test measures and listener ratings of articulation defectiveness." J. Speech Hearing Res., 3, 1960, 303-19.

Kerlinger, Fred N. Foundations of Behavioral Research. New York: Holt, Rinehart, and Winston, Inc., 1964.

McDonald, E. T. A Deep Test of Articulation. Pittsburgh: Stanwix House, 1964.

McDonald, E. T. A Screening Test of Articulation. Pittsburgh: Stanwix House, 1968.

Morrison, Shelia. "Measuring the severity of articulation defectiveness." J. Speech Hearing Dis., 20, 1955, 347-51.

Nelson, J. T. "A study of misarticulation of /s/ in combinations with selected vowels and consonants." Master's Thesis, University of Iowa, 1945.

Perrin, E. H. "The rating of defective speech by trained and untrained observers." J. Speech Hearing Dis., 19, 1954, 48-51.

Peterson, Harold, andBuros, Oscar. Mental Measurements Yearbook. New York: The Gryphon Press, 2, 1972, 952.

Powers, Margaret. "Functional disorders of articulation: symtomatology and etiology." Handbook of Speech Pathology and Audiology. Edited by Lee Edward Travis, New York: Appleton-Century-Crofts, 1971.

Roe, M. V. and Milisen, R. "The effect of maturation upon defective articulation in elementary grades." J. Speech Hearing Dis., 7, 1942, 37-45.

Schneider, C. "Phonetic Context and Articulation Ability." Master's Thesis, Department of Speech Pathology and Audiology, Eastern Illinois University, 1973.

Sherman, Dorothy and Cullinan, W. L. "Several procedures for scaling articulation," J. Speech Hearing Res., 22, 1957, 698-706.

Sherman, D. and Morrison, S. "Reliability of individual ratings of severity of defective articulation." J. Speech Hearing Dis., 20, 1955, 352-58.

Siegel, Sidney. Nonparametric Statistics for the Behavioral Sciences. New York: McGraw-Hill Book Company, Inc., 1956.

Stetson, R. H. Motor Phonetics: A Study of Speech Movements in Action. 2nd. edition. Amsterdam: North-Holland Pub. Co., 1951 (for Oberlin College).

Templin, Mildred C. "Spontaneous versus imitated verbalization in testing articulation in preschool children." J. Speech Hearing Dis., 12, 1947, 293-300.

Templin, M. and Darley, F. The Templin-Darley Tests of Articulation. 2nd. ed., Iowa City: University of Iowa, 1969.

Van Riper, C. and Erickson, R. "A Predictive Screening Test of Articulation." J. Speech Hearing Dis., 34, 1969, 214-19.

Voelker, C. H. "Phonetic distribution in formal American Pronounciation." J. Acoust. Soc. Amer., 5, 1934, 242-46.

Winitz, Harris. Articulatory Acquisition and Behavior. New York: Appleton-Century-Crofts, 1969.

Wood, K. S. "Measurement of progress in correction of articulatory speech defects." J. Speech Hearing Dis., 14, 1949, 1971-74.

Wright, H. N. "Reliability of evaluations during basic articulation and stimulation testing." J. Speech Hearing Dis., Monogr.

APPENDIX

OBSERVERS' SCORE SHEET

Scale of Articulation Defectiveness

least highest

1. ____
2. ____
3. ____
4. ____
5. ____
6. ____
7. ____
8. ____
9. ____
10. ____

INSTRUCTIONS

You are asked to judge a group of sentences which are read by children. You are to judge each group of sentences read by each child in relation to a seven-point scale of articulation defectiveness. Articulation defectiveness is defined as the degree to which the misarticulations interfere with the communication process. Make your judgments solely on articulation defectiveness, not on the basis of reading difficulty.

The scale is one of equal intervals from one to seven, with one representing the lowest degree of articulation defectiveness and seven representing the highest degree. The interval four is the middle between one and seven in degree with the other numbers following at equal distances along the scale. Do not attempt to place segments between any of these seven points, but only at these points.

After hearing the group of sentences read by each child, you will record immediately the number of the scale position you think the sample should have. You will record your scale number to the right of each group number on your answer sheet. Notice that you will start at the top of the column and work down toward the bottom.

Before you record any judgments, listen to the entire set of grouped sentences to acquaint yourself with the experimental task and to acquaint yourself with the range of samples with respect to the degree of articulation defectiveness, which you are asked to judge.

After you have acquainted yourself with the task and the range, make a judgment on every sample. If you are somewhat doubtful, make a guess as to the most suitable scale position.

Are there any questions? Ready...

SENTENCE ARTICULATION INVENTORY
FOR THE PHONEME /r/

____ 1. We played tug-of-war and ran foot races.
F [ɔr] /A I [ræ] /A I [re] /A

____ 2. I was born in March.
F [rn] /A F [rtʃ] /A

____ 3. Is that red rose real?
I [rɛ] /A I [ro] /A I [ri] /A

____ 4. Start at the large rock.
F [rt] /A F [rdʒ] /A I [ra] /A

____ 5. Run around the tree and come back to this mark.
I [rʌ] /A I [rəu] /A I [tr] /A F [rk] /A

____ 6. Bring some wood so we can make a fire.
I [br] /A F [aɪr] /A

____ 7. Ride north until you reach Green Street.
I [raɪ] /A F [rθ] /A I [gr] /A I [str] /A

____ 8. Carry only three glasses at a time.
F [ær] /A I [θr] /A

____ 9. You must have looked in the wrong room.
I [rɔ] /A I [ru] /A

____ 10. Henry brought some fresh fruit from the farm.
F [rɪ] /UA I [fr] /A F [rm] /A

____ 11. We keep our horse in the country.
F [aur] /A F [rs] /A I [tr] /UA

____ 12. I must press your dress.
I [pr] /A F [ɹr] /A I [dr] /A

____ 13. Several hundred people were at the party.
I [rə] /UA I [dr] /UA F [ar] /A

____ 14. We live near the railroad.
F [ɪr] /A I [ro] /UA

____ 15. Will you prepare dinner for me?
I [pr] /UA I [or] /A

____ 16. This chair is really hard.
F [ɛr] /A F [rd] /A

____ 17. We can't cross the river in the spring.
I [kr] /A I [rɪ] /A I [spr] /A

Number Correct
46

SUMMARY OF MISARTICULATION PATTERNS FOR THE PHONEME /r/

VOWEL COMBINATIONS

I/A

M.F.*	Con.
Front	ri rɪ re rɛ ræ
Back	ra rɔ ro ru
Central	rʌ
Diphthongs	raɪ raʊ

I/UA

M.F.	Con.
Front	rɪ
Back	ro
Central	rə

F/A

M.F.	Con.
Front	ɪr ɛr æ r
Back	ar ɔ r or ʊr
Diphthong	aɪr aʊr

BLENDS

I/A

M.F.	Con.
Stop	pr br tr dr kr gr
Fricative Stop	spr str
Fricative	fr fθ

I/UA

M.F.	Con.
Stop	pr br dr

F/A

M.F.	Con.
Stop	rt rd rk
Fricative	rs rθ rtʃ
Affricative	rdʒ
Nasal	rm rn

* Manner of Formation

1. We played tug-of-war and ran foot races.
2. I was born in March.
3. Is that red rose real?
4. Start at the large rock.
5. Run around the tree and come back to this mark.
6. Bring some wood so we can make a fire.
7. Ride north until you reach Green Street.
8. Carry three glasses at a time.
9. You must have looked in the wrong room.
10. Henry brought some fresh fruit from the farm.
11. We keep our horse in the country.
12. I must press your dress.
13. Several hundred people were at the party.
14. We live near the railroad.
15. Will you prepare dinner for us?
16. This chair is really hard.
17. We can't cross the river in the spring.

SENTENCE ARTICULATION INVENTORY
FOR THE PHONEME /s/

____ 1. I will be seven years old next summer.
I [sɛ]/A F[kst]/A I[sʌ]/A

____ 2. Our house is across the street from the school.
F[aus]/A F[ɔs]/A I[str]/A I[sk]/A

____ 3. One side of this box is a square.
I[saɪ]/A F[ks]/A I[skw]/A

____ 4. Let's race my horse against yours.
F[es]/A F[rs]/A F[nst]/A

____ 5. I want a special dress for the spring dance.
I[sp]/A F[ɛs]/A I[spr]/A F[ns]/A

____ 6. Perhaps the storm kept us from going to sleep.
F[ps]/A I[st]/A F[ʌs]/A I[sl]/A

____ 7. It's s ppose to be a surprise party.
F[ts]/A I[sɹ]/UA

____ 8. Save me a small piece of cake if possible.
I[se]/A I[sm]/A F[is]/A F[as]/A

____ 9. Ask Mother where the new snow suit is.
F[sk]/A I[sn]/A I[su]/A

____ 10. Miss Price spoke in a soft voice.
F[ɹs]/A F[aɪs]/A I[sɔ]/A F[ɔɪs]/A

____ 11. She will explain the story to the class.
I[spl]/A F[æ s]/A

____ 12. We sat on a bench in the city park.
I[sæ]/A I[sɪ]/A

____ 13. All the best has been sold.
F[st]/A I[so]/A

____ 14. What else did you want besides sweet corn?
F[ls]/A I[sw]/A

____ 15. It is also necessary to practice very hard.
F[so]/UA I[sɛ]/UA F[ʌs]/UA

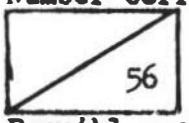
____ 16. I suggest we serve vario s cookies also.
I[sə]/UA I[sɝ]/A F[əs]/UA F[so]/UA

____ 17. The scene of the forest was nice.
I[si]/A F[st]/UA

____ 18. Another word for 'a short distance' is close.
F[ns]/UA F[os]/A

____ 19. The soil in the south does not produce well.
I [sɔɪ]/A I [sau]/A F [us]/A

Number correct



Possible correct

SUMMARY OF MISARTICULATION PATTERNS
VOWEL COMBINATIONS

I/A

M.F.*	Con.
Front	si sɪ se sɛ sæ
Back	sɔ so su
Central	sʌ sɝ
Dipthongs	saɪ sau sɔɪ

I/UA

M.F.	Con.
Front	sɛ
Back	so
Central	sɜ sə

F/A

M.F.	Con.
Front	is ɪs es ɛs æs
Back	as ɔs os us
Central	ʌs
Dipthongs	aɪs aus ɔɪs

F/UA

M.F.	Con.
Front	ɪs
Central	əs

BLENDS

I/A

M.F.	Con.
Stop	sp sk st*
Stop and Glide	spl spr skw str*
Nasal	sm sn
Glide	sl sw

F/A

M.F.	Con.
Stop	sk st* ps ts ks*
	kst nst*
Nasal	ns
Glide	rs ls

F/UA

M.F.	Con.
Stop	st
Nasal	ns

* Manner of Formation

1. I will be seven years old next summer.
2. Our house is across the street from the school.
3. One side of this box is a square.
4. Let's race my horse against yours.
5. I want a special dress for the spring dance.
6. Perhaps the storm kept us from going to sleep.
7. It's suppose to be a surprise party.
8. Save me a small piece of cake if possible.
9. Ask Mother where the new snow suit is.
10. Miss Price spoke in a soft voice.
11. She will explain the story to the class.
12. We sat on a bench in the city park.
13. All the best has been sold.
14. What else did you want besides sweet corn?
15. It is also necessary to practice very hard.
16. I suggest we serve various cookies also.
17. The scene of the forest was nice.
18. Another word for 'a short distance' is close.
19. The soil in the south does not produce well.

SENTENCE ARTICULATION INVENTORY
FOR THE PHONEME /l/

____ 1. There are twelve children in my English class.
F[ɛlv]/A F[ld]/A I[gl]/UA I[kl]/A

____ 2. Please give the girl a glass of milk.
I[pl]/A F[rl]/A I[gl]/A F[lk]/A

____ 3. The family went to an island to live.
I[lɪ]/UA I[lə]/UA I[lɪ]/A

____ 4. I'll explain the rule to Bill.
F[aɪl]/A I[spl]/A F[ul]/A F[rl]/A

____ 5. I did not realize you built it yourself.
I[laɪ]/UA F[lt]/A F[lf]/A

____ 6. There was a beautiful large lake on our land.
F[ɔl]/UA I[lɔ]/A I[le]/A I[læ]/A

____ 7. We will all probably sleep on the floor.
F[ɔl]/A I[bl]/UA I[sl]/A I[fl]/A

____ 8. They must learn to help themselves.
I[lɜ]/A F[lp]/A F[lvs]/A

____ 9. Get in line for follo -the-leader.
I[laɪ]/A F[ɑl]/A I[li]/A

____ 10. Who else did ell in health?
I[ɛl]/A F[ɛl]/A F[lθ]/A

____ 11. The soil is almost black in color.
F[ɔɪl]/A I[ɔl]/A I[bl]/A F[ʌl]/A

____ 12. Will he allow us to lo er the sail?
I[lau]/A I[lo]/A F[el]/A

____ 13. Did you lose the letter from London?
I[lu]/A I[lɛ]/A I[lʌ]/A

____ 14. The whole lot was part of the deal.
F[ol]/A I[la]/A F[il]/A

____ 15. He shall have to look for it.
F[æl]/A I[lʊ]/A

____ 16. Books are full of knowledge.
F[ʊl]/A F[lɪ]/UA

Number correct

52

Possible correct

SUMMARY OF MISARTICULATION PATTERNS FOR THE PHONEME /l/

VOWEL COMBINATIONS

I/A

M.F.*	Con.
Front	li lɪ lɛ le læ
Back	lɑ lɔ lo lʊ lu
Central	lʌ lɝ
Dipthong	laɪ laʊ

I/UA

M.F.	Con.
Front	lɪ
Central	la
Dipthong	laɪ

F/A

M.F.	Con.
Front	il ɪl el ɛl æl
Back	ɑl ɔl ol ʊl ul
Central	ʌl ɝl
Dipthong	aɪl ɔɪl

F/UA

M.F.	Con.
Front	ɪl
Back	ɔl
Central	əl

BLENDS

I/A

M.F.	Con.
Stop	bl pl kl gl
Fricative and Stop	spl
Fricative	sl fl

I/UA

M.F.	Con.
Stop	gl bl

F/A

M.F.	Con.
Stop	ld lt lk lp
Fricative	ls lf lθ
Affricative & Fricative	lvs
Affricative	lv

1. There are twelve children in my English class.
2. Please give the girl a glass of milk.
3. The family went to an island to live.
4. I'll explain the rule to Bill.
5. I did not realize you built it yourself.
6. There was a beautiful large lake on our land.
7. We will all probably sleep on the floor.
8. They must learn to help themselves.
9. Get in line for follow-the-leader.
10. Who else did well in health?
11. The soil is almost black in color.
12. Will he allow us to lower the sail?
13. Did you lose the letter from London?
14. The whole lot was part of the deal.
15. He shall have to look for it.
16. Books are full of knowledge.

SCREENING FORM OF THE SENTENCE ARTICULATION INVENTORY

1. The red roses are near the tire.
2. The short horse was afraid of the train.
3. This house sat empty all summer.
4. The small stick rested against the strong box.
5. The class saw the wild animal.
6. The school will probably be built soon.
7. Allow Bill to complete his letter last.

Sentences 1-2: /r/ phoneme
Sentences 3-4: /s/ phoneme
Sentences 5-7: /l/ phoneme

FREQUENCY OF OCCURRENCE OF PHONETIC CONTEXTS FOR /r/, /s/, /l/ IN SCREENING FORM OF THE SAI

Hi	Moderate	Lo
I[rɛ] F[ɪr] F[rt] I[tr]	I[ro] F[ɚr] I[fr]	F[aɪr] F[rs]
F[ɪs] I[st] F[st]	I[s ʌ] I[str]	F[aus] I[sɛ] I[sm] F[nst] F[ks]
I[lɛ] F[ɪl] I[kl] F[ld]	I[læ] F[ul] F[ɔl] I[bl] F[lt]	F[laʊ] I[pl]

Table 6.--Matrix Showing Ranks for the Defective /r/ Group

Subjects	Sentence Articulation Inventory	Templin-Darley Test of Articulation	Goldman-Fristoe Test of Articulation	Fisher-Logemann Test of Articulation	Developmental Test of Articulation	McDonald Screening Deep Test of Articulation	McDonald Deep Test of Articulation	Bryngleson-Glaspey Test of Articulation	Predictive Screening Test of Articulation	Laradon Articulation Scale	Trained Listener Judgments	Untrained Listener Judgments	Sum of Ranks
DO_F	2	1	1.5	1.5	3	1.5	3	2	2	2.5	3	2	25
CC	5	5	5	5	5	5	5	4.5	5	5	5	5	59.5
RS	1	2	1.5	1.5	1	1.5	1.5	1	1	1	1.5	1	15.5
PC	4	4	4	4	4	3.5	4	3	4	4	4	3	45.5
DO_M	3	3	3	3	2	3.5	1.5	4.5	3	2.5	1.5	4	34.5
											Total		180.0

Kendall W = .82

Table 7.--Matrix Showing Ranks for the Defective /s/ Group

Subjects	Sentence Articulation Inventory	Templin-Darley Test of Articulation	Goldman-Fristoe Test of Articulation	Fisher-Logemann Test of Articulation Competence	Developmental Test of Articulation	McDonald Screening Deep Test of Articulation	McDonald Deep Test of Articulation	Bryngleson-Glaspey Test of Articulation	Predictive Screening Test of Articulation	Laradon Articulation Scale	Trained Listener Judgments	Untrained Listener Judgments	Sum of Ranks
CL	3.5	1	1	1	1	1	2	1	1	2	3	2	19.5
LH	2	2	2	4	2	2	4	2	2	2	4	4	32
TT	1	3.5	3.5	2	3	3.5	1	3.5	3	2	1	1	28
DK	5	5	5	5	4	5	5	5	5	5	5	5	59
SM	3.5	3.5	3.5	3	5	3.5	3	3.5	4	4	2	3	41.5
											Total		180.0

Kendall W = .63

Table 8.--Intercorrelation Matrix Summarizing Kendall's Tau-Correlation Between Ranks for /r/

	Templin-Darley Test of Articulation	Goldman-Fristoe Test of Articulation	Fisher-Logemann Test of Articulation	Developmental Test of Articulation	McDonald Screening Deep Test of Articulation	McDonald Deep Test of Articulation	Bryngleson-Glaspey Test of Articulation	Predictive Screening Test of Articulation	Laradon Articulation Scale	Trained Listener Judgments	Untrained Listener Judgments
Sentence Articulation Inventory	.80	.90	.90	.80	.80	.80	.70	1.0	.90	.80	.80
Templin-Darley Test of Articulation		.90	.90	.70	.80	.40	.60	.80	.70	.50	.60
Goldman-Fristoe Test of Articulation			.90	.70	.80	.60	.90	.90	.80	.60	.70
Fisher-Logemann Test of Articulation				.70	.80	.60	.90	.90	.80	.60	.70
Developmental Test of Articulation					.60	.90	.30	.60	.90	.90	.60
McDonald Screening Deep Test of Articulation						.50	.70	.80	.80	.40	.80
McDonald Deep Test of Articulation							.90	.70	.80	.90	.50
Bryngleson-Glaspey Test of Articulation								.80	.60	.40	.90
Predictive Screening Test of Articulation									.90	.70	.80
Laradon Articulation Scale										.70	.70
Trained Listener Judgments											.50

____= Correlation reached the designated magnitude of 0.80