

1971

Comparison of Articulation Severity Ratings of /s/ and /r/ by Lower-, Middle-, and Upper Socioeconomic Groups

Stephen R. Alcorn
Eastern Illinois University

Recommended Citation

Alcorn, Stephen R., "Comparison of Articulation Severity Ratings of /s/ and /r/ by Lower-, Middle-, and Upper Socioeconomic Groups" (1971). *Masters Theses*. 3994.
<https://thekeep.eiu.edu/theses/3994>

This is brought to you for free and open access by the Student Theses & Publications at The Keep. It has been accepted for inclusion in Masters Theses by an authorized administrator of The Keep. For more information, please contact tabruns@eiu.edu.

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.

July 30, 1971
Date

Author

I respectfully request Booth Library of Eastern Illinois University not allow my thesis be reproduced because _____

Date

Author

COMPARISON OF ARTICULATION SEVERITY RATINGS

OF /s/ AND /r/ BY LOWER-, MIDDLE-, AND

UPPER SOCIOECONOMIC GROUPS

(TITLE)

BY

STEPHEN R. ALCORN

THESIS

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF

MASTER OF SCIENCE

IN THE GRADUATE SCHOOL, EASTERN ILLINOIS UNIVERSITY
CHARLESTON, ILLINOIS

1971

YEAR

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

August 2, 1971
DATE

ADVISER

August 2, 1971
DATE

ACKNOWLEDGEMENTS

The writer would like to express his appreciation to the many individuals who have helped make this study possible.

Special appreciation goes to Dr. Lynn Miner, Mrs. Linda Huddlestun, Mr. Ben Ober, the writer's Aunt Ruth and his wife Monya for the help in "lightening the load" while carrying out this study.

Appreciation also goes to Dr. Jerry Griffith and Dr. Wayne Thurman for their help and guidance in the writing of this research.

The writer also says "thank you" to the children who served as subjects and the people who participated as judges. Without these groups this study would not have been possible.

TABLE OF CONTENTS

Chapter	Page
ACKNOWLEDGEMENTS	i
LIST OF TABLES	iii
I. STATEMENT OF THE PROBLEM	1
II. REVIEW OF THE LITERATURE	9
III. SUBJECTS, PROCEDURES, EQUIPMENT	27
IV. RESULTS AND DISCUSSION	34
V. SUMMARY AND CONCLUSIONS	48
APPENDIX I	53
BIBLIOGRAPHY	56

LIST OF TABLES

Table		Page
1.	Location of Judging Sessions, Sound Pressure Level of the Room, and the Number of Sessions at Each Location	32
2.	Reliability Levels and Number of Judges for Upper-, Lower-, and Middle Socioeconomic Status Populations	34
3.	Values of Skewness, Kurtosis, Mean, and Standard Deviation for Lateral and Frontal Lisps	37
4.	Mean Scale and <u>t</u> Values for Urban and Rural Populations for Overall and Specific Phoneme Errors	40
5.	Range of Mean Scale Value for Upper-, Middle-, and Lower Socioeconomic Groups	44

CHAPTER I

STATEMENT OF PROBLEM

Evaluation of articulation ability must be determined by a listener.

This means that determining a person's ability to articulate phonemes in conversational speech is primarily a perceptual event. Young (1969) states that "measurement of a speech disorder is primarily a perceptual event and the observer's response necessarily represents the final validation for any measurement."

Previous research has utilized observers to make judgments of articulation severity on a global basis (Reid, 1947; Curry, et. al., 1943; Morrison, 1954; Perrin, 1954; Sherman and Moodie, 1957; Jordan, 1970; Sherman and Morrison, 1955; Siegel, 1962; Stitt and Huntington, 1963; and DeMuth, 1969). For some clinical and research purposes it may be desirable to have utterances rated for acceptability of specific, individual phonemes. Although prior studies have shown that observers can rate global aspects of articulation with a high degree of reliability, this does not necessarily mean that they can also do this for individual phonemes. Obviously, when a speech segment consists of multiple phonemic events, the raters are provided with more information than they are when a segment consists of only one disordered phoneme. It could be argued that

the more information, the more pronounced the disorder within a speech segment; thus the more observers would be expected to agree in their ratings. There is need, therefore, to assess the abilities of groups of observers to rate reliably speech segments for specific phonological attributes of communication.

The method of quantification of perceptions of observers is important for the purpose of statistical evaluation. Psychological scaling is the method used to quantify judgments of articulatory ability. Research by Sherman and Morrison (1955) and Sherman and Moodie (1957) has indicated that the scaling technique of equal-appearing intervals (Edwards, 1957) can reliably be used to assess observer ratings of articulation defectiveness. It seems that equal-appearing intervals (EAI) could be used to assess observer ratings of acceptability of specific phonemes. Since EAI is used to assess observer ratings of overall articulation defectiveness, it would seem that EAI could also be utilized in scaling acceptability of specific phonemes.

The usual criterion for assessing validity of scaling methods is the amount of observer agreement. If judges rate stimuli in a similar manner, this indicates the use of essentially the same criteria when assigning particular scale values to stimuli. This also indicates that the stimuli had basically the same perceptual impact on the observers. Repeated differences in assignment of scale values to stimuli would indicate differences in perceptual impact on the observers and in the referential system utilized by observers.

Several factors may be involved when considering the perceptual impact that acceptability of articulation has on the listener. Factors such as type of error, consistency, and intelligibility help to determine the perceptual impact of the stimuli upon the listener. It would be safe to assume that different listeners may have different internalized standards of acceptability. Differences might be even more likely when considering particular groups of listeners. Johnson, Darley, and Spriestersbach (1963) imply that the speech clinician often makes important assumptions based upon particular groups of listeners to which a person with a speech defect belongs. They made the following statement concerning socioeconomic level and the person with a speech defect:

"As we have previously considered, the individual and his communicative attitudes and behavior are importantly influenced by the family to which he belongs, and the neighborhood and larger social setting in which he lives. You can appreciate more fully the communicative difficulties of a person and their personal and social implications for him if you take account of the general social and economic level on which he experiences these difficulties and their consequences (p. 305)."

Johnson, Darley, and Spriestersbach (1963) indicate the need for the speech clinician to be concerned about the individual in relation to his socioeconomic level. Therefore, an investigation which studies the perceptual impact of particular misarticulated phonemes upon different groups of observers from varied socioeconomic levels seems warranted.

DeMuth (1969) investigated the assumption that the speech clinician's perception of a child's speech represents similar perceptions of other listeners within a child's environment. However, he found the validity of

this assumption was questionable. More specifically, he reported statistically significant differences at the .01 level between articulation severity scale ratings for speech clinicians, mothers, and teachers. These statistical differences indicate that the assumption of similar perceptions between speech clinicians and others in the child's environment does not always hold true.

A question which the speech clinician must answer regarding each child he examines is--whether the child needs speech therapy. In more general terms, the question would be: How does the speech clinician select a case load? Decisions are usually made by the speech clinician with little, if any, help from the parents or teachers. The clinician usually makes the decision as to a child's need for therapy. After the clinician decides a child needs speech therapy, parents and teachers are informed of this decision. Sometimes, but not always, parent and teacher conferences are held prior to initiating therapy. In other words, the therapist has made the decision to include a child in therapy with little outside help, although children are sometimes referred for speech therapy by parents or teachers. The assumption that the speech clinician's perception of the child's speech is similar to perceptions of others in the child's environment may not be true in all instances, as previously indicated (DeMuth, 1969). This would seem to be the case where the writer has had parents say to him, "I didn't realize my child had a speech problem," or even more bluntly, "My child doesn't have a speech problem." This may be

an extreme example; however, as Eisenson and Ogilvie (1963, p.5) have noted, many parents do not consider their child's articulatory defect as serious because they have become accustomed to his articulation errors and do not notice them. The writer has even had some teachers say, "I didn't realize John had a speech problem." This could possibly be caused by the type of adjustment indicated by Eisenson and Ogilvie, or, since a teacher listens to many children in the course of a day, she (the teacher) could not possibly be "tuned in" to the way every child speaks. Another consideration might be that the listener perceives the child's speech as quite acceptable when intentionally listening to or "tuning in" to the child's speech. An indication here might be that the speech clinician needs to consider the perceptual impact of a child's speech upon other observers when considering case load selection. This may also mean that the speech clinician may need to revise his own concept of "right" and "wrong" phoneme production when considering other listeners in the child's environment.

It has previously been indicated that consideration of socioeconomic level can often be an important factor to the speech clinician when considering speech defects and since it has been assumed that the speech clinician's perception of speech is representative of other listeners in a child's environment, it would seem important to analyze how different socioeconomic groups perceive speech problems. For example, if the particular socioeconomic group to which a child belongs does not view the child as having a speech defect, or if they find his speech to be very acceptable, then the

child has no speech defect. On the other hand, if a child's socioeconomic group finds his speech to be less acceptable, then the child has a speech defect because this is the particular group with whom he associates most and who influence him most. It might be easier to say that a speech defect is present when the individual's speech creates a problem for him and/or attracts the attention of people with whom he associates most.

It may be possible to use the speech perceptions of the socioeconomic groups to which a child belongs as a criterion for case load selection. This would give the speech therapist a better basis for decisions as to inclusion in or exclusion from therapy. The socioeconomic group to which the child belongs would, in a sense, be helping to decide whether the child needs speech therapy. Another possible use of socioeconomic group speech perceptions might be a criterion for termination of therapy at a point where the individual fits a certain level of acceptability as determined by the group from which he comes. The implication of the importance of socioeconomic level has led to the following question for research: "Is there any difference in the way upper-, middle-, and lower socioeconomic groups rate acceptability of articulation of specific phonemes?"

The idea of this study grew out of the writer's own frustration over "which child needs speech therapy?" and his own curiosity as to whether or not different socioeconomic groups differed with respect to criteria for acceptability of speech. Since many of the children making up this writer's

case load come from what could be considered lower socioeconomic families, the writer was interested to see whether their criteria for speech acceptability was different from middle-, and upper socioeconomic groups. Children with /s/ and/or /r/ problems make up approximately 85 percent of the writer's articulatory case load. The /s/ and the /r/ are also two of the most commonly misarticulated phonemes, in addition to being the most frequently heard phonemes (Van Riper, 1938; Travis, 1931; and Henrikson, 1948); therefore, these phonemes were chosen for use in this experiment.

A previous study by DeMuth (1969) indicated a significant difference in the way speech pathologists and teachers, and speech pathologists and mothers rate articulation severity. His most tenable hypothesis for interpreting his results for this study was that internalized standards of the judges were different, thus affecting their use of the perceptual continuum as presented in equal-appearing intervals form. He suggested the need for using caution in utilizing judges in scaling experiments, depending on the interest of the research design. As in the DeMuth study, the present investigator is interested in the internal standards of different groups of people.

In view of the preceding discussion involving the importance of socioeconomic level to the speech therapist and the possible implications for case load selection, the problem for consideration in this study was: Do judgments of degree of acceptability of /s/ or /r/ by upper-, middle-, and lower socioeconomic groups differ significantly from one another?

The design of this study was set up to answer the following questions:

1. Can various socioeconomic levels reliably rate misarticulations of /s/ and /r/?
2. Is there a statistically significant difference among upper-, middle-, and lower socioeconomic groups for judgments of acceptability of /s/ production by grade school children?
3. Is there a statistically significant difference among upper-, middle-, and lower socioeconomic groups for judgments of acceptability of /r/ production by grade school children?
4. What is the effect of population density upon judgment of articulatory proficiency?
5. What is the relationship between the Warner Index of SES ratings and educational achievement?

CHAPTER II

REVIEW OF LITERATURE

Review of Socioeconomic Status

Influences of Environment and Socioeconomic Status Upon Language

Development: Previous studies have indicated the importance of socioeconomic groups to speech and language ability. Harms (1961) noted that ways of talking, speech patterns, and status dialects develop through social group membership. Bloomfield (1933) observed that a person talks more like those people with whom he most often communicates and less like those with whom he communicates least. Templin (1957) indicated that the speech and language developed by the individual are dependent upon two factors: (1) his own capabilities and (2) the environment from which he comes. The influence exerted by the environment involves several factors in determining the pattern of speech development; Templin mentions three: specific language spoken, amount and quality of conversation, and factors such as socioeconomic status. Templin (1957) stated that "socioeconomic status levels indicated, broadly, the background which may tend to foster or discourage the development of certain language skills."

Biesanz and Biesanz (1964) indicated that socioeconomic class differences give rise to status groups. The status groups are informal social groups, whose members consider each other as equals because of similar attitudes and modes of behavior. Individuals within the status group belong to the same organizations, share the same leisure activities, encourage inter-marriage among their children, and enjoy similar amounts of prestige in the community. Income and occupation are two factors which sociologists indicate as determining socioeconomic status. These factors make a certain "style of life" possible for each group, and in turn, each group helps to determine how its members react in various situations.

Since the child begins life with an ascribed socioeconomic status, namely, that of his family, the socioeconomic status of the family is important when considering speech and language development. The parent-child relationship seems to be of particular importance.

McCarthy (1954) indicated that children from lower socioeconomic levels did not ask as many questions as those from upper socioeconomic levels. She implied that this may have happened because children from lower socioeconomic levels receive fewer and less adequate answers. Milner's study (1951) showed that children from lower socioeconomic homes did not often engage in two-way conversations with their parents. Although lower socioeconomic status itself cannot be said to be causally related to poorer speech of lower class children, McCarthy (1954) stated that patterns of family life similar to those of children in the Milner study

are more prevalent in lower socioeconomic groups. These patterns are important in speech and language development.

The reinforcement and feedback from the child's environment is another important factor to be included when considering speech and language ability. Daniel and Giles (1966) indicated that reinforcement and feedback from the environment are important to language stimulation. They wrote that reinforcement of speech and language patterns may stimulate the child's interest in and motivation for speech and language development. Deutsch (1963) emphasized that in speech and language development ". a great deal of exposure to language is presupposed. Such exposure involves training, experimenting with, and identifying objects and having corrective feedback, listening to a variety of verbal material, and just observing adult language usages. Exposure of children to this type of experience is one of the great strengths of the middle-class home, and concomitantly represents a weakness in the lower-class home. The lower-class home is not a verbally oriented environment."

Other studies have also indicated the importance of corrective feedback (Ausubel, 1964; Bloon, et. al., 1965). These studies have assumed that in cases where language is well developed "corrective feedback" has taken place. In order for "corrective feedback" to take place, the child must make errors and the adults must recognize these errors. Research seems to indicate that "corrective feedback" is not taking place at the lower socioeconomic level.

Hurst (1970) has indicated "that socialization depends on the acquisition of language and that language acquisition depends on socialization; the former is substantiated by the fact that social conformity is dependent on the linguistic (dialectal) conformity and the latter, at least in part, by the fact that there are linguistic (dialectal) differences among the social classes."

Since there are linguistic differences among social class, there may be differences in the way different social classes view speech and language abilities. Several investigators have indicated differences in speech patterns among varying groups of people. Cazden (1965) indicated that the patterning of speech activities differs from society to society or from group to group within a society. Shriner (1969) stated that speech problems can exist within a speech community as well as between speech communities. He said that persons interested in communication and concerned with socially acceptable speech must study those aspects conveying patterns not thought to be socially acceptable by the listeners within the environment. Information obtained by gathering listener judgments might help investigators decide on the kinds of "corrections" that should be made. Shriner indicated that there is a need to determine which phonological deviations have the most negative effect upon listeners. Towman (1969) suggested that judgments of "appropriate" or "correct" usage, when considering language, should be the decision of the people involved in the situation.

Articulation and Socioeconomic Status: Irwin (1948, 1952) studied the effect of family occupational status and age upon sound frequency. The study compared two groups of infants: (1) infants from business, clerical and professional homes and (2) infants from homes of laboring families. He found that during the last year of the infancy period ($1\frac{1}{2}$ - $2\frac{1}{2}$ years of age) children from professional homes utilized sounds with greater frequency than did children from laboring homes. This difference was significant at the .05 level of confidence, $F=6.59$. After one and one-half years of age, the phoneme type curves separate in favor of the professional group. The phoneme frequency curve shows a similar but more pronounced shift. It is during this period that parental stimulation seems to be an important variable in speech development.

Becky (1942) and Irwin (1942) found that children with speech and language retardation usually belonged to the lower socioeconomic groups. Irwin (1942) indicated that children with delayed speech had parents with inferior educational backgrounds when compared to parents of children with normal speech development. The critical ratio for the educational factor as studied by Irwin (1942) was 2.7 (which indicates that 99.9 times out of 100, the outcome would not be caused by chance). A study by Weaver, Furbee, and Everhart (1960), which in general supported the earlier studies of Becky and Irwin, indicated that parental occupation was generally related to speech maturation. Weaver, et. al. (1960) also found a significant difference at the .01 level of confidence (chi square = 28.80) when considering

paternal occupation. Everhart (1956), however, found no significant differences between boys and girls, with or without articulation problems, when making comparisons on the basis of parental occupational classification.

Templin (1957) found consistent differences in the performance of upper and lower socioeconomic groups. She found significant differences in total articulation ability at the .01 level of confidence in the following age categories: 4.0 years ($t = 2.09$), 4.5 years ($t = 2.13$), and 7.0 years ($t = 2.50$). The upper socioeconomic group received higher scores at each age level and for all language measures when compared to the lower socioeconomic group. Templin suggested that further research concerning the language development of children and their social acceptability is indicated.

Friedlander (1968) studied the articulatory and intelligibility status of socially disadvantaged preschool children. The children utilized in this study came from one of three family backgrounds: those from families with a Spanish language background; those from native white families; and those from native Negro families. Friedlander stated that "it is generally recognized that the adverse reactions of the listener to the communication difficulties of a speaker may create serious emotional and social problems for the speaker. For these reasons, testing the articulation of preschool or school children should be followed by therapy designed to reduce in quantity and quality those deviations which create adverse reactions in the listener." This study, however, used five speech pathologists to judge intelligibility and articulation. It hardly seems that the reactions of

five expert judges could be considered to reliably represent the reactions of other listeners in the child's environment. The judgments of expert judges would probably not be as important to the child as would judgments of people with whom the child communicates most.

The Friedlander study (1968) indicated that the occupation of the father and aggregate family income were not significant in articulation, intelligibility, or verbal proficiency status of the children. He found few variables in the family data that correlated significantly and meaningfully with the articulation and intelligibility performance of the children.

Shriner (1970) indicated that culturally disadvantaged children learn the language of the region in which they live and seem to function in that particular speech community with little difficulty. However, when these children cross speech-community boundaries, the contrast with the culturally advantaged causes them to become socially different. Shriner said that language problems should be considered more fundamental than speech problems. Gussow (1965) has stated that : "what is undoubtedly and unfortunately true is that a good deal more effort has been expended on modifying the pronunciation and syntax of lower-class speech than has been expended on improving language functioning for these children." Shriner further indicated that there was little information on the kinds of "corrections" that should be emphasized in speech therapy when considering social acceptance. Nothing is known as to which phonological deviations, which syntactical "errors," or which lexical substitutions have the

most negative effect on the listener. Shriner indicated that if the language behavior is not improved, emphasis on correct speech would appear to be questionable.

Review of Methods for Articulatory Scaling

Overview of Methods for Quantifying Articulation Defectiveness:

Previous investigators have emphasized the importance of listener judgments concerning speech defectiveness. It is necessary, when considering acceptability of speech patterns, to have a way to quantify the degree of acceptability of individual speech samples. This means that an investigator must consider the method which will best enable him to collect data from listeners.

In the past, numerous scales have been constructed to quantify articulatory defectiveness. Early tests and scales involved methods such as taking frequency counts of sound errors, deriving percentages of speech sounds produced correctly, and weighting according to developmental order.

Roe and Milisen (1942) used frequency counts to determine mean scores for articulation errors. Curry, Kennedy, Wagner, and Wilke (1943) used phonographic recordings to obtain observer reactions to articulation in scaling degree of deficiency. The psychological scaling technique employed was the method of paired comparisons. Reid (1947) devised a rating scale for articulation defectiveness using a reverse developmental order. The earliest learned phonemes were given the highest number and the error

total was subtracted from the total possible score. Templin (1947) made comparisons of two methods of articulation testing. The scores obtained were based on percentages of phonemes produced correctly. Wood (1949) used an articulation index based on relative frequency of occurrence in American English of various phonemes and different positions in which phonemes occur. Each phoneme was weighted in relation to its frequency of occurrence. Snow and Milisen (1954) obtained articulation scores by assigning a value of 1.0 to 5.0 to each phoneme in each position. Correct articulation was given a value of 1.0; a mild distortion was a 2.0; a severe distortion was a 3.0; a substitution was a 4.0; and an omission was a 5.0. The mean of the rank order values for all phonemes in all positions comprised the articulation score. Wright (1954) used a seven-point rating scale involving four levels of distortion.

Choice of Scaling Method: More recent research has relied upon the psychological scaling method of equal-appearing intervals when obtaining judgments of articulation from listeners. Thurstone and Chave (1929) originally described the scaling method of equal-appearing intervals. They made the assumption that the listener's attitude toward anything being scaled would not affect reliability. Edwards (1957) stated that with the method of equal-appearing intervals, the listener need make only one comparative judgment for each stimulus.

Young and Downs (1968) indicated that most speech related dimensions can be rated by psychological scaling methodology. The first study of

speech disorders, using the method of psychological scaling, involved the rating of auditory characteristics of stuttering (Sherman and Lewis, 1951). The severity of auditory characteristics of stuttering was rated on a nine-point equal-appearing intervals scale, utilizing 40 graduate students from a clinical psychology course as observers. They concluded that the severity of stuttering could be measured on the basis of what the listeners heard alone by utilizing the scaling method of equal-appearing intervals. Sherman and Trotter (1956), and Cullian, Prather, and Williams (1963) also studied severity of stuttering, utilizing the rating scale of equal-appearing intervals.

Psychological scaling methods have also been used with voice quality and language disorders. Sherman and Linke (1952) utilized a seven-point equal-appearing intervals scale to determine whether vowel variations within controlled speech samples had any effect on perception of harshness. They utilized 35 students in speech pathology as judges in scaling perceived harshness. The results indicated that categories of vowels could be rated in relation to the perceived harshness with an r of .97 using the seven-point equal-appearing intervals scale. Rees (1958) found that the seven-point equal-appearing intervals scale could be utilized to study the influences of selected consonant environments, vowels, and vowel initiation on perceived harshness. Shriner (1967) examined relationships between selected language measures and psychological scale values derived from language samples obtained from children of specific age categories. The method utilized was that of equal-appearing intervals. The scale was

one of seven points, with one representing least development of language and seven representing most development. He indicated that "in devising scales of developmental sequence, complexity, or correctness and assigning numbers to points on the scale, the experimenter should be aware of the differences between ordinal and interval-scale measurements. Further experimentation with transformations or psychological scaling procedures may help to develop a weighting method with equal units that will eventually prove worthwhile in clinical evaluation." Sherman and Silverman (1968) utilized naive judges to rate language samples, using different scaling procedures: equal-appearing intervals, direct magnitude - estimations, and successive intervals. They found that the simpler computational factors involved in the method of equal-appearing intervals, made it a satisfactory scaling procedure for language samples.

Investigation of psychological scaling methods for use in scaling articulation defectiveness has also been done. Morrison (1955) obtained ratings of severity of articulation defectiveness using a nine-point equal-appearing intervals scale. She was interested in the reliability of measures obtained by this method. The observers used were 12 advanced students in speech pathology and 40 undergraduate students in a basic speech course. Three lengths of sample segments were used: five, ten and 15 second segments. Two sets of median scale values were gathered for the segments for evaluation of reliability of the scale values and preciseness of the values in placing segments along the severity scale. The results

indicated that judgments of trained and untrained observers for segments of five and ten seconds in length were highly reliable and precise. The reliability coefficient for the ten second segments was 0.98 and 0.97 when compared to five second segments. A follow-up study by Sherman and Morrison (1955) utilized the nine-point equal-appearing intervals scale to determine whether reliable scale values of articulation severity could be obtained from an individual observer. The observers listened to taped samples of articulation disorders, then made their judgments on the samples. Mean scale values of severity were computed. They found that the entire range of the continuum was used with no piling up of scale values at the extreme ends of the scale. In addition, they found that the mean scale values of severity were precise when placing one-minute samples along the continuum.

Sherman and Moodie (1957) indicated a need for a comparison of psychological scaling methods to determine whether high correlation of scale values could be obtained. They made a comparison of successive intervals, equal-appearing intervals, constant sums, and pair comparisons as scaling methods of articulation defectiveness. The investigators were interested in correlations between sets of scale values, and presence or absence of interval consistency within sets of scale values. The scale values obtained for equal-appearing intervals indicated that this method was relatively easy to compute. There was also close agreement between the methods of equal-appearing intervals and successive intervals when

considering internal consistency. This indicated that equal-appearing intervals is the preferred method when scaling short segments of speech involving articulation defectiveness.

Guilford (1954) indicated several advantages to the utilization of the method of equal-appearing intervals; some of which would be applicable to the present study:

- (1) It can be used with naive observers having a minimum of training.
- (2) It can be used when presenting a large number of stimuli.
- (3) Many experimenters feel that the best judgments are made when stimuli are present individually.

The method of equal-appearing intervals allows the observer to make judgments on the basis of the observer's own internal standards concerning the type of stimuli being rated. For the purpose of the present study, the anchor points for the rating scale will be "most acceptable" and "least acceptable." The measurement which is obtained is the observer's internal standards in relation to his own idea or attitude of least to most acceptable. It is important, however, that prior to scaling, the end points are tied down by allowing observers to listen to the entire range of samples before making their judgments.

Selection of Observers: Another related problem with which previous investigators have been concerned is the question: "Can naive listeners be used as observers?" Previous studies have been concerned with the use of untrained observers in evaluating speech defectiveness.

A previously mentioned study by Morrison (1955) utilized both trained and untrained observers. The results showed that differences between median scale values of the two observer populations were not significant. Perrin (1954) utilized the method of pair comparisons to investigate whether trained and untrained judges could rate functional articulation defects. He found no significant difference in the way trained and untrained judges made evaluation of articulation defects. Siegel (1962) was concerned with comparability of articulation examiners. He used two experienced examiners (graduate students in speech pathology) and two inexperienced examiners (women who had been classroom teachers). Judgments were made of correct, incorrect, or unscorable responses of children to a modified Templin-Darley articulation test. The inexperienced judges received training after their first listening session, while experienced judges received no training. Siegel found that inexperienced judges correlated highly ($r = 0.92$) before receiving training; and that they also correlated well with experienced judges. DeMuth (1969) utilized two groups of untrained listeners to rate severity of articulation disorders, using the equal-appearing intervals scale. The obtained r for each group exceeded 0.97.

Previous studies (Morrison, 1955; Sherman and Morrison, 1955; Sherman and Moodie, 1957) have indicated that the psychological scaling method of equal-appearing intervals lends itself most effectively to the scaling of articulation defectiveness; other studies (Siegel, 1962; Perrin, 1954; Sherman, 1955) have also indicated that untrained judges can

reliably scale articulation defectiveness. On the basis of these studies, this writer decided that both the equal-appearing intervals method and untrained judges could be utilized in the present study.

Quantifying Socioeconomic Status

Objective ratings seem to be the preferred method of quantifying socioeconomic status. Bergel (1962, p. 262) indicated that "these ratings are based either on a single criterion (wealth, income, occupation, position) or a combination of several criteria (multiple correlation). Most of these criteria are economic categories. The validity of the ratings rests on the tacit or explicit assumption that social classes are either identical with, or closely related to, systems of economic stratification. It is only consistent with this assumption that many authors using economic factors for their ratings speak of "socioeconomic status."

Those studies using a single criterion usually get their material from already existing compilations, such as income or labor statistics. When using multiple correlations, the investigator must decide on that combination of criteria which will serve his purpose.

The Minnesota Scale for Paternal Occupation is an example of a single criterion scale, classifying persons into one of the eight classes, according to parental occupations. It has been used as a method for scaling socioeconomic status.

Bergel (1962) indicates, however, that computations of socioeconomic status based on a single factor are not reliable, because status is a complex phenomenon. A multiple correlations approach is preferable.

There are several factors which can be utilized when considering a multiple correlations scales. Factors, such as the following may be considered: education, occupation, income, source of income, area lived in, and type of dwelling. Bergel (1962) indicates that income seems to be one of the least revealing factors, and it's often over-emphasized when considering socioeconomic status. He further indicates that occupation seems to be the most important single factor in determining social class.

Several multiple correlations scales are available for use in studies considering socioeconomic status. Two of the more frequently utilized are: 1) the Chapin scale (1935), 2) Index of Status Characteristics (Warner, et. al., 1949).

Chapin's scale is also called "Chapin's Living Room Scale." He was convinced that he could "judge the social status of a person" if he knew the person's "cultural possessions, effective income, material possessions, and social participation" (Bergel, 1962, p. 263). Chapin felt that these four factors could be determined by looking at the individual's living room. Bergel (1962, p. 263) objects to measuring cultural possessions or cultural expressions by observing the furnishings in a living room. He says that "Non material culture, we firmly believe, cannot be conceived in quantitative terms, nor is the mere possession of cultural 'matter' a reliable criterion for a person's culture."

The Index of Status Characteristics by Warner, Meeker, and Eells (1949) is a multiple correlations scale considering four factors:

occupation of bread winner, source of income, house type and size, and area lived in or education of bread winner. Turmin (1967) and Bergel (1962) have indicated that the problem with this scale is that there is not quantitative data offered for support, and that it is difficult to say whether the status orders of various communities of various sizes are comparable to those delineated by Warner for his population.

The evaluation of any socioeconomic scale is often based on two factors: 1) correct statistical procedures and 2) validity of the selected symbols. Bergel (1962) indicates that Chapin, Warner, and practically all other authors, using different scales, have been attacked on both grounds. Actually, the chief objection directed against the principles underlying all types of socioeconomic scales is that status is something that is accorded by the community. Any investigation into the status of an individual should consider what the community thinks of the person. The socioeconomic scales are based upon what an individual thinks of himself, not what the community thinks, and herein is the basic problem. Another objection is based on the objective nature of these tests (education, income, house type). Bergel (1962) indicated that the hereditary character of status is neglected and should be included. However, setting up a scale which would take into consideration an individual's hereditary status, and how the community views a particular individual, would be a very lengthy and cumbersome task which in itself is not practical.

Despite the objections to the use of any of the present multiple correlations scales, Bergel (1962) has concluded that the Index of Status characteristics is a fairly reliable instrument of class distribution in a "community of small or modest proportions has to be determined."

However, this method is not adequate when considering large communities or the nation as a whole.

Since the Warner Index has been indicated by Bergel (1962) and Gordon (1958) to be as good, if not better than, other multiple correlations scales, it will be utilized as the method of classifying socioeconomic status in the present study.

CHAPTER III

SUBJECTS, PROCEDURE, EQUIPMENT

Subjects: Fifty-seven public elementary school children (grades 1 through 3; thirty seven males and twenty females), whose chronological ages ranged from 5 years, 11 months to 9 years, 5 months; served as subjects. Each subject had been diagnosed by the experimenter, a speech correctionist teaching for the Kaskaskia Special Education District in four small communities (Sandoval, Patoka, Odin, and Selmaville, Illinois) as having an articulation disorder involving either the /s/ or /r/. There were fourteen children with lateral lisps, twenty-one children with frontal lisps and twenty-two children with /r/ problems, such as w/r substitution as in "red," omissions of the final /ɹ/ and medial /ɹ/, and vowel substitution, such as /bʊd/ for /bɹd/, and /beʊ/ for /beɹ/.

All subjects had misarticulations in varying degrees, in that some subjects misarticulated a particular phoneme more consistently, while others were not consistent in their misarticulations. Twenty-two of the subjects were receiving speech therapy at the time this study took place. All subjects came from the waiting list or case load of the experimenter, with the exception of three children with lateral lisps, who came from the

case loads of two speech correctionists in the Centralia City Schools, Centralia, Illinois. The three subjects from the Centralia City Schools were all receiving speech therapy at the time of inclusion in this study.

Recording of Stimuli: A speech sample of approximately two minutes was taped from each subject. Connected speech samples were studied by Faircloth and Faircloth (1970), who indicated "analysis of connected speech describes a person's habitual articulatory behavior more appropriately than does single word testing." Samples of connected speech are also more representative of the conditions under which listeners would ordinarily hear children speaking.

Continuous speech was elicited from the children by selecting approximately five of fifteen verbal directives prepared by Miner (1970) for use in eliciting LCI samples. The specific verbal directives used were: (1) "Tell me about your mother," (2) "Tell me about your family," (3) "Tell me about Christmas," (4) "Tell me about your favorite story," and (5) "Tell me about your favorite T.V. program." An approximate distance of two to three feet was maintained between the subject and the microphone. Each child faced the experimenter and microphone, and spoke directly into the microphone. Responses were recorded on an Ampex Recorder, model number 60, on 1.5 Mil Acetate Magnetic Recording Tape, with tenzar backing at a tape speed of seven and one-half inches per second. Each subject spoke for approximately two minutes.

Preparation of Stimuli: A ten second segment of continuous speech was extracted from the original two minute recording. Each representative segment which was selected for use in this study was then cut from the original two-minute sample. The fifty-seven 10 second segments were spliced together in random order with seven second inter-stimulus intervals between each sample. Randomization was achieved by placing the fifty-seven 10 second samples in a box, mixing them up and then drawing them out one at a time, placing them in sequential order. Next a table of random numbers was used to select the sample order for splicing. A seven second inter-stimulus interval as used in the DeMuth study (1969) was utilized in this experiment to allow for judging time.

The ten second segment length was selected on the basis of the Morrison (1955) study. Morrison found reliable scale values of articulation severity for five and ten second segments. She had used both naive and expert judges in this study.

The ten second segments which were extracted by the experimenter were representative of the speech defect (either /s/ or /r/) of the subjects. Each segment contained a minimum of three misarticulations (either /s/ or /r/) per segment.

Description of Scaling Method: A seven-point equal-appearing intervals scale of acceptability was used, with one representing least acceptable and seven representing most acceptable production of either /s/ or /r/, as judged by the listener. The experimenter felt that seven points would be sufficient for the rating of the specific phonemes. The method of equal-appearing intervals was used, since it is indicated as being the most useful scale for articulation studies (Sherman and Moodie, 1957).

Selection of Judging Panel: The judges for this experiment were representative of three different socioeconomic levels for the Centralia, Sandoval, Odin, Patoka, and Selmaville areas. The three socioeconomic status (SES) levels used were lower-, upper-, and middle classes. The number of judges representing each class was determined by the use of the principle of sequential sampling (Silverman, 1967), with a predetermined alpha level of 0.05. On the basis of this criterion, the lower class contained thirty-nine judges; middle class, thirty judges, and upper class, nineteen judges.

Socioeconomic level was determined on the basis of the Index of Status Characteristics (Warner, Meeker, and Eells, 1949). The factors employed with this index were: Factor I - occupation of the bread winner of the family; Factor II - source of income; Factor III - house size and type; and Factor IV - for the purposes of this study Alternative (b) was used - education of bread winner.

The information on each of the four factors was rated on a seven-point scale by the examiner and multiplied by a prescribed weight (Johnson, Darley, and Spriestersbach, 1963, p. 306). The total of the values placed the judges in one of three socioeconomic levels. This index is set up to represent five different socioeconomic levels: Upper, Upper-middle, Lower-middle, Upper-lower, and Lower-lower. The present study utilized only three levels; therefore, upper-middle, and Lower-middle were combined as Middle, and Upper-lower and Lower-lower were combined as Lower.

The cutoff values for each group were as follows: upper, 12-22; middle, 23-51; and lower, 52-84.

Judges were also divided into two geographic groups on the basis of population density. These two groups were urban and rural dwellers. The procedure for separating judges into urban and rural groups was the same as that used by Miner (1970). A judge living in the city limits or in a housing development of more than four families adjacent to the city limits, was classified as an urban dweller. Any judge not fitting the description was classified as an rural dweller. The rationale for this procedure stems from the U. S. Census Bureau, Department of Commerce.

Parents of children serving as subjects were not used as judges in this study.

Hearing acuity of judges was allowed to operate as a random variable since the experimenter was interested in a sampling of the general population.

Presentation of Stimuli: The judges first filled out an information sheet for purposes of socioeconomic classification by the examiner (see Appendix I). Next, the examiner read the instructions to the observers as they followed along on their own instructions sheet (see Appendix I).

A one page answer sheet was prepared with numbers from one to fifty-seven, corresponding to each sample on the tape. Each judge assigned a scale value next to the corresponding number of the stimulus being rated. A sample of the observers answer sheet is also included in Appendix I.

Five different locations were used for presentation of the tape recorded responses of the judges. This was done since the judges came from different communities within a 20 mile radius in South-Central Illinois. The sound level of ambient noise in each location was measured prior to presentation of the taped samples with a sound level meter as a guard against possible interference. The locations, sound level of each location, and number of sessions at each location is indicated in Table 1. For the present investigation, the average signal-to-noise ratio was 65/50, an acceptable level for listening to tape recordings.

TABLE 1.--Location of Judging Sessions, Sound Pressure Level (SPL) of the Room, and the Number of Sessions at Each Location

Location	SPL (C Scale)	Number of Sessions
First Christian Church Library (Centralia, Illinois)	40 dB	8
Odin Public School Library (Odin, Illinois)	50 dB	1
American Legion Hall (Patoka, Illinois)	55 dB	1
Sandoval Grade School Classroom (Sandoval, Illinois)	45 dB	3
Central City Lion's Club Shelter House (Central City, Illinois)	50 dB	1
Patoka High School Classroom (Patoka, Illinois)	50 dB	1

The size of the judging groups varied from individual to groups of fifteen, depending upon availability of observers. Stimuli were presented in sound field with the control of the recorder set at a 65 dB level for the person farthest from the speaker, as measured by the sound level meter. Each session lasted approximately an hour, with only one judging session required for each judge.

Analyses of Judges' Ratings: A total of 5, 016 judgements were available for analysis. The scale values assigned by the observers were transferred to IBM cards for statistical computation. The reliability of scale values of the socioeconomic groups was determined by an intra-class correlation coefficient for averages (Winer, 1962). A one-way analysis of variance was computed on an IBM 360 computer for comparison of differences between lower-, middle-, and upper socioeconomic groups for rating acceptability of /s/ and /r/ productions. A t test was used to assess the effect of population density upon judgments of articulatory proficiency. A correlation measure, the triserial r , was used to determine whether there is any relationship between the Index of Status Characteristics and educational achievement (Heath and Downie, 1965, p. 194).

CHAPTER IV

RESULTS AND DISCUSSION

Results

The purpose of this chapter is to report the statistical computations and interpret the results of the present study. Five questions were posed at the outset of this investigation.

1. Can observers representing various socioeconomic levels reliably rate misarticulations of /s/ and /r/?

An intraclass correlation coefficient for averages (Winer, 1962, p. 128) was computed for the mean scale value ratings of each of three judging populations. The number of judges representing each class was determined by use of the principle of sequential sampling (Silverman, 1968) with a pre-determined alpha level of 0.05. The obtained \bar{r}_{ave} for each population is indicated in Table 2.

TABLE 2.--Reliability Levels and Number of Judges for Upper-, Lower-, and Middle Socioeconomic Status Populations.

Group	N	Reliability Level (\bar{r}_{ave})
Lower	39	0.97
Middle	30	0.95
Upper	19	0.95

The obtained r_{ave} for each population met the predetermined confidence level of 0.95 as indicated in Table 2. The magnitude of these values indicate that if the experiment were repeated with another random sample of upper-, middle-, and lower socioeconomic judging populations using the same number of judges and set of stimuli, the correlation between mean ratings obtained from the populations would again meet the predetermined confidence level of 0.95. That is to say, within each judging population, the stimuli rank ordered themselves in a similar manner.

The answer to this question is that different socioeconomic groups can reliably rate acceptability of specific phonemes. There is now evidence indicating observers can reliably scale both overall articulation ability and specific phonemes. The data for this study also agree with the findings of Siegel (1962), Morrison (1955) and Perrin (1954), indicating that inexperienced observers can be used to scale articulation with high reliability.

2. Is there a statistically significant difference among upper-, middle-, and lower socioeconomic groups for judgments of acceptability of /s/ production by grade school children?

A one way analysis of variance was used to determine whether there was a statistically significant difference in the way socioeconomic groups scaled thirty-five stimuli for /s/ production. The resulting F -ratio was 0.05 ($df=2/102$). This value was interpreted to mean that there were no statistically significant differences in the way different socioeconomic groups scaled the /s/ stimuli. In other words, the different socioeconomic groups perceived the /s/ stimuli in a similar manner. This is noteworthy,

in that, there is now data to show how at least one particular speech community views particular phoneme defectiveness, specifically /s/. More specifically, socioeconomic status was not found to be a relevant variable in considering misarticulations.

Since no differences were found among socioeconomic status groups, the data were pooled for further analyses. The /s/ errors consisted of two basic types: frontal lisps and lateral lisps. The question arises, were there any differences within /s/ errors, that is, were lateral lisps and frontal lisps scaled in a significantly different way?

In order to answer this question, measures of skewness and kurtosis were utilized. Skewness and kurtosis are two ways to describe the curve of a distribution. Griffin (1962, p. 113) described skewness as the "asymmetry" of a curve and kurtosis as the "extent to which a unimodal frequency curve is peaked." Downe and Heath (1965, p. 25) described skewness in terms of the direction in which the tail of a curve is extended. A distribution is said to be positively skewed when the tail extends to the right side of the graph and negatively skewed when the tail extends to the left. Skewness and kurtosis can be used to compare a distribution to the hypothetical normal distribution.

In the case of the present study, the distribution to be compared to the normal distribution was the grand mean scale values of acceptability for frontal lisps and the grand mean scale values of acceptability for lateral lisps. A grand mean scale for each /s/ stimuli was computed by combining

the mean scale values of the three socioeconomic groups for each /s/ stimuli. This information was then key punched on IBM cards for analysis on the IBM 360 computer. The resulting values were measures of skewness and kurtosis for the distributions of lateral lisp scores and frontal lisp scores.

Griffin (1962, p. 114) indicated a value of $\pm .50$, or within that range is necessary for both measures of skewness and kurtosis, if an approximation of a normal distribution is to be expected. The distribution for frontal lisps was symmetrical but platykurtic; the distribution for the lateral lisps was positively skewed and platykurtic. The values for skewness and kurtosis are indicated in Table 3.

TABLE 3.--Values of Skewness, Kurtosis, Mean, and Standard Deviation for Lateral and Frontal Lisps.

Distribution	Skewness*	Kurtosis*	$\bar{M}SV$	SD
Frontal Lisps	-0.36	0.63	4.70	1.16
Lateral Lisps	-1.08	-0.79	4.39	0.99

* $\pm .50$ - values must fall at this value or less to be considered normal distribution.

The requirements of a normal distribution in terms of skewness and kurtosis are that: 1) It is symmetrical and 2) its curve is mesokurtic (normal). When considering the values in Table 3, neither the distribution for frontal lisps or that for the lateral lisps can be considered to be a normal distribution. Although the distribution on frontal lipss was

symmetrical as indicated by the value for skewness (-.36), the distribution's curve was less peaked (platykurtic) than the normal curve. The value for the kurtosis (peakedness) of the distribution was 0.63, which is not within the $\pm .50$ limit. The distribution of frontal lisps, therefore, cannot be considered normal. The values indicated for the distribution of lateral lisps did not meet the $\pm .50$ limit for either skewness or kurtosis. This distribution was asymmetrical with a flattened peak (platykurtic).

A comparison of the mean scale values and standard deviations of frontal and lateral lisps along with the values for skewness and kurtosis indicated that the listeners scaled frontal lisps as being more acceptable than lateral lisps.

3. Is there a statistically significant difference among upper-, middle-, and lower socioeconomic groups for judgments of acceptability of /r/ production by grade school children?

A one way analysis of variance was also used to determine whether there was a significant difference in the way different socioeconomic groups scaled twenty-two stimuli for /r/ production. The resulting F -ratio was 0.13 (df=2/63) as computed on the IBM 360 computer. An F value of 19.47 (df=2/63) is needed at the 0.05 level of confidence for a significant difference to be indicated. This value was also interpreted to mean that there was no statistically significant difference in the way the different socioeconomic status groups scaled the /r/ stimuli. In other words, the different socioeconomic status groups perceived the /r/ stimuli in a similar manner. This again is noteworthy as indicated in question number two,

that is, there is now data to show how at least one speech community views particular phoneme defectiveness when considering /r/ production. Unlike the /s/ stimuli, the /r/ stimuli in this study did not represent more than one distinct group of error type; therefore, further analysis of differences within /r/ stimuli was not indicated.

Measures of skewness and kurtosis were utilized for /r/ as had been done with /s/. These measures were used to determine how scale values for /r/ distributed themselves. A value of 0.92 was found for skewness indicating that the curve was positively skewed. A value of 0.40 for kurtosis indicated that the distribution was mesokurtic (normal). The mean scale value for the /r/ distribution was 3.53 with a standard deviation of 1.05. A comparison of the mean scale value for /r/ to the mean scale values of frontal and lateral lisps indicated that the /r/ stimuli was less acceptable to the listeners than frontal and lateral lisps.

4. What is the effect of population density upon judgment of articulation proficiency?

A t-test was utilized to determine whether there was a significant difference between the way urban and rural dwellers scaled /s/ and /r/ production. The mean scale value for rural dwellers was 4.27 and for urban dwellers it was 4.45. A significant difference at the .05 level of confidence would be 2.00 (df=56). The resulting t-value as computed on the IBM 360 computer was -2.12 (df=56), statistically significant beyond the .05 level of confidence.

The t test was utilized instead of the Mann-Whitney U because with a population as large as the urban population (N=70) you would get comparable results using either test; therefore, it stands to reason that an experimenter should use the more powerful test (the t test). The resulting t value also indicated that using the t test is supportable since it is significant beyond the .05 level.

In an effort to further analyze the overall effects of the significant difference in the overall mean scale values of the urban and rural populations, the stimuli were divided into three phoneme error groups: frontal lisps, lateral lisps, and /r/. Computer analyses were done to determine the mean scale values of each phoneme error group for the urban and rural populations. A t test was then applied to the mean scale values of the urban and rural populations to determine which type of phoneme error was causing the significant difference in overall mean scale values for the two judging populations. The obtained mean scale values for the urban and rural populations are indicated in Table 4.

TABLE 4.--Mean Scale and t Values for Urban and Rural Populations of Overall and Specific Phoneme Errors

Stimuli	Urban	Rural	<u>t</u>
Overall	4.45	4.27	-2.12*
Frontal	4.62	4.99	5.28**
Lateral	4.39	4.44	0.50
/r/	3.55	3.50	-0.41

* t value significant beyond .05 level of confidence ($t_{.05}=2.00$)

** t value significant at the .001 level of confidence ($t_{.001}=3.850$)

Table 4 indicates that the t value for the mean scale value of the frontal lisps is significant at the .001 level of confidence. A significant difference at the .001 confidence level would be 3.85 (df=20). The resulting t value for the frontal group was 5.28 (df=20). This indicates that the significant difference in the mean scale values for the frontal lisps accounts for the significant difference in the overall mean scale values for the rural and urban populations.

It is interesting to note that the rank order of acceptability for the three phoneme error groups when considering urban versus rural populations is the same as that for the total population (urban and rural combined). In other words, frontal lisps were scaled most acceptable by both groups of observers followed by lateral lisps and then /r/ in that order.

When considering the three phoneme error distributions as a whole, the mean scale values indicated that the urban population is more accepting of the errors ($\overline{MSV} = 4.45$) than the rural population ($\overline{MSV} = 4.27$), but in terms of specific phoneme error distributions a significant difference was found only on the frontal lisp distribution with the rural population more accepting of this error than the urban population.

Two questions need to be answered at this point:

1. Why did the listeners rank order the acceptability of phoneme errors in the order in which they did?
2. Why was a significant difference found between urban and rural populations on the frontal lisp with rurals more accepting than urbans?

The question concerned with rank order of phoneme error acceptability can be answered in terms of distinctive feature theory. Chomsky and Halle (1968, p. 177) have compiled a table of distinctive features of English phonemes. This table indicates distinctive features present or absent in English. It is the opinion of this researcher that the frontal lisp (θ/s) was more acceptable to observers because the θ/s differs from the $/s/$ by only one distinctive feature, namely stridency. The lateral $/s/$ distortions, on the other hand, is not a phoneme which normally occurs in English. While it still has some distinctive features in common with the $/s/$, it might still be considered an allophonic variant of $/s/$. Distinctive feature theory would hypothesize that the difference between a lateral $/s/$ and an acceptable $/s/$ production is greater than the degree of difference in a θ/s substitution. The $/r/$ was probably least acceptable to the observers because of two specific distinctive features which the $/r/$ and $/s/$ do not have in common. The $/r/$ is vocalic and voiced. The $/s/$ lacks these two features. It would seem that errors with voiced phonemes would be more noticeable than with voiceless phonemes. The $/r/$ is voiced while the θ/s and distorted $/s/$ are not voiced. Black (1952) has also indicated that when considering English as a foreign language, vowels contribute more to intelligibility than consonants. The $/r/$ has a vocalic distinctive feature and the θ/s and distorted $/s/$ do not, which means in light of what Black has indicated, the $/r/$ should call more attention to itself than the other two phoneme errors. It is for this reason that the $/r/$ was scaled by the observers as less acceptable than the frontal and lateral lisps.

It is interesting to note that the manner in which the phoneme errors were rank ordered in terms of acceptability is contradictory to the developmental order of misarticulation of phonemes as indicated by Roe and Milisen (1942). Milisen (1954) indicated the importance of error type as an indicator of severity. He concluded that articulation errors can be ranked according to severity. Severity can be rank ordered from most severe to least severe according to Milisen, in the following order: 1) omissions, 2) substitutions, and 3) distortions. It may be that distinctive feature theory needs to be considered in ranking severity of phoneme error production.

The reader will recall another question raised from the findings of this investigation, namely, why was a significant difference found between the urban and rural populations on frontal lisps with rurals more accepting than urban dwellers? The difference can be explained statistically on the basis of the significant difference in central tendency exhibited by the two populations for the frontal lisp, however, any further interpretations of this finding await further research.

5. What is the relationship between the Warner Index of SES ratings and educational achievement?

A correlation measure, the triserial r (Downe and Heath, 1965, p. 194), was utilized to determine whether there was any relationship between the Index of Status Characteristics and educational achievement. The resulting triserial r was -0.03 , which indicates no appreciable relationship

between the two measures. This indicates that for the purposes of this study, educational level alone would not have been a sufficient indicator of socioeconomic status. These results support Bergel's (1962) contention that a single factor computation is not reliable when assessing socioeconomic status.

Discussion

The major finding of this experiment was that socioeconomic status is an irrelevant variable when scaling acceptability of /s/ or /r/ production. In other words, the upper-, middle-, and lower socioeconomic status groups assigned similar ratings when scaling /s/ or /r/ production. One explanation for this finding might be that the stimuli did not represent a range of acceptability for the purpose of scaling. This would not be the case, however, since the judges utilized the entire range in scaling stimuli. Table 5 reports the range of mean scale values for the three socioeconomic populations used in this study.

TABLE 5.--Range of Mean Scale Value for Upper-, Middle-, and Lower Socioeconomic Groups

SES	Low Mean Scale Value	High Mean Scale Value
Upper	1.79	6.58
Middle	2.23	6.07
Lower	1.87	6.54

A second hypothesis of the results was that three different socioeconomic groups were not represented, therefore, no differences in scale values could be expected. For example, a person who would supposedly be classified in the middle socioeconomic groups may not have been significantly different when comparing his "life style" (i.e., social grouping and living conditions) to a person classified in the upper socioeconomic group. The classifications were made on the basis of the directions for classification by the Index of Status Characteristics (Warner, 1949). All classifications were made in a consistent manner according to the directions; therefore, the criteria for classifying individuals in socioeconomic groups was met for this index. Even if three different socioeconomic groups (upper, middle, and lower) were not represented, there were at least two groups: 1) Upper-Middle, and 2) Lower-Middle. Any weakness in the socioeconomic groupings would be in the sensitivity of the index utilized to differentiate socioeconomic groups. It has already been indicated that this index is as good, if not better than, other multi-correlation indexes used to differentiate socioeconomic groupings (Bergel, 1962; Termin, 1962).

A third hypothesis and probably the most tenable explanation, is that there is actually no difference in the way different socioeconomic groups scale acceptability of /s/ and /r/ production, and that three different judging populations, based on socioeconomic status were represented in this study. This would be the case at least when considering socioeconomic groups in the South-Central Illinois area. When considering the populations

of judges in this study, acceptability of speech in relation to /s/ and /r/ production would have basically the same impact on all listeners regardless of socioeconomic group. This is consistent with two other studies in the same region (Shriner, Miner, 1968; James, 1967) which found socioeconomic status to be an irrelevant variable in considering transformational and morphological skills of children.

The principle implication of this study is that the experimenter can utilize the scale values of his judges to help in making decisions as to case load selection when considering /s/ and /r/ production. A logical step in this direction might be the development of a master tape for the /s/ and /r/ stimuli, using the original tape from this study. Although the distributions for /s/ and /r/ stimuli were not normally distributed, samples can be found on the original tape whose mean scale values are close to the seven interval points on the scale. The speech clinician could then listen to the samples representing different scale values and learn to compare the speech of other children with /s/ or /r/ problems to these samples. A scale value could be assigned to other children with /s/ or /r/ errors. When considering a child with a frontal lisp, the speech clinician will have to consider whether the child is an urban or rural dweller, because this study has indicated that there is a significant difference in the way the two populations scale the frontal lisp. The rural dweller was more accepting of the frontal lisp; therefore, the speech clinician will have to take this into consideration. It should be kept in mind that this procedure is only one of

many tools which can be utilized in case load selection. This scale should be used in conjunction with other methods used to help make the decision as to case load selection.

Furthermore, this study can be utilized as a guideline for setting up other master tapes for /s/ and /r/ values in other areas. It could again be a useful tool for other clinicians in decision making, which concerns therapeutic involvement. In other words, this study could be carried out in other regions in order to determine a set of scale values, which is representative of the "lay" listeners in that region, thus the speech clinician would have an additional tool to help him in case load selection when considering /s/ and /r/. Depending upon the region and the specific phoneme disorders found primarily in that region, the speech clinician may want to find scale values of phonemes other than /s/ and /r/. There is no reason why this study cannot be done for other phonemes.

CHAPTER V

SUMMARY AND CONCLUSIONS

Five questions were considered in this study:

1. Can observers representing various socioeconomic levels reliably rate misarticulations of /s/ and /r/?
2. Is there a statistically significant difference among upper-, middle-, and lower socioeconomic groups for judgments of acceptability of /s/ production by grade school children?
3. Is there a statistically significant difference among upper-, middle-, and lower socioeconomic groups for judgments of acceptability of /r/ production by grade school children?
4. What is the effect of population density upon judgment of articulation proficiency?
5. What is the relationship between the Warner Index of Status Characteristics and educational achievement?

A review of the literature indicated the importance of environmental influences including socioeconomic status upon the development of speech and language skills. None of the studies reviewed were found to be directly related to the questions posed in this investigation. The question of importance was: How do various socioeconomic groups scale acceptability of /s/ or /r/ production?

Justification for this study was based on the concept that scale values from different groups of listeners could be helpful in setting up guide lines

for case load selection; that is to say, the scale values of the groups would serve as a criterion for case load selection. It was thought that various socioeconomic groups might have different standards of what constitutes acceptable phoneme production. This investigation was concerned with the reactions of judges to specific phoneme misarticulation, rather than a global overview of articulation.

The method of investigation was to obtain samples of conversational speech of children with misarticulations of /s/ or /r/. The samples were then presented to three judging panels representative of upper-, middle-, and lower socioeconomic groups as determined by the Index of Status Characteristics (Warner, 1949). The judges were then asked to rate each sample for acceptability of either the /s/ or /r/.

Fifty-seven subjects diagnosed by the experimenter as having either an /s/ or /r/ articulation disorder were recorded on tape. Conversational speech was elicited by using the verbal directives established for use in evoking language samples for the Length-Complexity Index (LCI) (Miner, 1970). Ten second segments were then selected and spliced together in random order for scaling by the three socioeconomic groups.

The fifty-seven stimuli obtained were then rated by each judging panel on a seven-point equal-appearing intervals scale. The resulting values were then placed on IBM cards and the data were analyzed by computer.

In answer to question one, can observers representing various socioeconomic levels reliably rate misarticulations of /s/ and /r/, all three

judging populations met the predetermined alpha level of 0.05. This was interpreted to mean that each of the three socioeconomic groups could reliably scale acceptability of articulation.

In answer to questions two and three, is there a statistical significant difference among upper-, middle-, and lower socioeconomic groups for judgments of acceptability of /s/ or /r/ production by grade school children, significant differences were not found between any two of the socioeconomic groups. The null hypothesis for this study was supported. Several hypotheses were formulated to account for this:

1. The stimuli did not represent a range of acceptability with sufficient width for the purpose of scaling.
2. The listeners did not represent three different socioeconomic groups.
3. Socioeconomic status is truly an irrelevant variable in considering acceptability of /s/ or /r/ production; and three distinct socioeconomic status groups were represented as measured by the Index of Status Characteristics.

The third hypothesis appeared to be the most tenable explanation of the results of this study. It indicated that there were three different socioeconomic groups represented as indicated by the index utilized; therefore, the results indicated no significant differences in the way various socioeconomic groups perceived acceptability of /s/ and /r/ production. This would be the case at least in the South-Central Illinois area. The results indicate that speech from the standpoint of acceptability for /s/ and /r/ production is given similar scale values by the three socioeconomic

status groups. In other words, socioeconomic status is an irrelevant variable when considering acceptability of /s/ or /r/ production, at least in this geographic region.

The fourth question for consideration was what is the effect of population density upon judgment of articulation proficiency. There was a significant difference between the mean scale values of urban and rural populations. The difference between the mean scale values was a result of the way the urban and rural populations scaled the frontal lisp. Although overall the urban population was more accepting than the rural population, in terms of specific phoneme errors a significant difference was found only on the frontal lisp with rurals more accepting than urbans. The investigator is unable to explain the reason why this occurred.

The last question to be considered in this study was: what is the relationship between the Warner Index of Status Characteristics and educational achievement. The correlation measure utilized indicated no relationship between the index and the educational level of the three groups of listeners. This seems to support the idea that a single factor is not reliable when determining socioeconomic status because socioeconomic status is a complex phenomenon (Bergel, 1962).

Implications for Further Research

There are several implications for further studies which have been brought about as a result of the present study:

1. A study comparing the scale values of the present study to scale values of speech pathologists.
2. Preparation of master training tapes for /s/ and /r/ production.
3. Studies in areas other than South-Central Illinois for /s/ and /r/ scale values.
4. Studies involving scale values for phonemes other than /s/and /r/.
5. Further study concerning acceptability scaling of omissions, substitutions, and distortions.

APPENDIX I

- I. Name: _____
- II. Sex: _____
- III. Address: _____
- IV. How long have you lived at this address: _____
- V. Occupation: Yourself _____
Spouse _____

VI. Source of income (Check the appropriate number below):

- ___ 1. Savings and investments, inherited; 50 percent or more of income
- ___ 2. Savings and investments, gained by earner (not retirement pensions)
- ___ 3. Profits and fees—including higher executives who share profits
- ___ 4. Salary or commission, including retirement earned thereby
- ___ 5. Wages based upon hourly rates or piece-work; time card personnel
- ___ 6. Private aid or assistance; may be supplemented by part-time work
- ___ 7. Public relief

VII. House Size (Check appropriate number) VIII. Condition of House

- | | |
|--|-------------------------|
| ___ 1. Small house (5 rooms, excluding bathrooms) | ___ 1. Good condition |
| ___ 2. Medium house (7 rooms, excluding bathrooms) | ___ 2. Medium condition |
| ___ 3. Large house (10 rooms, excluding bathrooms) | ___ 3. Fair condition |
| | ___ 4. Poor condition |

If your home does not fit any of the above descriptions, explain:

IX. Education: (number of years of schooling completed):

Yourself _____ Spouse _____

Instructions to Observers

As children learn to talk, some learn to talk better than others. Two sounds which seem to give children a lot of trouble are the "s" and "r", the "s" sound as in "say", "last", and "horse" and the "r" sound as in "red", "bird", and "bear." You will be listening to some short tape recorded samples of children who are pronouncing either the "s" or "r" with varying degrees of acceptability. You are asked to judge each sample in relation to a seven-point scale of degree of acceptability for either "s" or "r" pronunciation. In other words, you are asked to answer this question:

How well does each child pronounce his "s"s or "r"s?

Following these will be 57 speech samples to be rated on a seven-point scale. These speech samples were obtained by requesting children to talk about some things of interest to them. Each sample is approximately 10 seconds long.

Make your judgment on the basis of the total speech sample. Avoid being influenced by grammatical correctness. Do not give a rating based upon a judgment of vocabulary usage. Don't pay any attention to the child's voice, but only to the way he pronounces either his "s" or "r" words.

The scale is one of equal intervals- from 1 to 7, with 1 representing least degree of acceptability and 7 representing most acceptable; 4 represents the midpoint between 1 and 7 with respect to acceptability; the others fall at equal distances along the scale. Do not attempt to place samples between any two points of the seven points, but only at these points: 1, 2, 3, 4, 5, 6, or 7.

Each speech sample is preceded by a number. Your task will be to record your judgment on your answer sheet to the left of the identifying number of the speech sample.

Before you record any judgments, you will first be given an opportunity to listen to the speech samples in order to acquaint yourself with the task and the range of samples, with respect to degree of acceptability. We will do this now.

After you have acquainted yourself with the range and the task, 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?

Name

Observer's Answer Sheet

Scale samples as either 1,2,3,4,5,6,7

_____ 1.	_____ 21.	_____ 41.
_____ 2.	_____ 22.	_____ 42.
_____ 3.	_____ 23.	_____ 43.
_____ 4.	_____ 24.	_____ 44.
_____ 5.	_____ 25.	_____ 45.
_____ 6.	_____ 26.	_____ 46.
_____ 7.	_____ 27.	_____ 47.
_____ 8.	_____ 28.	_____ 48.
_____ 9.	_____ 29.	_____ 49.
_____ 10.	_____ 30.	_____ 50.
_____ 11.	_____ 31.	_____ 51.
_____ 12.	_____ 32.	_____ 52.
_____ 13.	_____ 33.	_____ 53.
_____ 14.	_____ 34.	_____ 54.
_____ 15.	_____ 35.	_____ 55.
_____ 16.	_____ 36.	_____ 56.
_____ 17.	_____ 37.	_____ 57.
_____ 18.	_____ 38.	
_____ 19.	_____ 39.	
_____ 20.	_____ 40.	

BIBLIOGRAPHY

Books

- Bergel, Ernest E. Social Stratification. New York: McGraw-Hill Book Company, Inc., 1962.
- Berry, Mildred F., and Eisenson, Jon. Speech Disorders. New York: Appleton-Century-Crofts, Inc., 1956.
- Biesanz, John, and Biesanz, Mavis. Modern Society. Englewood Cliffs: Prentice-Hall, Inc., 1964.
- Bloom, B. S.; Davis, A.; and Hess, R. Compensatory Education for Cultural Deprivation. New York: Holt, Rinehart, and Winston, 1965.
- Chomsky, Noam, and Halle, Morris. The Sound Patterns of English. New York: Harper and Row Publishers, 1968.
- Downie, N. M., and Heath, R. W. Basic Statistical Methods. New York: Harper and Row Publishers, 1965.
- Edwards, Allen L. Techniques of Attitude Scale Construction. New York: Appleton-Century-Crofts, Inc., 1957.
- Eisenson, Jon, and Ogilvie, Mardel. Speech Correction in the Schools. New York: The Macmillan Company, 1965.
- Gordon, Milton M. Social Class in American Sociology. Durham, North Carolina: Duke University Press, 1958.
- Griffin, John I. Statistics: Methods and Application. New York: Holt, Rinehart, and Winston, 1962.
- Guilford, J. P. Psychometric Methods. New York: McGraw-Hill Book Company, 1936.
- Guilford, J. P. Fundamental Statistics in Psychology and Education (4th ed.) New York: McGraw-Hill Book Company, 1956.

- Johnson, Wendell; Darley, Frederic L.; and Spriestersbach, D. C. Diagnostic Methods in Speech Pathology. New York: Harper and Row Publishers, 1963.
- Passow, A. H. (ed.). Education In Depressed Areas. New York: Columbia University, 1963.
- Siegel, Sidney. Nonparametric Statistics For the Behavioral Science. New York: McGraw-Hill Book Company, Inc., 1956.
- Travis, Lee Edward (ed.). Handbook of Speech Pathology. New York: Appleton-Century-Crofts, Inc., 1938.
- Tumin, Melvin M. Social Stratification. Englewood Cliffs: Prentice-Hall, Inc., 1957.
- Warner, W. L.; Meeker, M.; and Eella, K. Social Class in America. Chicago: Science Research Associates, 1949.
- Winer, B. Statistical Principles in Experimental Design. New York: McGraw-Hill Book Company, Inc., 1962.

Articles

- Altman, Mary F. "An Investigation of Scaled Severity of Whistled "S." Unpublished Master's Thesis, Department of Speech Pathology and Audiology, Eastern Illinois University, 1968.
- Ausubel, D. P. "How Reversible Are the Cognitive and Motivational Effects of Cultural Deprivation?" Urban Education, 1, 1964, 16-38.
- Becky, R. E. "A Study of Certain Factors Related to Retardation of Speech." Journal of Speech Disorders, 7, 1942, 223-249.
- Black, J. W. "Accompaniments of Word Intelligibility." Journal of Speech and Hearing Disorders, 17, No. 4, 1952, 409-418.
- Bloomfield, L. "Literate and Illiterate Speech." American Speech, 2, 1927, 432-439.
- Cazden, Courtney B. "Subcultural Differences in Child Language." Merrill Palmer Quarterly, 12, 1966, 185-219.

Cullinan, W. L.; Prather, Elizabeth M.; and Williams, Dean E. "Comparison of Procedures for Scaling Severity of Stuttering." Journal of Speech and Hearing Research, 6, No. 2, 1963, 187-194.

Curry, R.; Kennedy, L.; Wagner, L.; and Wilke, W. "A Phonographic Scale for the Measurement of Defective Articulation." Journal of Speech Disorders, 8, No. 2, 1943, 123-126.

Daniel, Artie A., and Giles, Douglas E. "A Comparison of the Oral Language Development of Head Start Pupils With Non-Head Start Pupils." Micro-film, ED 010 848, August, 1966.

DeMuth, Blaise. "Comparison of Articulation Severity Ratings By Mothers, Teachers, and Speech Pathologists." Unpublished Master's Thesis, Department of Speech Pathology and Audiology, Eastern Illinois University, 1969.

Deutsch, M. "The Disadvantaged Child and the Learning Process." in A. H. Passow (ed.), Education in Depressed Areas, New York: Columbia University, 1963, 163-179.

Everhart, R. W. "A Study of Certain Factors Related to Retardation of Speech." Speech Monographs, 23, 1956, 75-77.

Faircloth, Marjorie A., and Faircloth, Samuel R. "An Analysis of the Articulatory Behavior of a Speech-Defective Child in Connected Speech and in Isolated-Word Responses." Journal of Speech and Hearing Disorders, 35, No. 1, 1970, 51-61.

Friedlander, George H. "Reports on the Articulatory and Intelligibility Status of Socially Disadvantaged Preschool Children." Micro-film, ED 014 321, 1965.

Gussow, J. "Language Development in Disadvantaged Children." IRCD Bulletin, 1965, 1-5.

Harms, L. S. "Listener Comprehension of Speakers of Three Status Groups." Language and Speech, 4, 1961, 109-112.

Henrikson, E. H. "An Analysis of Woods Articulation Index." Journal of Speech and Hearing Disorders, 13, 1948, 233-235.

Hurst, Charles G., Jr. "Basic Factors Relating to Development of A Dialect By Disadvantaged Children" in Jerry Griffith and Lynn Miner (ed.) The First Lincolnland Conference on Dialectology, University of Alabama Press, 1970, 45-61.

- Irwin, O. C. "Reliability of Infant Speech Sound Data." Journal of Speech Disorders, 10, 1945, 227-235.
- Irwin, O. C. "Infant Speech Sounds and Intelligibility." Journal of Speech Disorders, 10, 1945, 293-295.
- Irwin, O. C. "The Effect of Family Occupational Status and of Age on Speech Sound Frequency." Journal of Speech and Hearing Disorders, 13, 1948, 320-323.
- Irwin, O. C. "Speech Development in the Young Child." Journal of Speech and Hearing Disorders, 17, 1952, 269-278.
- James, Sylvia E. "Transformational Skills of Culturally Disadvantaged and Culturally Advantaged Children." Unpublished Master's Thesis, Department of Speech Pathology and Audiology, Eastern Illinois University, 1968.
- Jordan, Evan P. "Articulation Tests Measures and Listeners Ratings of Articulation Defectiveness." Journal of Speech and Hearing Research, 3, No. 3, 1960, 303-319.
- McCarthy, Dorothea. "Factors That Influence Language Growth." Elementary English, 29, 1952, 421-428.
- McCarthy, Dorothea. "Language Disorders and Parent-Child Relationships." Journal of Speech and Hearing Disorders, 19, 1954, 514-523.
- Milisen, R. "A Rationale For Articulation Disorders." Journal of Speech and Hearing Disorders, Monograph Supplement No. 4, 1954, 4-17.
- Milner, Esther. "A Study of the Relationship Between Reading Readiness in Grade One School Children and Patterns of Parent-Child Interaction." Child Development, 22, 1951, 95-112.
- Miner, L. E. "A Normative Study of the Length-Complexity Index for Five Year Old Children." Unpublished Doctoral Dissertation, University of Illinois, 1970.
- Morrison, Sheila. "Measuring the Severity of Articulation Defectiveness." Journal of Speech and Hearing Disorders, 20, No. 4, 1955, 347-351.
- Perrin, Elinor H. "The Rating of Defective Speech by Trained and Untrained Observers." Journal of Speech and Hearing Disorders, 19, No. 1, 1954, 49-51.
- Rees, Maryjane. "Some Variables Affecting Perceived Harshness." Journal of Speech and Hearing Research, 1, No. 2, 1958, 155-168.

- Reid, Gladys. "The Efficiency of Speech Re-education of Functional Articulation Defectives in the Elementary Schools." Journal of Speech Disorders, 12, No. 3, 1947, 301-313.
- Roe, Vivian, and Milisen, R. "The Effect of Maturation Upon Defective Articulation in Elementary Grades." Journal of Speech Disorders, 7, No. 1, 1942, 37-50.
- Sherman, Dorothy. "Reliability and Utility of Individual Ratings of Severity of Audible Characteristics of Stuttering." Journal of Speech and Hearing Disorders, 20, No. 1, 1955, 11-16.
- Sherman, Dorothy and Lewis, Don. "Measuring the Severity of Stuttering." Journal of Speech and Hearing Disorders, 16, No. 4, 1951, 320-326.
- Sherman, Dorothy, and Linke, Eugene. "The Influence of Certain Vowel Types on Degree of Harsh Voice Quality." Journal of Speech and Hearing Disorders, 17, No. 4, 1952, 401-408.
- Sherman, Dorothy, and Moodie, Catherine E. "Four Psychological Scaling Methods Applied to Articulation Defectiveness." Journal of Speech and Hearing Disorders, 22, No. 4, 1957, 606-607.
- Sherman, Dorothy, and Morrison, Sheila. "Reliability of Individual Ratings of Severity of Defective Articulation." Journal of Speech and Hearing Disorders, 20, No. 4, 1955, 352-356.
- Sherman, Dorothy, and Trotter, William D. "Correlation Between Two Measures of the Severity of Stuttering." Journal of Speech and Hearing Disorders, 21, No. 4, 1956, 426-429.
- Sherman, Dorothy, and Silverman, Franklin H. "Three Psychological Scaling Methods Applied to Language Development." Journal of Speech and Hearing Research, 11, No. 4, 1968, 837-841.
- Shriner, Thomas H. "Sociolinguistics and Language," in Travis, L. E. (ed.) Handbook of Speech Pathology. New York: Appleton-Century-Crofts, Inc., 1969 (in preparation).
- Shriner, Thomas H. "Social Dialect and Language," in Jerry Griffith and Lynn Miner (eds.) First Lincolnland Conference on Dialectology, University of Alabama, 1970, 63-80.
- Shriner, T. H., and Miner, L. E. "Morphological Structures in the Language of Disadvantaged and Advantaged Children." Journal of Speech and Hearing Research, 11, 1968, 605-610.

- Siegel, Gerald M. "Experienced and Inexperienced Articulation Examiners." Journal of Speech and Hearing Disorders, 27, No. 1, 1962, 29-35.
- Silverman, Franklin H. "Intraclass Correlation Coefficient as an Index of Reliability of Median Scale Values for Sets of Stimuli Rated by Equal-Appearing Intervals." Perceptual Motor Skills, 26, 1968, 878.
- Snow, Katherine, and Milisen, Robert. "The Influence of Oral Versus Pictorial Presentation Upon Articulation Test Results." Journal of Speech and Hearing Disorders, Monograph Supplement No. 4, 29-36.
- Stitt, Clyde L., and Huntington, Dorothy A. "Reliability of Judgments of Articulation Proficiency." Journal of Speech and Hearing Research, 6, No. 1, 1963, 49-56.
- Templin, Mildred C. "Spontaneous Versus Imitated Verbalizations in Testing Articulation in Preschool Children." Journal of Speech Disorders, 12, No. 3, 1947, 293-300.
- Templin, Mildred. "Certain Language Skills in Children." The Institute of Child Welfare Monograph Series, No. 26, 1957.
- Templin, Mildred C. "Relation of Speech and Language Development to Intelligence and Socioeconomic Status." Volta Review, 60, 1958, 331-334.
- Towman, Peggy. "Oral Language of the Disabled Child." Journal of Learning Disabilities, 2, No. 6, 1969, 330-338.
- Van Riper, C. "Persistence of Baby Talk Among Children and Adults." Elementary School Journal, 38, 1948, 672-675.
- Weaver, Carl H.; Furbee, Catherine; and Everhart, R. W. "Paternal Occupational Class and Articulatory Defects in Children." Journal of Speech and Hearing Disorders, 25, 1960, 170-175.
- Wood, K. S. "Measurement of Progress in Correction of Articulatory Speech Defects." Journal of Speech and Hearing Disorders, 15, 1949, 171-174.
- Wright, H. N. "Reliability of Evaluations During Basic Articulation and Stimulation Testing." Journal of Speech and Hearing Disorders, Monograph Supplement No. 4, 1954, 19-27.

Young, Martin A. "Observer Agreement: Cumulative Effects of Rating Many Samples." Journal of Speech and Hearing Research, 12, No. 1, 1969, 135-143.

Young, Martin A. "Observer Agreement: Cumulative Effects of Repeated Ratings of the Same Samples and of Knowledge of Group Results." Journal of Speech and Hearing Research, 12, No. 1, 1969, 144-155.

Young, Martin A., and Downs, Tom D. "Testing the Significance of the Agreement Among Observers." Journal of Speech and Hearing Research, 11, No. 1, 1968, 5-17.