THE WATTS ARTICULATION TEST FOR SCREENING:

AN EVALUATION OF A SCREENING TEST

by

SHIRLEY ANN WATTS, B.A. in Ed.

A THESIS

IN

SPEECH PATHOLOGY AND AUDIOLOGY

Submitted to the Graduate Faculty
of Texas Tech University in
Partial Fulfillment of
the Requirements for
the Degree of

MASTER OF SCIENCE

IN

SPEECH PATHOLOGY AND AUDIOLOGY

Approved

August, 1971

ACKNOWLEDGMENTS

My sincere appreciation is extended to Dr. Earlene Tash Paynter for her assistance in directing the writing of this thesis.

I wish to thank Dr. James T. Yates for his suggestions and interest. I wish to acknowledge Jan Taylor for her assistance in testing the subjects for this study. I am also grateful to the two graduate students who made the tape recording for use in this investigation. To Bob Haynes and Bill Coberly, who served as my statistics consultants, I wish to extend my appreciation.

I am grateful to Max Manley and the administrators of the Lubbock Public Schools for their cooperation in securing the subjects for this study.

TABLE OF CONTENTS

ACKNOWLEDGMENTS	ii
LIST OF TABLES	v
LIST OF ILLUSTRATIONS	vii
I. INTRODUCTION	1
Development of Articulation Testing	4
Discussion	22
Statement of Purpose	26
II. METHODS AND PROCEDURES	28
Selection of Subjects	29
Test Description	29
Templin-Darley Screening Test	30
Watts Articulation Test for Screening	31
Examiner Reliability	40
Administration of Tests	41
Statistical Analysis	43
III. RESULTS AND DISCUSSION	45
Administration Time	45
Familiarity of Test Words	53
Interest of Tests	56
Identification of Children	59
Value of the Record Chart	62
IV. SUMMARY AND CONCLUSIONS	64
Purpose of Study	64
Procedures	65

Results	67
Conclusions	71
Implications	71
LIST OF REFERENCES	73
APPENDIX	76

LIST OF TABLES

Table	Page
1. Cut-off Scores for the Templin-Darley Screening Test	31
2. List of "WATS" Stimulus Words and the Sounds Tested	34
3. List of "WATS" Stimulus Questions and Statements and Desired Responses	35
4. Watts Articulation Test for Screening Record Chart	37
5. Cut-off Scores for the Watts Articulation Test for Screening	38
6. Total Administration Times of the Templin-Darley Test and the Watts Articulation Test for Screening	46
7. Results of T-Test Computed on the Administration Times of the Templin-Darley Test and the Watts Articulation Test for Screening	47
8. Totals, Means, and Standard Deviations in Minutes of the Administration Times of the Templin-Darley Screening Test and the Watts Articulation Test for Screening for Each Subgroup	48
9. Results of T-Test Computed on the Administration Times of the Templin-Darley Screening Test and the Watts Articulation Test for Screening for Each Subgroup	51
10. Total Number of Unfamiliar Stimulus Words and List of the Five Most Unfamiliar Stimulus Words from the Templin-Darley Screening Test and the Watts Articulation Test for Screening	54
11. Results of the Chi Square Computed on the Total Number of Familiar and Unfamiliar Templin-Darley Test and "WATS" Stimulus Words Arranged in a 2 X 2 Contingency Table	56

12.	Results of the Chi Square Computed on the Five Most Unfamiliar Templin-Darley Test and "WATS" Stimulus Words Arranged in a 2 X 2 Contingency Table	57
13.	Number of Subjects Who Preferred the Templin- Darley Test and the Number Who Preferred the Watts Articulation Test for Screening	58
14.	Subjects Identified by the Templin-Darley Screening Test and the Watts Articulation Test for Screening for Diagnostic Evaluation .	60
15.	Descriptive Statistics of the Pearson Product- Moment Coefficient of Correlation Computed on the Templin-Darley and "WATS" Test Scores .	62

LIST OF ILLUSTRATIONS

Figure	Page
1. The Watts Articulation Test for Screening	32
2. Seating Arrangement for the Watts Articulation Test for Screening	39

CHAPTER I

INTRODUCTION

The public school speech clinician has traditionally been involved in assessing articulatory skills of children in an attempt to identify those with defective articulation. Two primary types of tests, screening and diagnostic tests, have been employed. A screening test is used to identify "satisfactory or unsatisfactory articulation" (Templin, 1953) while a diagnostic test is used to determine the "accuracy and inaccuracy of articulation of specific speech sounds" (Templin, 1953).

According to the Midcentury Whitehouse Conference report, the total estimate of the incidence of speech defects among children in the United States between the ages of 5 and 21 years was 5 per cent. Functional articulation disorders constituted 3 per cent of the total estimate (Kirk, 1962). In the school-age population the incidence of articulation disorders has been reported from 2.5 per cent (Dunn, 1963) to 7.0 per cent (Mills and Streit, 1942). The relative frequency of the different types of speech defects is difficult to determine accurately; however, the great majority of these are most likely defects of articulation (Van Riper, 1963).

These articulation disorders constitute a large portion

of the case load of the public school speech clinician. Kirk (1962) has estimated some 80 to 85 per cent of the speech clinician's cases will display defects of articulation. Johnson, Darley, and Spriestersbach (1963) stated that public school speech clinicians reported 81 per cent of their average case load consisted of children with functional articulation problems. Anderson (1953) estimated some 75 per cent of the total speech defects among public school children are various types of articulation disorders, including delayed speech.

In selecting the case load, the speech clinician usually conducts a speech screening survey to identify speech defective children who need diagnostic evaluation. After these children are selected, the clinician administers a diagnostic test to determine the type and severity of the disorder. From the information gathered in this examination, the public school clinician determines whether or not the child needs speech therapy.

Often, the speech survey involves screening large numbers of children in a relatively short period of time. The task is further complicated by the fact that the tests available for screening articulatory skills are time consuming and many times diagnostic in nature. Screening procedures, theoretically, are designed to separate individuals according to skills, personality, or aptitudes. They are

not designed to measure the type or degree of impairment; they are primarily employed to identify only the existence of impairment. Articulation screening tests specifically serve to identify those children who display faulty articulation while diagnostic tests provide a more thorough measurement to determine the type and severity of the disorder.

To date, screening tests of articulation have been so similar in content and method of administration to diagnostic tests that public school speech clinicians have been unable to differentiate between the two. As a result, the clinician has attempted to conduct a speech screening survey using diagnostic tests. Numerous hours have been spent turning picture cards and transcribing misarticulated speech sounds phonetically. Valuable time and needless effort have been wasted because screening tests have failed to simply identify children with articulation defects. Screening tests have assumed a diagnostic nature, losing sight of the original premise from which they were developed.

The purpose of the present study was to develop and evaluate a screening test of articulation called the Watts Articulation Test for Screening (WATS) which would satisfy the following criteria: (1) identify children who need diagnostic examination, (2) eliminate unnecessary stimulus materials, (3) save time, (4) provide a simple form for

recording responses, (5) provide stimulus words which are familiar to school-age children, and (6) provide an interesting articulation screening test for school-age children.

Development of Articulation Testing

Wood (1957) states that articulation is,

. . . literally the state of being united by a joint or joints; in speech, the production of individual speech sounds in connected discourse; the movement and placement during speech of the organs which serve to interrupt or modify the voiced or unvoiced air stream into meaningful sounds; the speech function performed largely through the movements of the lower jaw, lips, tongue, and soft palate.

For use in the present study, the term articulation was conceived as a process whereby speech sounds are produced through certain movements of the speech organs; namely, the tongue, lips, jaw, teeth, soft palate, and cheeks.

Disorders of articulation have been defined (Dunn, 1963; Kirk, 1962; and Van Riper, 1963) as those deviations which involve substitutions, omissions, distortions, and additions of consonants or vowels or both. A substitution is a type of error in which a child substitutes one sound for another. An omission is an error in which a child omits a sound. A distortion is an attempt to produce the correct sound which results in a distorted imitation. An addition is an error in which additional consonants or vowels are produced in words or between them. One or all four types of errors can exist in any degree of severity within the speech pattern

of a given child (Dunn, 1963). Disorders of articulation are commonly classified as mild, moderate, or severe.

In an effort to identify and diagnose children with communication disorders, many authors have devised various types of articulation tests, including screening, diagnostic, proficiency, developmental, and prognostic tests. Van Riper and Irwin (1958) state that,

An articulation test, as the term is generally used, is a technique employed to measure the general phonemic capacity of an individual. . . . The purpose of the articulation test, then, is to compare the phonemes that are actually used by an individual with the phonemic structure of his language group.

The nature of the articulation test varies according to the purpose of the measurement (Darley, 1964). If the speech clinician wishes to determine the general adequacy of a child's articulation, a screening test is usually employed. The ideal articulation screening test should identify 100 per cent of those children with articulation defects and refer those who need diagnostic evaluation. The test should be simple and quick to administer and score; and it should be interesting to children. If the clinician wants to obtain a more detailed description and evaluation of the child's articulatory skills, a diagnostic test is commonly used (Templin and Darley, 1960). A diagnostic test should gather information about a child's ability to produce a wide range of speech sounds in a variety of positions and phonetic

contexts (Templin and Darley, 1960). Close examination of previously constructed articulation tests will reveal many deficiencies which are characteristic of the tests used by the public school speech clinician in the speech screening survey.

Among the first to standardize an articulation test were Templin and Darley (1960). In the 176-item test, there are 57 cards which contain single line drawings to elicit a spontaneous response for each test word. The first 16 cards constitute the Screening Test of 50 items. The sentence portion of the test, consisting of 89 sentences containing the test words, is used to obtain responses from older children for whom pictures would be inappropriate (Templin and Darley, 1960).

The sounds in the 176-item test were taken from previous research conducted by Templin in 1947 in which 100 preschool and kindergarten children were tested on 113 sound elements. The 176-item test includes 25 different consonant sounds, 12 vowels, and 6 diphthongs distributed among eight sound classifications (Templin and Darley, 1960). The Screening Test contains 22 single consonants, 22 two-consonant blends, 4 three-consonant blends, and 2 vowels.

Spaces on the scoring form designated by a double line are for recording responses to the Screening portion of the test. The five types of symbols used are: (1) a [✓] for

correct responses, (2) phonetic symbols for substitutions, (3) a [-] for omissions, (4) an [X] for distortions, and (5) an [NR] for no response (Templin and Darley, 1960). For scoring consonant blends, the same code is used.

Test norms were developed on 480 children, ages 3 through 8 years, as part of a study of language development in children (Templin, 1953). The children were white, monolingual singletons of normal intelligence, displaying no gross evidence of hearing loss. They were enrolled in fourteen public schools and twenty-one nursery schools and selected to be representative of the United States urban population (Templin and Darley, 1960). The diagnostic and screening tests were both found valid and reliable measures (Templin, 1947a, 1953).

In 1969, Templin and Darley revised the original Templin-Darley Test of Articulation (1960) into a condensed set of 141 items. The first 50 items may be used as a Screening Test; another set of 43 items constitutes the Iowa Pressure Articulation Test; and other sets of items permit separate testing of single consonants in the initial and final positions, consonants in clusters, vowels, and diphthongs (Templin and Darley, 1969).

The Revised Test still uses 57 picture cards; however, the pictures are now, for the most part, in color. Most of the stimulus words are those used in the first edition of

the test. The new words in the Revised Test were chosen because "they could be more clearly and interestingly illustrated and still be familiar to young children" (Templin and Darley, 1969). Questions and stimulus statements are written on the backs of the cards in a manner similar to the original test to enable the clinician to elicit spontaneous responses from the child (Templin and Darley, 1969).

A sentence test of 141 sentences incorporates two test words for each sound element tested in the picture test. Single consonants in the medial position have been omitted from the Revised Test except for a part of the previously devised Screening Test and Iowa Pressure Articulation Test. Consonant clusters are incorporated into the Revised Test and the final consonant /r/ sound is not included in the new form (Templin and Darley, 1969).

Revisions were also made in the Test Form used to record responses. The first 50 items are separated from the other items on the form to designate the Screening Test. Consonant singles, pressure consonants, /l/ clusters, /s/ clusters, and vowels and diphthongs each occupy separate sections on the new Test Form (Templin and Darley, 1969).

Norms are furnished for the evaluation of performance on the Revised Diagnostic Test, and the two previously devised tests, and the nine separate groupings of sound elements (Templin and Darley, 1969). Cut-off scores on the

50-item Screening Test, reproduced from Templin (1957), are provided in tabular form in both original and revised Templin-Darley Tests (Templin and Darley, 1969).

Another diagnostic test, the McDonald Deep Test was published in 1964. In this test, each consonant sound is preceded by each of the other consonant sounds and followed by a vowel. Then, each consonant sound is preceded by each of the vowel sounds and followed by each of the other consonant sounds. A word combination like cupsun is used to test the /s/ sound in the initial position. The /s/ is preceded by the consonant sound /p/ and followed by the vowel sound /ʌ/. To test the /s/ sound in the final position, a word combination such as housepipe is employed. The /s/ is preceded by the vowel sound /ʌ/ and followed by the consonant sound /p/. McDonald used these combinations of 25 consonant sounds and 10 vowels to investigate a variety of patterns of "overlapping ballistic movements" (McDonald, 1964) on the production of the sound being studied.

The Deep Test was constructed in both picture and sentence form. The picture test is made up of two sets of pictures mounted side by side in a way which permits each picture to be individually manipulated by the examiner (McDonald, 1964). Pictures on the left contain the sounds to be tested in the final position while those on the right contain the sounds to be tested in the initial position.

Pictures used for the Deep Test were those readily identifiable by 100 kindergarten children. Only those pictures with the highest recognition percentages were selected for use in the test (McDonald, 1964).

The Sentence Form of the Deep Test was constructed for the more complex consonants. The vocabulary level employed was that of easy reading for the average third grader (McDonald, 1964).

Responses are recorded by writing the phonetic symbol of the sound being tested in brackets at the top of a column of consonants and vowels. This sound is then combined with the other sounds in the column. The number of correct responses is divided by the number of phonemes tested and multiplied by 100 to determine the percentage of correct articulations (McDonald, 1964). No normative data is available concerning the consistency with which each sound is correctly articulated at the different age levels.

The Goldman-Fristoe Test of Articulation was developed in 1969. This test contains a Sounds-in-Words Subtest of 35 pictures, a Sounds-in-Sentences Subtest of 9 pictures, and a Stimulability Subtest located on 2 cards following the Sounds-in-Sentences Subtest. Key words are coded by color and number. Recording responses involves locating each test sound in the appropriate response matrix and recording the various symbols specified in the test booklet

which represent the types of errors (Goldman and Fristoe, 1969).

Test-retest reliability was determined for the first two Subtests. There were 37 children with articulatory defects, ages 4 to 8 years who were tested by eight speech pathologists. All the speech pathologists met the American Speech and Hearing Association's requirements for the Certificate of Clinical Competence (Goldman and Fristoe, 1969). All subjects were tested twice by the same examiner with one week intervening between each test. Test-retest comparisons were made on the basis of recurrence of error. For the Sounds-in-Words Subtest, the median agreement was 95 per cent. Median agreement for the Sounds-in-Sentences Subtest was 94 per cent. Agreement according to the type of production was 89 per cent for the Sounds-in-Words Subtest and 96 per cent for the Sounds-in-Sentences Subtest (Goldman and Fristoe, 1969). Information concerning the test-retest reliability for the Stimulability Subtest has not yet been obtained (Goldman and Fristoe, 1969).

On the basis of presence or absence of error, median percentage agreement for interjudge reliability for the Sounds-in-Words Subtest was 92 and the median percentage agreement for the type of production was 88. Intrajudge reliability determined for both presence or absence of error and type of production was a median percentage agreement of

91. Interjudge and intrajudge reliability information has not been obtained for the Sounds-in-Sentences Subtest (Goldman and Fristoe, 1969).

Another investigator, William Clark Rodgers (1968) designed the Picture Articulation and Language Screening Test (PALST). This test is an attempt to provide the public school speech clinician with an effective screening test to determine articulation and language abilities (Rodgers, 1968).

The PALST is divided into three units. Unit one is the Initial Screening portion, unit two is the Sound Evaluation portion, and unit three is the Stimulus Materials Section. Each of these three units may be used to evaluate either or both articulation and language abilities (Rodgers, 1968).

The picture stimuli are presented to the child in order that the clinician can evaluate the child's articulation of certain selected sounds and judge the child's ability to formulate certain verbal expressions (Rodgers, 1968). The first 6 picture cards elicit the language sample and the articulation sample. The 3 other picture cards elicit the child's production of /s/, /ʃ/, and /tʃ/ in isolation.

To score the articulation portion of the PALST, omissions, substitutions, and distortions are recorded in the standard manner (Rodgers, 1968). Errors are recorded in the initial and final positions. A numerical score, based on

clinician's judgment of that child's articulatory ability, is designated for the articulation section of the test. The score is placed in the "additional observations" column (Rodgers, 1968). Guidelines used in the scoring procedures are provided by the author. He suggests that a score of 1 be given to a child having many defective sounds. Children with lesser degrees of severity of problems would receive higher scores up to 5 (Rodgers, 1968). Although a child receiving a score of 5 would have an articulation problem, prognosis for spontaneous improvement without therapy would appear good and he would be placed on the waiting list. A child receiving a score of 1 or 2 would have poor prognosis for spontaneous improvement; therefore, he would be enrolled in therapy (Rodgers, 1968).

Articulation tests, characteristically, have measured only 1 test sound per word (Templin and Darley, 1960, 1969; Hejna, 1959; and Carter and Buck, 1958). Irwin and Musselman (1962) devised an articulation test to check more than 1 sound in a word. The two conditions employed in testing subjects were face-to-face and tape recordings. Sounds tested in twenty-seven words, represented by original India ink line drawings, consist of all the consonants in the initial and final positions, all the vowels, and all the diphthongs (Irwin and Musselman, 1962).

All subjects were enrolled in speech therapy and had

normal hearing, normal intelligence, and no indication of cerebral palsy, brain damage, cleft palate, or stuttering. There were 14 children, ages 4 to 8 years, who served as subjects for the face-to-face testing situation. There were 14 children, ages 5 to 7 years, who were recorded on tape. There were 8 children common to both parts of the study (Irwin and Musselman, 1962).

The Compact Picture Test and the conventional type of articulation test, where a separate word is used for each sound tested, were administered alternately to each child. Experienced and inexperienced clinicians evaluated the taped speech responses to the two tests. No significant difference was found between the two types of tests in the ranking of children according to total number of errors (Irwin and Musselman, 1962). No significant difference was found between experienced and inexperienced judges. The original Compact Picture Test developed may be used to evaluate more than 1 sound in a word with reliability by both experienced and inexperienced judges (Irwin and Musselman, 1962).

A unique idea in articulation testing was presented by Goldman, Fristoe, and Thomas (1965) when they studied the value of a light projected test. Colored slides of objects which are familiar to young children were used to obtain a measure of articulatory proficiency. This test was given to children in four groups: (1) functional misarticulation,

(2) misarticulation with mental retardation, (3) cerebral palsy, and (4) delayed language development (Goldman, et al., 1965). Comparisons of responses to the light projected test and the Templin-Darley Test and the Bryngleson-Glaspey Test were made. Results indicated that the slide test was more quickly administered than the Templin-Darley Test and about equal in time to the Bryngelson-Glaspey (Goldman, et al., 1965). More spontaneous responses in naming the pictures were obtained with the slide procedure than with the Templin-Darley or Bryngelson-Glaspey Test. Other reported advantages to the slide test were higher interest value, less distractions, better focusing of attention through light and color, and increased stimulus size (Goldman, et al., 1965).

Goldman and Fristoe (1967) attempted to make modifications in the original test. A portable filmstrip viewer with its own rear projection screen was devised to eliminate the bulky equipment employed to project the stimulus material. There are 35 pictures, representing objects familiar to young children, which make up the complete Filmstrip Articulation Test (Goldman and Fristoe, 1967). A total of 44 responses are obtained by asking eight questions about the pictures in addition to the names of the 35 objects. A total of 72 sounds, in various positions in words, are tested. Often one word is utilized to elicit several of these test sounds (Goldman and Fristoe, 1967).

A coding system for recording responses was devised in which numbers, printed above the names of the stimulus pictures, indicate the phonemes being tested. These numbers refer the examiner to the appropriate blanks on the developmental chart (Goldman and Fristoe, 1967).

Other studies (Barker, 1960; Barker and England, 1962; and Fudala, 1970) have supplied numerical scores to represent an individual's articulatory proficiency. In Barker's original study in 1960, two Articulation Scores were devised. A picture articulation test was administered to 45 subjects ranging in age from 6 to 12 years, to determine the effectiveness of the Articulation Scores. All the articulation test sounds were given numerical values related to their probable frequency of occurrence in American speech (Barker, 1960). The values of all misarticulated sounds were added together and subtracted from 100 per cent. The arithmetic score represented the adequacy of the individual's articulation. There were ten trained judges who rated the speech of the 45 subjects. A coefficient of correlation of .94 was obtained between the Articulation Scores and the judges' ratings.

Modifications in the original Articulation Scores were made in 1962 by Barker and England. A detailed scale adding consonant blends and separating consonant and vowel /r/ sounds was devised. A simplified scale was also constructed

for use by speech clinicians (Barker and England, 1962). The detailed and simplified scales were then used to recompute the data from the 45 subjects of Barker's first study conducted in 1960.

All error values in the consonant and vowel columns were totaled and then added together. According to Barker and England (1962), the grand total was then subtracted from 100 to obtain the Arizona Articulation Proficiency Score (AAPS). Developmental ages for the sounds tested were based on Templin's 1957 study of 480 children (Fudala, 1970). The correlations between the judges' ratings in Barker's (1960) original study and the values obtained from the detailed and simplified scales were .93 and .92 respectively (Barker and England, 1962). The authors concluded that each of the scales was a valid measure of articulatory proficiency.

In 1970, Janet Barker Fudala made revisions in the AAPS. The major changes are: (1) elimination of the medial position, (2) average AAPS Scores for each age level 3 through 11, (3) addition of a Sentence Test, (4) addition of consonant symbols after phonetic symbols, (5) addition of 2 stimulus picture cards to elicit spontaneous speech, (6) guidelines for a gross measure of language screening for children 4 years and older, (7) a four page Protocol Booklet for recording responses, and (8) a Survey Form for articulation records of 10 children (Fudala, 1970).

The original Picture Test Cards make up, for the most part, the revised AAPS. Those pictures used to stimulate medial sounds were eliminated, the order of picture presentation has been changed, and some substitutions for original pictures have been made (Fudala, 1970). Each card still contains only 1 stimulus picture. All but 2 pictures are black and white line drawings. These 2 pictures are black line drawings on a red background and a green background.

Test-retest reliability of .96 was obtained by nineteen public school speech clinicians testing 105 children. Validity was determined for the AAPS by ten judges who rated two 10-second tape recorded samples of the spontaneous speech of 45 children. These ratings were correlated with AAPS Scores obtained for the same 45 children on a picture articulation test and a coefficient of correlation of .92 was obtained (Fudala, 1970). The author concluded that the AAPS measures whatever listeners evaluate in defective articulation.

Norms for ages by which 90 per cent of the children tested had mastered each sound were determined from the data of 702 children from average elementary schools in Seattle. The children, ages 3 to 11 years, displayed no gross deviations in mental ability, emotional stability, hearing or neurological functioning (Fudala, 1970).

New scoring procedures have been incorporated in the revised AAPS. Only errors are recorded in the following manner:

(1) record substitutions, (2) record omissions by a [-], (3) record distortions by an [X] (Fudala, 1970). Corresponding values are circled for each error and totaled for the Consonant Score and the Vowel Score. These two Scores are added together to obtain the Total Sound Score, which, in turn, is subtracted from 100.0 to obtain the AAPS Total Score (Fudala, 1970). This score indicates the percentage of sounds articulated correctly. Degrees of articulatory proficiency are listed in Table 1 of the Protocol Booklet (Fudala, 1970).

Hejna (1959) designed an articulation test to assess consonant sounds on a developmental scale. Sounds are tested according to various chronological age levels by which at least 90 per cent of all children use the sounds correctly (Hejna, 1959). Each sound tested has a specific age level at which it is correctly produced by 90 per cent or more children. The final /b/, /g/, and /z/ sounds are exceptions to the 90 per cent rule (Hejna, 1959). These sounds should be produced correctly at the age levels specified by approximately 75 per cent of all children. The developmental age levels are approximate, and are based on "normative data bearing on this subject" (Hejna, 1959). To date, the normative data Hejna refers to has not been specified.

There are 26 cards, containing from 2 to 5 pictures each, which are used to test consonant sounds in the initial, medial,

and final positions with the exception of the /ŋ/ sound which is tested only in the medial and final positions (Hejna, 1959). The /ʒ/ and /hw/ sounds are omitted from the test. There are 14 consonant blends tested at Developmental Age Levels 6, 7, and 8 years (Hejna, 1959).

Defective sounds are marked in columns one, two, and three according to the position in the word. The child is asked to produce the sound in isolation and the response is marked in the "Iso." column (Hejna, 1959). The errors on the Hejna Developmental Articulation Test (1959) are scored as substitutions, g/d; omissions, -/d; or distortions, Dist./d.

Attempts have been made by several other investigators (Carter and Buck, 1958, and Van Riper and Erickson, 1969) to devise prognostic or predictive tests of articulation. These tests are designed to identify those children with articulation defects who can be expected to overcome their errors without therapy from those who will need professional help to correct their errors.

Carter and Buck (1958) devised a prognostic test which measured the child's articulation responses to a picture test, an imitative test, and a nonsense-syllable test. The picture test consists of 114 pictures representing 13 sound elements in each of the three word positions. For the imitative test the child is asked to repeat the same words

used in the picture test. The nonsense-syllable test employs the same sounds in the first two tests, only in nonsense-syllables (Carter and Buck, 1958).

First grade subjects were given the picture test, the imitative test and the nonsense-syllable test in that order. If the child's articulation improved significantly on the nonsense-syllable test, Carter and Buck (1958) concluded that it was possible to predict that the child would improve his defective articulation without therapy.

Another prognostic test, Van Riper and Erickson's (1969) Predictive Screening Test of Articulation (PSTA) attempts to distinguish first grade children who would overcome their articulation errors without therapy from first grade children who would not overcome their errors. The 47-item test includes the child's imitative responses to stimulus words, isolated phonemes, nonsense-syllables, and words in sentences (Van Riper and Erickson, 1969). The child's ability to discriminate a misarticulated /r/ sound and clap his hands in replication of a simple rhythm pattern are also part of the PSTA (Van Riper and Erickson, 1969).

Each item in the PSTA is scored on a pass-fail basis. The resulting PSTA score is the number of items passed (Van Riper and Erickson, 1969). In 1965, and again in 1967, the PSTA was administered to 180 first grade children in Michigan. The mean PSTA score for those children who were

found free of misarticulations in 1967 was 35.7. The mean score for those who still had misarticulations was 29.8 (Van Riper and Erickson, 1969). The authors have suggested a cut-off score for the PSTA in order to give the speech clinician a guideline as to which children to accept for speech therapy and which children to reject. The cut-off score of 34 was selected to prevent predicting some children will overcome defective articulation without therapy when, in actuality, they will not. Secondly, the score was selected to prevent predicting that some children will still have defective articulation when they enter the third grade when, indeed, they will not have defective articulation (Van Riper and Erickson, 1969).

No data is available to validate any interpretation of the PSTA results or those obtained by Carter and Buck in 1958. The PSTA should be regarded as the first step in the development of prognostic articulation testing according to Van Riper and Erickson (1969).

Discussion

The deficiencies inherent in the articulation tests mentioned above make it impractical for the school speech clinician to conduct an effective speech screening survey. One of the most obvious deficits in test materials is that most tests (Templin and Darley, 1960, 1969; Hejna, 1959; Irwin and Musselman, 1962; Carter and Buck, 1958; McDonald,

1964; Barker and England, 1962; Fudala, 1970; Rodgers, 1968; Van Riper and Erickson, 1969; and Goldman and Fristoe, 1969) employ a variety of pictures on plates or cards to elicit spontaneous speech responses from children. These pictures, ranging in number from 9 (Rodgers, 1968) to 114 (Carter and Buck, 1958), require continual manipulation by the examiner while he is in the process of listening to, evaluating, and recording responses. The use of such materials is impractical in a screening survey designed to distinguish "acceptable from unacceptable articulation" (Templin, 1947b). Time is lost in the process of handling picture cards; yet, two of the most fundamental criteria of an effective screening test are that it be rapid and simple to use (Barker, 1960; Templin and Darley, 1960; Irwin and Musselman, 1962; Riley, 1965; and Van Riper, 1963). Thus, it appears that many screening articulation tests have assembled a much larger number of picture stimuli than is actually necessary to obtain an adequate sample of articulation.

As cited earlier, Goldman and Fristoe (1967) attempted to eliminate the manipulation of cards or objects as stimulus material by devising a Filmstrip Articulation Test using a portable filmstrip viewer. This type of test seems to have some merit as far as the immature, easily distracted, or mentally retarded child is concerned (Goldman and Fristoe, 1969). However, using a portable filmstrip viewer involves

a projector, a filmstrip, a recording form, the ability to operate a projector, and an available room conducive to viewing film. It appears that the elimination of handling objects as stimulus materials has not been accomplished to a significant degree.

A second weakness found in articulation tests (Templin and Darley, 1960, 1969; and Hejna, 1959) is that some of the stimulus words employed appear unfamiliar to and inappropriate for testing the spontaneous responses of young school-age children. In the Hejna Developmental Articulation Test, the words dog-house, spider-web, jump-rope, orange-juice, snowman, and scooter are not found in Part I "List of Words Occurring at Least Once per 1,000,000 Words" of The Teacher's Word Book of 30,000 Words (Thorndike and Lorge, 1944). Several other words such as television, toothbrush, zebra, and bib occur relatively infrequently (i.e., from 1 to 7 times per 1,000,000 words) according to Thorndike and Lorge (1944). Both the original and revised forms of the Templin-Darley Test contain the words snowman, shredded wheat, and zipper, which are not found in Part I "List of Words Occurring at Least Once per 1,000,000 Words" (Thorndike and Lorge, 1944). Since publication of the Thorndike and Lorge List, it appears reasonable to assume that the words television, toothbrush, orange-juice, jump-rope, and zipper would now be considered more familiar to school-age children.

Still another problem found in articulation testing is that tests (Templin and Darley, 1960; and Irwin and Musselman, 1962) have used line drawings rather than colored pictures to elicit responses. Line drawings employed as stimulus materials would appear to be of little interest to most young children.

A primary characteristic of the cited screening tests of articulation is that they have become so detailed and complex they no longer resemble screening devices at all, but diagnostic tests instead. Van Riper and Erickson's (1969) Predictive Screening Test of Articulation (PSTA) not only includes the child's ability to respond to words, isolated phonemes, nonsense-syllables, and words in sentences, but also his ability to recognize a misarticulated /r/ sound and to clap his hands in a simple rhythmic pattern. The Templin-Darley Test is also quite detailed. The Screening Test examines the child's production of 26 blends which appears to be an unnecessarily large number to obtain an adequate speech sample in a screening survey.

Another aspect of screening tests of articulation that is particularly detrimental to the speech survey is the method of recording responses. Many tests employ individual score sheets (Templin and Darley, 1960, 1969; Goldman and Fristoe, 1969; and Hejna, 1959) which become cumbersome when screening large numbers of children.

The method of recording errors usually involves using various symbols to represent substitutions, omissions, distortions, or additions. Most tests (Templin and Darley, 1960, 1969; Hejna, 1959; Rodgers, 1968; and Fudala, 1970) utilize this type of recording procedure. Goldman and Fristoe (1969) have suggested a system in which the coding of responses involves locating each test response in relation to color and number in the Response Matrix and, then, recording each response according to presence or absence of error or type of production. If only presence or absence of error is considered, the examiner leaves the cell blank to indicate a correct response, marks an [X] to indicate an error, and marks a [/] to indicate no response was elicited (Goldman and Fristoe, 1969). This method appears useful as a means to record responses on a screening test; however, locating the correct response matrix cell for each test sound according to color and number may become time consuming because the examiner is faced with the task of manipulating picture cards while attempting to evaluate and record responses.

Statement of Purpose

Several problems in screening tests of articulation remain unsolved: (1) test materials have proven cumbersome and time consuming, (2) specific words used to stimulate responses have been unfamiliar to and inappropriate for use

with school-age children, (3) screening devices have become too detailed and diagnostic in nature, and (4) the methods of recording responses have involved the use of complex symbols and recording forms.

The purpose of the present study was to develop the Watts Articulation Test for Screening (WATS) to meet the needs of the public school speech clinician more effectively and to determine the validity of the WATS by evaluating it against the Revised Templin-Darley Screening Test. It was believed that the use of the WATS would facilitate the speech screening survey in six ways by: (1) identifying children who need diagnostic examination, (2) utilizing one stimulus picture to eliminate unnecessary stimulus materials, (3) saving time, (4) providing a simple method for recording responses, (5) providing stimulus words which are familiar to young school-age children, and (6) providing school-age children with an interesting articulation screening test.

CHAPTER II

METHODS AND PROCEDURES

Previously constructed articulation tests have proven inadequate for use in the public school speech screening survey. The purpose of the present experiment was to develop and evaluate a screening test which could be used more effectively by the public school speech clinician. The test was specifically designed to identify children who need diagnostic evaluation, employ test words which are familiar to school-age children, be simple and quick to administer and score, and be interesting to school-age children.

A widely used screening test of articulation served as a standard to which the newly developed test was compared. Both screening tests were administered individually to 90 elementary school children. The selection of subjects and the order of presentation of the tests were randomly determined.

Various statistical measures were carried out on the data with respect to the time required to administer the tests, the familiarity of the stimulus words, the proportion of subjects who preferred the new test to the standardized test, and the scores obtained on the two tests.

Selection of Subjects

A total of 90 subjects participated in this study. Of these subjects, 60 were Caucasian, 7 were Negroid, and 23 were Mexican-American. There were 40 males and 50 females ranging in age from 6 years, 6 months to 10 years, 4 months. No subject displayed any gross visual or auditory impairment.

The subjects were chosen from three elementary schools in Lubbock, Texas. The selection of schools was made on the basis that School 1 appeared to represent the upper socio-economic level, School 2 appeared to represent the middle level, and School 3 appeared to represent the lower socio-economic level of Lubbock. From each school, a first, second, and third grade classroom were selected at random by the speech clinician of that particular school.

Letters requesting parental approval to include a child in the study were distributed to all the children in each of the classrooms. From those letters which were returned one week later, a list of names of potential subjects was compiled. Using a table of random numbers, 10 subjects in each grade from each school were selected to serve as subjects. There were 30 subjects from each school; 10 in the first grade, 10 in the second grade, and 10 in the third grade.

Test Description

There were two articulation tests employed in this study. The Watts Articulation Test for Screening (WATS) was compared

to the Templin-Darley Screening Test (1969) which served as the standard.

Templin-Darley Screening Test

The Revised Templin-Darley Screening Test utilizes 19 cards bound together in booklet form with the Templin-Darley Diagnostic Test (1969). The 19 cards contain 50 pictures representing fifty stimulus words. Only 1 sound is tested within each word. These sounds include 4 three-consonant clusters, 22 two-consonant clusters, 22 single consonants and 1 vowel sound (Templin and Darley, 1969).

The Templin-Darley Test is administered by presenting picture cards in order to obtain a spontaneous speech response. Stimulus questions and statements are printed on the back of each picture card to aid the examiner in eliciting spontaneous responses. In some instances the examiner is required to point to the desired picture.

Responses are recorded in individual Articulation Test Form booklets. Various symbols representing the type of error are employed. A [✓] indicates a correct response; a phonetic symbol represents a substitution; a [-] indicates an omission; an [X] means a distortion; [NR] indicates no response; and [NE] indicates nasal emission.

Norms for the Screening Test were based on data collected by Templin in 1957, as part of a study of language development

in children (Templin and Darley, 1969). The first edition of the Templin-Darley Diagnostic Tests was administered to 480 children, ages 3 through 8. The subjects were white, monolingual singletons of normal intelligence who displayed no gross evidence of hearing impairment (Templin and Darley, 1969). Cut-off scores on the 50-item Screening Test are provided for eight age levels and can be seen in Table 1.

TABLE 1
CUT-OFF SCORES FOR THE TEMPLIN-DARLEY
SCREENING TEST

C.A.	3	3 1/2	4	4 1/2	5	6	7	8
Cut-off Scores	12	18	23	26	31	34	39	44

Watts Articulation Test for Screening

The WATS is designed around a brightly colored picture of a farm scene painted on a 16 inch by 20 inch canvas board. The canvas, seen in Figure 1, was painted with acrylics for durability and mounted on a picture stand for easy manipulation. The complete test, which includes the canvas board and record charts, is carried in a large portfolio.

There are 42 sound elements elicited from twenty-eight stimulus words. Some of the stimulus words include the names of farm animals, colors, and parts of the body. A list of WATS stimulus words and the sounds tested can be seen in

Fig. 1.--The Watts Articulation Test for Screening

Table 2. Close examination of Part I "List of Words Occurring at Least Once per 1,000,000 Words" of The Teacher's Word Book of 30,000 Words (Thorndike and Lorge, 1944) revealed the following: of the twenty-eight WATS stimulus words, seventeen occurred 100 times or more per 1,000,000 words, four occurred at least 50 times, one occurred 49 times, one occurred 47 times, two occurred 44 times, one occurred 43 times, one occurred 27 times, and one word, zipper, was not listed.

The 42 sounds tested include 35 consonant sounds in either one, two, or all three word positions, 6 two-consonant blends, and 1 vowel /ʃ/ sound. Many of the WATS stimulus words test more than 1 sound. Previous studies (Irwin and Musselman, 1962; and Fristoe and Goldman, 1968) have indicated that accurate and reliable judgments can be made when testing several sounds within the same stimulus word.

Administration of the WATS is quite simple. The examiner is required to present several stimulus questions or statements in order to obtain spontaneous speech responses. At the same time it is necessary that the examiner point to desired pictures on the canvas board. The examiner would say, "As I point to the picture tell me what color it is." Suggested stimulus questions and statements and the desired responses are reproduced in Table 3. The questions and statements are arranged in order to elicit a natural flow of

TABLE 2

LIST OF "WATS" STIMULUS WORDS AND THE SOUNDS TESTED

1. <u>H</u> OR <u>S</u> E	15. RED <u>E</u>
2. <u>T</u> REE	16. <u>Y</u> EL <u>L</u> OW
3. <u>S</u> K <u>Y</u>	17. <u>F</u> I <u>V</u> E
4. <u>C</u> LOU <u>D</u> S	18. <u>S</u> E <u>V</u> EN
5. <u>C</u> OW	19. FEATH <u>E</u> R
6. <u>S</u> HEEP <u>E</u>	20. <u>B</u> OY
7. <u>D</u> UCK <u>E</u>	21. <u>N</u> O <u>S</u> E
8. <u>R</u> AB <u>B</u> IT	22. <u>T</u> EETH <u>E</u>
9. <u>C</u> HICK <u>E</u> N	23. FINGER
10. <u>P</u> IG	24. <u>T</u> HUMB
11. <u>B</u> ROWN	25. <u>L</u> EG
12. <u>G</u> REEN	26. SHIR <u>T</u>
13. <u>B</u> LACK	27. <u>Z</u> IP <u>P</u> ER
14. <u>W</u> HIT <u>E</u>	28. <u>G</u> ATE

TABLE 3

LIST OF "WATS" STIMULUS QUESTIONS AND STATEMENTS
AND DESIRED RESPONSES

-
1. This is a picture of a little boy and his farm animals.
As I point to the picture I want you to tell me its name.
(a) horse (f) sheep
(b) tree (g) duck
(c) sky (h) rabbit
(d) clouds (i) chicken
(e) cow (j) pig
 2. Now, as I point to the picture tell me what color it is.
(a) brown (d) white
(b) green (e) red
(c) black (f) yellow
 3. Count these baby chicks. (point to group of five chicks)
(a) five
 4. How many do you see altogether? (point to all seven chicks)
(a) seven
 5. What is this? (point to feather)
(a) feather
 6. What is he called? (point to little farm boy)
(a) boy
 7. As I point to the boy I want you to name what you see.
(a) nose
(b) teeth
(c) finger
(d) thumb
 8. What is the horse standing on? (point to horse's leg)
(a) leg
 9. What is he wearing? (point to little boy's shirt)
(a) shirt
 10. What does he fasten his pants with? (point to zipper on suit)
(a) zipper
 11. What does the little boy open to let the animals out of the pasture? (point to gate)
(a) gate
-

responses and to provide a definite pattern for the examiner to follow in administering the WATS. Once the examiner becomes proficient in following this pattern, he will no longer need to refer to the printed stimulus questions and statements.

In order to simplify the recording process, the Record Chart was constructed. It enables the examiner to record the incorrect responses of 15 children easily and rapidly. As seen in Table 4, the Chart contains information as to name, age, sex, grade, date, errors, whether further testing is indicated, and any additional comments. The Error column is divided into three subcolumns representing the three word positions. If a child made an error in producing the /r/ sound in the initial position, it would be necessary to record only the /r/ in the "I" subcolumn. This would indicate that the initial /r/ sound was produced incorrectly. The WATS is a screening test, not a diagnostic test; therefore, the type of error is not recorded.

This particular chart-type of record form was designed on the basis of the examiner's experience in dealing with separate test forms for individual children. Individual forms are time consuming to score, cumbersome, and unnecessary for the purposes of a screening survey.

The WATS score is determined by adding all the incorrect responses recorded on the Record Chart and subtracting this number from 42. The resulting number is then applied to the

TABLE 4

WATTS ARTICULATION TEST FOR SCREENING
RECORD CHART

[illegible]

WATS cut-off scores which were determined from previous research (Templin, 1957) concerned with the development of articulation of speech sounds in words. The WATS cut-off scores can be seen in Table 5. These scores were established for the purpose of identifying those children who require diagnostic evaluation. If further testing is necessary, a mark is placed in the Recheck column on the Record Chart.

TABLE 5
CUT-OFF SCORES FOR THE WATTS ARTICULATION
TEST FOR SCREENING

C.A.	3	3 1/2	4	4 1/2	5	6	7	8
Cut-off Scores	16	17	32	35	35	38	41	41

For best results, testing should be conducted in a quiet, well-lighted, properly ventilated room. The examiner is seated to the subject's right in a chair of equal size. The seating arrangement is displayed in Figure 2. The WATS is placed on a table in front of the subject at eye level approximately 10 to 12 inches from his eyes. The Record Chart is placed to the right and slightly behind the canvas board to eliminate any distraction caused by the recording process. The examiner's left hand is then free to point to the desired pictures.

Fig. 2.--Seating arrangement for the Watts Articulation Test for Screening.

Examiner Reliability

Winitz (1969) in discussing examiner reliability has suggested that the highest examiner agreement occurs when judges score a recording of a simulated articulation defect. For the purpose of establishing examiner reliability in the present study, this procedure was employed.

Two female master's candidates in Speech Pathology at Texas Tech University made a recording on a Cassette cartridge tape recorder (RCA Model VS-335). The recording was made in an Industrial Acoustics Company Sound Treated Room (Model 403SP). One master's candidate presented the 19 picture cards of the Revised Templin-Darley Screening Test while the other simulated an articulation defect. Neither of these two individuals participated in the actual testing of the 90 subjects in this study.

The experimenter and another female graduate student in Speech Pathology at Texas Tech University served as examiners. Each examiner listened individually to the tape recorded articulation test in the AIC booth #403SP for three days in succession. Responses were recorded each time on separate Articulation Test Forms. Neither examiner was permitted to view her own previously scored Test Forms or those scored by the other examiner until the results were ready to be tabulated and analyzed.

According to Winitz (1969) evaluating scores in terms of the percentage of agreement of incorrect responses recorded tends to give a more accurate description of the "percentage agreement index" (Winitz, 1969) if a two-way scoring system such as correct-incorrect is used. This procedure and the evaluation of articulation errors in terms of the type of error were employed to determine interjudge reliability. Intrajudge reliability was established by evaluating the three scores obtained by each examiner in terms of the percentage of agreement within each examiner.

The following results were obtained. Interjudge reliability determined on the basis of the number of incorrect responses scored was 100 per cent. The same 18 out of 50 responses were judged as incorrect by both examiners. The percentage of agreement according to the type of incorrect response recorded was 94.4 per cent. Intrajudge reliability for both examiners was 100 per cent in terms of the number of incorrect responses scored and the type of error recorded.

Administration of Tests

The two articulation screening tests, the Templin-Darley Screening Test (1969) and the Watts Articulation Test for Screening, were administered individually to each of the 90 elementary school children. The order of presentation of the two tests was determined by using a table of random numbers.

Each test was administered in separate rooms at each school. The rooms, equipped with a table and two chairs, were quiet, well-lighted, and properly ventilated. This examiner administered only the WATS while the other examiner administered only the Templin-Darley Screening Test. This procedure was followed to eliminate biasing test results.

The length of administration time per individual child was determined for both articulation screening tests. The examiners were provided with stop watches and instructed to record an accurate account of the time involved in testing each child.

Both examiners attempted to obtain spontaneous responses from all 90 subjects. If, however, after a stimulus question or statement had been stated twice and the subject failed to give the desired response, the examiners were instructed to say the test word aloud. The subject was then asked to repeat the word. A record was kept of all those words which had to be cued by the examiners.

Several weeks after the subjects had been tested, this examiner returned to the three schools to ask the subjects which of the two screening tests they preferred. The tests were shown to the children again and a count was taken of the number who preferred the WATS and the number who preferred the Templin-Darley Test.

Statistical Analysis

A paired t-test for significant difference between two groups was carried out with respect to the time required to administer each of the two screening tests to all 90 subjects.

The test statistic is $t = \frac{\bar{X}_D}{S_D / \sqrt{N}}$ (Dixon and Massey, 1969).

The 90 children were subdivided according to race, sex, school, and grade level. A t-test employing the above test statistic was computed to analyze the differences between the administration times of the two screening tests for each subgroup.

A chi square was performed on the five most unfamiliar stimulus words and on the total number of unfamiliar stimulus words from each test to determine if there were significant differences between the two tests in terms of the frequency of cued words. The formula for computing a chi square in a

2 X 2 table is
$$\chi^2 = \frac{N \left(|ad - bc| - \frac{N}{2} \right)^2}{(a + b)(a + c)(b + d)(c + d)}$$

(Guildord, 1965).

In order to estimate the proportion of children who preferred the WATS, a standard normal test of hypotheses on proportions was carried out. The test statistic is

$$Z = \frac{\bar{X} - P_0}{\sqrt{P_0(1 - P_0)/N}}$$
 (Dixon and Massey, 1969).

The relationship between the Templin-Darley scores and the WATS scores was determined by a Pearson product-moment coefficient of correlation. The basic formula is

$$r_{xy} = \frac{\Sigma xy}{N_{oxy}} \text{ (Guilford, 1965).}$$

CHAPTER III

RESULTS AND DISCUSSION

The Watts Articulation Test for Screening was developed as an outgrowth of the examiner's experience in screening school-age children for speech disorders. Certain deficiencies were found in articulation tests; therefore, the development of a new screening instrument appeared warranted.

Various statistical procedures were carried out to determine if the purposes of the WATS had been fulfilled. The results are categorized according to the time required to administer both tests, the familiarity of the test words, the interest of the tests, the identification of children who need diagnostic evaluation, and the value of the record form.

Administration Time

A fundamental criterion of an effective screening test of articulation is that the test be administered in the shortest practical time. The time required to administer the Templin-Darley Test and the time required to administer the WATS were recorded in seconds for each subject in this study. Administration times of each test can be found for individual subjects in the Appendix. To determine which test, the Templin-Darley or the WATS, was administered more quickly to the 90 subjects, the total administration times

required for each test were calculated. As seen in Table 6, the total time necessary to administer Test 1 (Templin-Darley) to the 90 subjects was 530.066 minutes (8 hours, 50 minutes, 2 seconds). For Test 2 (WATS) a total of 198.615 minutes (3 hours, 18 minutes, 36 seconds) was required to screen the same 90 subjects.

TABLE 6

TOTAL ADMINISTRATION TIMES OF THE TEMPLIN-DARLEY
TEST AND THE WATTS ARTICULATION TEST
FOR SCREENING

Test	N	Hours	Minutes	Seconds	Total Minutes
# 1 Templin-Darley Test	90	8	50	2	530.006
# 2 Watts Articulation Test for Screening	90	3	18	36	198.615

A paired t-test (Dixon and Massey, 1969) was performed to determine if there was a significant difference between the administration times of the two tests. As seen in Table 7, the mean time required to administer Test 1 to each of the 90 subjects was 5.889 minutes as compared to the mean of 2.206 minutes for Test 2. The standard deviation for Test 1 was 2.312 minutes and 1.234 minutes for Test 2. The test statistic was 25.57 which is significant at the .01

level of confidence. The WATS was administered to the 90 subjects in significantly less time than the Templin-Darley Test.

TEST 7

RESULTS OF T-TEST COMPUTED ON THE ADMINISTRATION TIMES OF THE TEMPLIN-DARLEY TEST AND THE WATTS ARTICULATION TEST FOR SCREENING

Test	N	Mean	SD	t-Test
# 1 Templin-Darley Test	90	5.889	2.312	25.57*
# 2 Watts Articulation Test for Screening	90	2.206	1.234	

*Significant at the .01 level of confidence.

The 90 subjects were subdivided according to race, sex, school, and grade level. The total administration times of each test were calculated for the subjects in each of the four subgroups and are presented in Table 8. The total administration times of Test 1 for Caucasian, Negroid, and Mexican-American subjects were 299.566 minutes, 46.25 minutes, and 184.25 minutes respectively. Total times of Test 2 were 105.2 minutes, 19.4 minutes, and 74.016 minutes respectively for Caucasian, Negroid, and Mexican-American subjects. For male and female subjects, administration times of Test 1 were

TABLE 8

TOTALS, MEANS, AND STANDARD DEVIATIONS IN MINUTES OF THE ADMINISTRATION
TIMES OF THE TEMPLIN-DARLEY SCREENING TEST AND THE WATTS
ARTICULATION TEST FOR SCREENING FOR EACH SUBGROUP

Subgroup	Test #1 Templin-Darley Screening Test			Test #2 Watts Articulation Test for Screening		
Race	Total	Mean	SD	Total	Mean	SD
Caucasian	299.566	4.9927	1.3946	105.2	1.7533	.4840
Negroid	46.25	6.6071	1.1782	19.4	2.7714	1.3363
Mexican-American	184.25	8.0108	2.9957	74.016	3.2180	1.8222
Sex						
Male	254.583	6.3645	2.8664	96.2	2.405	1.5590
Female	275.483	5.5096	1.6853	102.416	2.0483	.8813
School						
School 1	132.25	4.4083	.8822	50.116	1.6705	.326
School 2	166.816	5.5605	1.6105	54.383	1.8127	.5996
School 3	231	7.7	2.6953	94.116	3.1372	1.6909
Grade						
Grade 1	211.166	7.0388	2.9604	88.633	2.9544	1.8107
Grade 2	172.483	5.7494	1.8168	54.466	1.9488	.4915
Grade 3	146.416	4.8805	1.3798	51.516	1.7172	.5132

254.583 minutes and 275.483 minutes respectively as compared to 96.2 minutes and 102.416 minutes on Test 2. Subjects enrolled in Schools 1, 2, and 3 required total times of 132.25 minutes, 166.816 minutes, and 231 minutes on Test 1 and 50.116 minutes, 54.383 minutes, and 94.116 minutes respectively on Test 2. For subjects in grades 1, 2, and 3, total administration times of Test 1 were 211.166 minutes, 172.483 minutes, and 146.416 minutes respectively as opposed to 88.633 minutes, 58.466 minutes, and 51.516 minutes on Test 2.

As seen in Table 8, mean times of 4.9927 minutes (SD = 1.3946), 6.6071 minutes (SD = 1.1782), and 8.0108 minutes (SD = 2.9957) were achieved by Caucasian, Negroid, and Mexican-American subjects respectively on Test 1 (Templin-Darley). On Test 2 (WATS), mean times of 1.7533 minutes (SD = .4840), 2.7714 minutes (SD = 1.3363), and 3.2180 minutes (SD = 1.8222) were required by Caucasian, Negroid, and Mexican-American subjects respectively. Mean times were 6.3645 minutes (SD = 2.8664) and 5.5096 minutes (SD = 1.6853) respectively for males and females on Test 1 as compared to 2.405 minutes (SD = 1.5590) and 2.0483 minutes (SD = .8813) respectively on Test 2. Subjects enrolled in Schools 1, 2, and 3 achieved mean times of 4.4083 minutes (SD = .8822), 5.5605 minutes (SD = 1.6105), and 7.7 minutes (SD = 2.6953) respectively on Test 1. On Test 2, the means were 1.6705 minutes (SD = .326), 1.8127 minutes (SD = .5996), and 3.1372

minutes (SD = 1.6909) respectively for subjects enrolled in Schools 1, 2, and 3. The subjects in grades 1, 2, and 3 required mean administration times of 7.0388 minutes (SD = 2.9604), 5.7494 minutes (SD = 1.8168), and 4.8805 minutes (SD = 1.3798) respectively on Test 1 as compared to 2.9544 minutes (SD = 1.8107), 1.9488 minutes (SD = .4915), and 1.7172 minutes (SD = .5132) on Test 2.

A paired t-test (Dixon and Massey, 1969) was computed to determine if there were significant differences between the mean administration times of the two tests for each subgroup. The results of this evaluation, as seen in Table 9 indicated that the t statistic computed for Caucasian subjects was 22.90 which is greater than $t_{.95}(59) = 1.671$. For Negroid subjects, the test statistic was equal to 11.05 which is greater than $t_{.95}(6) = 1.943$. The statistic 15.25 computed for Mexican-American subjects was greater than $t_{.95}(22) = 1.717$. A t statistic of 15.46, which is greater than $t_{.95}(39) = 1.684$, was obtained for males while a t of 22.44 obtained for females is greater than $t_{.95}(49) = 1.678$. For subjects enrolled in Schools 1, 2, and 3, the test statistics were 20.69, 17.05, and 17.56 respectively which are all greater than $t_{.95}(29) = 1.699$. The test statistics for grades 1, 2, and 3 were 17.03, 13.78, and 15.43 respectively which are all greater than $t_{.95}(29) = 1.699$. All the test statistics were greater than their corresponding t values at

TABLE 9

RESULTS OF T-TEST COMPUTED ON THE ADMINISTRATION TIMES
OF THE TEMPLIN-DARLEY SCREENING TEST AND THE
WATTS ARTICULATION TEST FOR SCREENING
FOR EACH SUBGROUP

Subgroup	t-Value	Test Statistic
RACE		
Caucasian	$t_{.95}(59) = 1.671$	22.90*
Negroid	$t_{.95}(6) = 1.943$	11.05*
Mexican-American	$t_{.95}(22) = 1.717$	15.25*
SEX		
Male	$t_{.95}(39) = 1.684$	15.46*
Female	$t_{.95}(49) = 1.678$	22.44*
SCHOOL		
School 1	$t_{.95}(29) = 1.699$	20.69*
School 2	$t_{.95}(29) = 1.699$	17.05*
School 3	$t_{.95}(29) = 1.699$	17.56*
GRADE		
Grade 1	$t_{.95}(29) = 1.699$	17.03*
Grade 2	$t_{.95}(29) = 1.699$	13.78*
Grade 3	$t_{.95}(29) = 1.699$	15.43*

*Significant at the .05 level of confidence.

the .05 level of confidence; therefore, the WATS was administered to the children in each subgroup in significantly less time than the Templin-Darley Test.

The results revealed identical patterns on Test 1 and Test 2 in terms of the average time required to administer each test to the children in each subgroup. On both the Templin-Darley Test and the WATS, Caucasian subjects achieved lower mean times than Negroid subjects who, in turn, exhibited lower average times than Mexican-American subjects. Males achieved higher mean times than females. Subjects from the upper socioeconomic level school displayed lower average times than subjects from the middle socioeconomic level school who, in turn, required lower mean times than subjects from the lower socioeconomic level school. First grade subjects displayed higher mean times than second grade subjects who, in turn, achieved higher mean administration times than third graders.

When the total time necessary to screen the 90 subjects in this study with the WATS was subtracted from the total time required to screen the same 90 subjects with the Templin-Darley Test, a difference of 5 hours, 31 minutes, and 26 seconds resulted. The public school speech clinician can save valuable time in the speech screening survey by employing the WATS, rather than the Templin-Darley Test, to screen school-age children for articulation disorders.

Familiarity of Test Words

Previously constructed articulation tests have employed stimulus words which are unfamiliar to school-age children. In an effort to provide test words that children could readily identify, the Watts Articulation Test for Screening (WATS) was designed around a picture of a farm scene.

In the present study, certain words had to be cued by both examiners during the administration of the tests. These words were considered unfamiliar to the subjects on the basis that the examiners were required to provide the test word to obtain the desired speech response. The data, seen in Table 10, is divided into three sections according to each elementary school that participated in this investigation. The total number of unfamiliar stimulus words and the five most unfamiliar stimulus words from each test are presented for the subjects from each of the three schools.

The Templin-Darley Test words were provided by the examiner a total of 52 times, 44 times, and 286 times at Schools 1, 2, and 3 respectively. The total number of times Templin-Darley Test words were cued was 382. WATS test words were provided 3 times, 0 times, and 43 times respectively at Schools 1, 2, and 3. The WATS words were cued a total of 46 times.

As indicated in Table 10, the five most unfamiliar Templin-Darley Test words were provided by the examiner 45

TABLE 10

TOTAL NUMBER OF UNFAMILIAR STIMULUS WORDS AND LIST
OF THE FIVE MOST UNFAMILIAR STIMULUS WORDS
FROM THE TEMPLIN-DARLEY SCREENING TEST
AND THE WATTS ARTICULATION TEST
FOR SCREENING

School 1	Screening Test	Five Most Unfamiliar Stimulus Words					Total
	Templin-Darley Test	Shredded Wheat	Smooth	Fire Engine	Music	Sprinkler	45
		21	4	8	8	4	
	WATS	Cloud	Gate	Feather	Tree	Sky	3
		0	3	0	0	0	
School 2	Screening Test	Five Most Unfamiliar Stimulus Words					Total
	Templin-Darley Test	Shredded Wheat	Smooth	Fire Engine	Music	Sprinkler	34
		19	5	4	4	2	
	WATS	Cloud	Gate	Feather	Tree	Sky	0
		0	0	0	0	0	
School 3	Screening Test	Five Most Unfamiliar Stimulus Words					Total
	Templin-Darley Test	Shredded Wheat	Smooth	Fire Engine	Music	Sprinkler	103
		28	23	18	14	20	
	WATS	Cloud	Gate	Feather	Tree	Sky	28
		9	6	6	4	3	

Total # of

Cued Words: Templin-Darley Screening Test = 382
 Watts Articulation Test for Screening = 46

times at School 1, 34 times at School 2, and 103 times at School 3. Only one of the five most unfamiliar WATS stimulus words was cued 3 times by the examiner at School 1. The examiner was not required to provide any WATS test words at School 2. At School 3, the examiner was required to say the five most unfamiliar WATS test words a total of 28 times. Both examiners were required to provide the greatest number of cues at Schools 3, 1, and 2 respectively.

To determine if there was a significant difference between the two tests in terms of the frequency of cued words, chi squares (Guilford, 1965) were computed on the total number of unfamiliar stimulus words and on the five most unfamiliar stimulus words from each test. As seen in Table 11, the results of the chi square performed on the total number of unfamiliar words from both tests ($\chi^2 = 124.01 > 6.63$) revealed a significant difference between the tests at the .01 level of confidence. The total number of unfamiliar Templin-Darley Test words were provided by the examiner a significantly greater number of times than the total number of unfamiliar WATS test words. As seen in Table 12, the results of the test performed on the five most unfamiliar stimulus words from each test ($\chi^2 = 138.38 > 6.63$) showed a significant difference between the Templin-Darley Test and the WATS at the .01 level of confidence. The five most unfamiliar Templin-Darley Test words had to be cued by the examiner

significantly more times than the five most unfamiliar WATS test words.

TABLE 11

RESULTS OF THE CHI SQUARE COMPUTED ON THE TOTAL NUMBER OF FAMILIAR AND UNFAMILIAR TEMPLIN-DARLEY TEST AND "WATS" STIMULUS WORDS ARRANGED IN A 2 X 2 CONTINGENCY TABLE

	Templin-Darley Screening Test	Watts Articulation Test for Screening	Marginal Totals
Unfamiliar Test Words	382 a	46 b	428 a + b
Familiar Test Words	4118 c	2474 d	6592 c + d
Total Number of Test Words	4500 a + c	2520 b + d	7020 N

$$\chi^2 = 124.01 > 6.63^*$$

*Significant at the .01 level of confidence.

Interest of Tests

Several weeks after the two articulation tests had been administered to the subjects, both tests were shown to the subjects again. To determine if the WATS was, indeed, interesting, 88 of the original 90 subjects were asked which of the two tests they liked better. Two subjects had moved; therefore, data for these 2 children could not be collected. A count was taken of the number of children who preferred the Templin-Darley Test and the number of children who preferred

the WATS. The results can be seen in Table 13. Of the 88 subjects, 14 preferred the Templin-Darley Test and 74 preferred the WATS.

TABLE 12

RESULTS OF THE CHI SQUARE COMPUTED ON THE FIVE MOST
UNFAMILIAR TEMPLIN-DARLEY TEST AND "WATS"
STIMULUS WORDS ARRANGED IN A 2 X 2
CONTINGENCY TABLE

	Templin-Darley Screening Test		Watts Articulation Test for Screening		Marginal Totals
Five Most Unfamiliar Test Words	182	a	31	b	213 a + b
Familiar Test Words	268	c	419	d	687 c + d
Total Number of Test Words	450	a + c	450	c + d	900 N

$$\chi^2 = 138.38 > 6.63^*$$

*Significant at the .01 level of confidence.

A standard normal test of hypotheses on proportions (Dixon and Massey, 1969) was performed. The hypothesis under test was that the proportion of children who preferred the WATS was greater than or equal to 90 per cent. The results of the test indicated that the Z value -1.87 is greater than -2.326; therefore, the hypothesis was accepted.

The fact that at least 90 per cent of the children tested preferred the WATS to the Templin-Darley Test indicated

that the WATS was more interesting to the subjects. Several children, selected at random, were asked why they liked the WATS better. Some of the typical responses were, "Because I like farm animals" and "Because I like to visit the farm."

TABLE 13

NUMBER OF SUBJECTS WHO PREFERRED THE TEMPLIN-DARLEY TEST AND
THE NUMBER WHO PREFERRED THE WATTS ARTICULATION TEST
FOR SCREENING

School	Screening Test	
	Templin-Darley Screening Test	Watts Articulation Test for Screening
SCHOOL 1		
Grade 1	3	7
Grade 2	4	6
Grade 3	<u>0</u>	<u>10</u>
	7	23
SCHOOL 2		
Grade 1	3	7
Grade 2	1	9
Grade 3	<u>0</u>	<u>10</u>
	4	26
SCHOOL 3		
Grade 1	2	7
Grade 2	0	9
Grade 3	<u>1</u>	<u>9</u>
	3	25
Total	14	74

Identification of Children

If a screening test is administered rapidly, contains familiar stimulus words, and is interesting to school-age children and, yet, does not accurately identify children who need diagnostic evaluation, the test is inadequate. The WATS was compared to the Templin-Darley Test to determine if it was as accurate in selecting children to receive more detailed examination.

The scores obtained by the 90 subjects on the Templin-Darley Test were applied to the Templin-Darley cut-off scores previously mentioned. The scores obtained by the same subjects on the WATS were applied to the previously cited WATS cut-offs. The children selected for diagnostic testing can be seen in Table 14.

Ten subjects identified by the Templin-Darley Test were also identified by the WATS. Of these 10 children, 6 were Caucasian, 0 were Negroid, and 4 were Mexican-American. The subjects were 7 males and 3 females. There were 4 children selected from School 1, 2 subjects selected from School 2, and 4 subjects selected from School 3. Of the 10 subjects, 4 were in the first grade, 5 in the second grade, and 1 in the third grade. Of the subjects selected, 2 were 6 years old, 3 were 7 years old, 2 were 8 years old, 1 was 9 years old, and 2 were 10 years old.

As seen in Table 14, the WATS identified 10 additional children for diagnostic testing which were not identified

TABLE 14

SUBJECTS IDENTIFIED BY THE TEMPLIN-DARLEY SCREENING TEST
AND THE WATTS ARTICULATION TEST FOR SCREENING
FOR DIAGNOSTIC EVALUATION

Subject	Race	Sex	School	Grade	Age	Test Scores	
						Templin-Darley	WATS
03	C	M	1	1	6	33	34
06	C	M	1	1	7	36	37
12	C	M	1	2	7	38	35
29	C	F	1	3	8	35	37
38	C	F	2	1	6	06	18
48	C	F	2	2	7	38	38
63	M-A	M	3	1	8	36	33
71	M-A	M	3	2	9	42	39
72	M-A	M	3	2	10	41	40
73	M-A	M	3	2	10	42	35
39*	C	F	2	1	7	48	40
61*	N	M	3	1	9	44	39
62*	M-A	M	3	1	6	40	33
64*	M-A	M	3	1	7	43	36
67*	N	F	3	1	8	44	36
68*	M-A	F	3	1	6	42	37
74*	N	F	3	2	8	48	40
77*	M-A	F	3	2	7	48	40
80*	M-A	F	3	2	9	45	40
85*	M-A	M	3	3	10	46	39

*Indicates those subjects identified by the WATS who were not identified by the Templin-Darley Test.

by the Templin-Darley Test. Of these subjects, 1 was Caucasian, 3 were Negroid, and 6 were Mexican-American. There were 4 male and 6 female subjects. None of the children were selected from School 1, 1 child was selected from School 2, and 9 children were selected from School 3. Of the 10 subjects, there were 6 in the first grade, 3 in the second grade, and 1 in the third grade. Of the 10 subjects identified, 2 were 6 years old, 3 were 7 years old, 2 were 8 years old, 2 were 9 years old, and 1 was 10 years old.

The WATS identified the same 10 children for diagnostic evaluation that the Templin-Darley Test identified. The fact that 10 additional subjects were selected by the WATS suggests that there is a differentiation between the cut-off scores of the two tests. A possible explanation for this discrepancy is that the WATS tests several early developmental sounds that are not tested in the Templin-Darley Test; therefore, the WATS cut-off scores are higher in most instances. This makes it more difficult for a child with a speech defect to escape detection by the WATS.

The Pearson product-moment coefficient of correlation (Guilford, 1965) was computed to measure the correlation or relationship between the Templin-Darley Test scores and the WATS scores. The results of this test are presented in Table 15.

TABLE 15

DESCRIPTIVE STATISTICS OF THE PEARSON PRODUCT-MOMENT
COEFFICIENT OF CORRELATION COMPUTED ON THE
TEMPLIN-DARLEY AND "WATS" TEST SCORES

Variable	Mean	SD	SD Error	r
Templin-Darley Screening Test	47.044	5.8615	.621321	.939
WATS	40.488	3.2472	.344211	

The mean for the Templin-Darley Test scores was 47.044 while the mean for the WATS scores was 40.488. The standard deviation was 5.8615 for the Templin-Darley scores and 3.2472 for the WATS scores. The standard error for the Templin-Darley was .621321 as compared to .344211 for the WATS scores. The coefficient of correlation between the Templin-Darley Test scores and the WATS scores was .94. This value indicated that there was a strong relationship between the two test scores achieved by the 90 subjects in this study. Subjects who were identified for diagnostic examination tended to make low scores on both screening tests and subjects who were not identified for testing tended to make high scores on both tests.

Value of the Record Chart

The last purpose in developing a new screening test of articulation was to construct a simple form for recording

responses. During the development of the WATS, several types of record forms were employed on an experimental basis. In a pilot study conducted by this investigator involving 101 Caucasian, Negroid, and Indian subjects ranging in age from 6 to 10 years, various record forms were utilized. The Record Chart evolved as the least complicated and most efficient form for recording responses to the WATS.

No statistical measures were computed to determine if the WATS Record Chart was more efficient than the Templin-Darley Articulation Test Form. The Chart, however, was found to be time saving, easy to score, and informative. The responses of 15 children were recorded on one Chart, thus, eliminating the numerous individual record forms. Recording only the sounds that were defective, rather than using a complex system of symbols, added to the effectiveness of the Chart. The value of the Chart for other speech clinicians remains to be determined.

CHAPTER IV

SUMMARY AND CONCLUSIONS

The majority of speech defective children in the school population display disorders of articulation. It was recognized that the public school speech clinician has attempted to identify these children by employing articulation tests which are inadequate in terms of (1) stimulus materials, (2) administration time, (3) familiarity of stimulus words, (4) interest for school-age children, and (5) method of recording responses.

Purpose of Study

The purpose of this investigation was to develop and evaluate a screening test of articulation which would eliminate the deficiencies found in previously cited articulation tests. The Watts Articulation Test for Screening was specifically designed to facilitate the speech screening survey by: (1) identifying children who need diagnostic evaluation, (2) utilizing one stimulus picture thereby eliminating unnecessary stimulus materials, (3) saving time, (4) providing a simple method for recording responses, (5) providing stimulus words which are familiar to young school-age children, and (6) providing school-age children with an interesting articulation screening test.

Procedures

The Templin-Darley Screening Test and the WATS were administered individually to 90 elementary school children, 30 in the first grade, 30 in the second grade, and 30 in the third grade. Of the 90 subjects, 60 were Caucasian, 7 were Negroid, and 23 were Mexican-American. There were 40 males and 50 females ranging in age from 6 years, 6 months to 10 years, 4 months. No subject displayed any gross visual or auditory impairment. These children were selected at random from three elementary schools. School 1 appeared to represent the upper socioeconomic level, School 2 appeared to represent the middle socioeconomic level, and School 3 appeared to represent the lower socioeconomic level of Lubbock, Texas.

Two female graduate students in Speech Pathology at Texas Tech University served as examiners. Interjudge and intrajudge reliability were determined by having the examiners score a tape recorded articulation test in which defective articulation responses to the Templin-Darley Screening Test were simulated by a third graduate student. Interjudge reliability determined on the basis of the incorrect responses scored was 100 per cent. The same 18 out of 50 responses were judged as incorrect by both examiners. The percentage of agreement according to the type of error recorded by the examiners was 94.4 per cent. Intrajudge

reliability for both examiners was 100 per cent in terms of the incorrect responses scored and the type of error recorded.

The Templin-Darley Test served as the standard test in this study against which the WATS was compared. The Templin-Darley is organized around 19 cards containing 50 pictures. These pictures represent fifty words in which 50 sound elements are tested. The WATS is designed around a 16 inch by 20 inch brightly colored canvas board depicting a farm scene. There are twenty-eight stimulus words which test 42 sound elements.

The order of presentation of the two tests was determined by using a table of random numbers. The tests were administered in separate rooms provided at each school. One examiner administered only the Templin-Darley Test while the other examiner administered only the WATS.

The examiners were instructed to cue a stimulus word if the subjects failed to give the desired response after a stimulus question or statement had been stated twice. A list of cued words from both tests was compiled to be used in the analysis of data. The examiners used stop watches to measure the time required to administer each test to the individual subjects.

Several weeks after the articulation tests had been administered to the 90 subjects, both tests were shown to

the subjects again. The children were asked which of the two tests they liked better. A count was taken to determine the number of subjects who preferred the Templin-Darley Test and the number of subjects who preferred the WATS.

Results

A paired t-test (Dixon and Massey, 1969) was computed to determine if there was a significant difference between the total administration times of the Templin-Darley Test and the WATS. The results indicated a mean time of 5.889 minutes was required to administer the Templin-Darley Test to each of the 90 subjects as compared to a mean of 2.206 minutes to administer the WATS. The test statistic 25.57 was significant at the .01 level of confidence. The WATS was administered to the 90 subjects in significantly less time than the Templin-Darley Test.

The 90 subjects were subdivided according to race, sex, school, and grade level. A paired t-test (Dixon and Massey, 1969) was computed to determine if there were significant differences between the average administration times of the two tests for the subjects in each subgroup. The results indicated that all the test statistics were greater than their corresponding t values at the .05 level of confidence. The WATS was administered to the children in each subgroup in significantly less time than the Templin-Darley Test.

The children in each subgroup displayed identical patterns on the two screening tests. Caucasians exhibited lower mean times than Negroes who, in turn, displayed lower mean times than Mexican-Americans. Males required higher mean times than females. Subjects enrolled in the upper socioeconomic level school achieved lower mean times than subjects enrolled in the middle socioeconomic level school who, in turn, required lower mean times than the children enrolled in the lower socioeconomic level school. First graders achieved higher mean administration times than second graders who, in turn, required higher mean times than third graders.

The difference between the total time necessary to screen the 90 subjects with the Templin-Darley Test and the time required to screen the same subjects with the WATS was 5 hours, 31 minutes, and 26 seconds. The public school speech clinician can save time by screening school-age children for speech disorders with the WATS, rather than with the Templin-Darley Test.

Chi square tests (Guilford, 1965) were computed on the total number of unfamiliar stimulus words from each test and on the five most unfamiliar stimulus words from each test. The chi square test was performed to determine if there were significant differences between the tests in terms of the frequency of cued words. The results of the chi square on

the total number of unfamiliar words from both tests ($\chi^2 = 124.01 > 6.63$) revealed a significant difference at the .01 level of confidence. The total number of unfamiliar Templin-Darley stimulus words were provided by the examiner a significantly greater number of times than the total number of unfamiliar WATS test words. The results of the test on the five most unfamiliar test words from each test ($\chi^2 = 138.38 > 6.63$) indicated a significant difference at the .01 level of confidence. The five most unfamiliar Templin-Darley test words were cued significantly more times than the five most unfamiliar WATS test words.

A standard normal test of hypotheses on proportions (Dixon and Massey, 1969) revealed that at least 90 per cent of the children tested preferred the WATS to the Templin-Darley Screening Test. These results suggest that the WATS was more interesting to the subjects tested than the Templin-Darley Test.

The scores obtained by the 90 subjects on the Templin-Darley Test and the scores obtained on the WATS were applied to the respective cut-off scores of each test. The Templin-Darley Test identified 10 children for diagnostic evaluation. The same 10 subjects were also selected by the WATS. In addition, the WATS identified 10 other children for diagnostic testing.

The fact that the WATS selected more subjects for detailed examination than the Templin-Darley Test suggests that there is a difference between the cut-off scores of the two tests. Several early developmental sounds tested in the WATS are not tested in the Templin-Darley Test; therefore, the WATS cut-off scores are higher in most instances. A speech defective child would, therefore, have more difficulty escaping detection by the WATS.

The Pearson product-moment coefficient of correlation (Guilford, 1965) was computed to measure the relationship between the Templin-Darley Test scores and the WATS scores. The results revealed a correlation coefficient of .94, which indicates a strong relationship between the scores. The correlation indicates that the manner of selecting subjects is similar for both tests. Subjects who are identified for detailed examination will tend to make low scores on both screening tests, while subjects who are not identified for testing will tend to make high scores on both tests.

No statistical measures were computed to determine if the WATS Record Chart was more efficient than the Templin-Darley Articulation Test Form. The Chart was found to be time saving, easy to score, and informative on the basis of experimental use alone. The value of the Chart, however, remains to be determined by each individual clinician.

Conclusions

Based on the results of the analysis of data for the present study, the following conclusions were made:

1. The WATS can be administered in significantly less time than the Templin-Darley Screening Test and, thereby, save time in the speech screening survey.
2. The WATS stimulus words were significantly more familiar to the subjects tested than the Templin-Darley Test words.
3. At least 90 per cent of the children tested liked the WATS better than the Templin-Darley Test.
4. The WATS employs earlier developmental sounds than the Templin-Darley Test; thus, the WATS cut-off scores will tend to identify more children for diagnostic evaluation.
5. The WATS scores correlated highly with the Templin-Darley Test scores; therefore, both tests select children in a similar manner for diagnostic testing.
6. The WATS Record Chart appeared time saving, easy to score, and informative.

Implications

Some of the goals of conducting research in the field of Speech Pathology are to identify, remediate, and prevent communication disorders. The present study was undertaken

for the purpose of developing and evaluating a new test which would aid the public school speech clinician in identifying children with articulation defects.

The results of this experiment indicated that time can be saved and children can be accurately identified for diagnostic evaluation when a screening test employing a single stimulus picture is used. For further research, investigating various themes for a picture test which would eliminate cultural bias might prove invaluable for assessing articulatory skills of groups other than middle class whites. Another suggestion for research is to construct a test that does not require the examiner to point to the desired stimulus picture.

LIST OF REFERENCES

- Anderson, V. A., Improving the Child's Speech. New York: Oxford University Press (1953).
- Barker, Janet O., A numerical measure of articulation. J. Speech Hearing Dis., 25, 79-88 (1960).
- Barker, Janet O., and England, G., A numerical measure of articulation: further developments. J. Speech Hearing Dis., 27, 23-27 (1962).
- Carter, Eunice T., and Buck, M., Prognostic testing for functional articulation disorders among children in the first grade. J. Speech Hearing Dis., 23, 124-133 (1958).
- Darley, F. L., Diagnosis and Appraisal of Communicative Disorders. Englewood Cliffs, New Jersey: Prentice-Hall (1964).
- Dixon, W. J., and Massey, F. J., Jr., Introduction to Statistical Analysis. (3rd ed.). New York: McGraw-Hill (1969).
- Dunn, L. M. (Ed.), Exceptional Children in the Schools. New York: Holt, Rinehart, and Winston (1963).
- Fristoe, Macalyne, and Goldman, R., Comparison of traditional and condensed articulation tests examining the same number of sounds. J. Speech Hearing Res., 11, 583-589 (1968).
- Fudala, Janet B., Arizona Articulation Proficiency Scale: Revised Manual. Los Angeles, California: Western Psychological Services (1970).
- Goldman, R., and Fristoe, Macalyne, The development of the filmstrip articulation test. J. Speech Hearing Dis., 32, 356-362 (1967).
- Goldman, R., and Fristoe, Macalyne, Goldman-Fristoe Test of Articulation. Circle Pines, Minnesota: American Guidance Service (1969).
- Goldman, R., Fristoe, Macalyne, and Thomas, Linda, Development and evaluation of a filmstrip articulation test. (ABSTRACT) Asha, 7, 387 (1965).

- Guilford, J. P., Fundamental Statistics in Psychology and Education. (4th ed.). New York: McGraw-Hill (1965).
- Hejna, R. F., Developmental Articulation Test. Ann Arbor: Speech Materials (1959).
- Irwin, Ruth B., and Musselman, Barbara, A compact picture articulation test. J. Speech Hearing Dis., 27, 36-39 (1962).
- Johnson, W., Darley, F. L., and Spriestersbach, D. C., Diagnostic Methods in Speech Pathology. New York: Harper and Row (1963).
- Kirk, S. A., Educating Exceptional Children. Boston: Houghton Mifflin (1962).
- McDonald, E. T., Articulation Testing and Treatment: A Sensory Motor Approach. Pittsburgh: Stanwix House (1964).
- Riley, G. D., Construction and standardization of a quick screening test. (ABSTRACT) Asha, 7, 387 (1965).
- Rodgers, W. C., The development of an articulation and language screening procedure. Unpublished master's thesis. California State College (1968).
- Templin, Mildred C., Spontaneous versus imitated verbalization in testing articulation in preschool children. J. Speech Hearing Dis., 12, 293-300 (1947a).
- Templin, Mildred C., A non-diagnostic articulation test. J. Speech Hearing Dis., 12, 392-396 (1947b).
- Templin, Mildred C., Norms on a screening test of articulation for ages three through eight. J. Speech Hearing Dis., 18, 323-331 (1953).
- Templin, Mildred C., Certain Language Skills In Children. Inst. Child Welfare Monogr. Series, No. 26. Minneapolis: University of Minnesota Press (1957).
- Templin, Mildred C., and Darley, F. L., The Templin-Darley Tests of Articulation. Iowa City: University of Iowa (1960).
- Templin, Mildred C., and Darley, F. L., The Templin-Darley Tests of Articulation. Iowa City: University of Iowa (1969).

- Thorndike, E. L., and Lorge, I., The Teacher's Word Book of 30,000 Words. New York: Teacher's College, Columbia Univ. (1944).
- Van Riper, C. G., Speech Correction: Principles and Methods (4th ed.). Englewood Cliffs, New Jersey: Prentice-Hall (1963).
- Van Riper, C. G., and Erickson, R., A predictive screening test of articulation. J. Speech Hearing Dis., 34, 214-219 (1969).
- Van Riper, C. G., and Irwin, J., Voice and Articulation. Englewood Cliffs, New Jersey: Prentice-Hall (1958).
- Winitz, H., Articulatory Acquisition and Behavior. New York: Appleton-Century-Crofts (1969).
- Wood, K. S., Terminology and Nomenclature. In L. Travis (Ed.), Handbook of Speech Pathology. New York: Appleton-Century-Crofts (1957).

APPENDIX

SUBJECT CHARACTERISTICS, TEST ADMINISTRATION TIMES,
NUMBER OF CUED WORDS, TEST SCORES,
AND TEST PREFERENCE

Subject	Age	Race	Sex	School	Grade	Time in Seconds		Number of Cued Words		Test Score		Test Preferred	
						Templin-Darley	WATS	Templin-Darley	WATS	Templin-Darley	WATS	Templin-Darley	WATS
01	6	C	M	1	1	325	129	03	00	50	42		✓
02	7	C	M	1	1	285	100	04	00	50	42		✓
03	6	C	M	1	1	245	103	01	00	33	34		✓
04	6	C	M	1	1	275	103	04	00	49	41	✓	
05	6	C	M	1	1	240	108	02	00	50	42		✓
06	7	C	M	1	1	395	169	03	01	36	37	✓	
07	6	M-A	F	1	1	345	090	02	00	49	42	✓	
08	7	C	F	1	1	290	091	03	00	48	41		✓
09	6	C	F	1	1	300	115	04	00	49	42		✓
10	7	C	F	1	1	280	082	03	00	50	42		✓
11	8	C	M	1	2	195	088	01	00	50	42		✓
12	7	C	M	1	2	300	095	03	00	38	35	✓	
13	8	C	M	1	2	330	116	02	00	50	42	✓	
14	7	C	M	1	2	270	122	01	00	49	41		✓
15	8	C	F	1	2	255	098	01	00	50	42		77 ✓

Subject	Age	Race	Sex	School	Grade	Time in Seconds		Number of Cued Words		Test Score		Test Preferred	
						Templin-Darley	WATS	Templin-Darley	WATS	Templin-Darley	WATS	Templin-Darley	WATS
16	9	C	F	1	2	210	092	01	00	50	42		✓
17	7	N	F	1	2	340	105	04	01	49	42	✓	
18	8	C	F	1	2	220	088	00	00	50	42		✓
19	8	C	F	1	2	270	127	01	00	50	42	✓	
20	8	C	F	1	2	250	112	01	00	49	42		✓
21	8	C	M	1	3	180	079	00	00	50	42		✓
22	8	C	M	1	3	245	087	01	00	50	42		✓
23	8	M-A	M	1	3	185	081	00	00	50	41		✓
24	8	C	F	1	3	235	096	01	00	50	42		✓
25	8	C	F	1	3	330	087	04	00	50	42		✓
26	8	C	F	1	3	265	109	02	01	50	42		✓
27	8	C	F	1	3	185	084	00	00	50	42		✓
28	9	C	F	1	3	235	073	00	00	50	42		✓
29	8	C	F	1	3	235	099	00	00	35	37		✓
30	9	C	F	1	3	220	079	00	00	50	42		✓

Subject	Age	Race	Sex	School	Grade	Time in Seconds		Number of Cued Words		Test Score		Test Preferred
						Templin-Darley	WATS	Templin-Darley	WATS	Templin-Darley	WATS	
31	7	C	M	2	1	350	119	04	00	50	42	✓
32	6	C	M	2	1	540	202	04	00	48	40	✓
33	7	M-A	M	2	1	300	097	01	00	50	42	✓
34	7	C	M	2	1	410	152	03	00	47	40	✓
35	6	C	M	2	1	320	109	02	00	49	42	✓
36	6	C	F	2	1	360	104	02	00	49	42	✓
37	6	C	F	2	1	300	090	01	00	46	40	✓
38	6	C	F	2	1	590	247	01	00	06	18	✓
39	7	C	F	2	1	285	098	00	00	48	40	✓
40	6	C	F	2	1	325	085	01	00	49	42	✓
41	7	C	M	2	2	475	107	04	00	50	42	✓
42	8	C	M	2	2	390	107	01	00	50	42	✓
43	9	C	M	2	2	320	097	02	00	49	42	✓
44	7	C	F	2	2	410	097	02	00	50	41	✓
45	8	C	F	2	2	255	094	01	00	50	42	✓

Subject	Age	Race	Sex	School	Grade	Time in Seconds		Number of Cued Words		Test Score		Test Preferred
						Templin-Darley	WATS	Templin-Darley	WATS	Templin-Darley	WATS	
46	7	C	F	2	2	220	102	01	00	50	42	✓
47	8	C	F	2	2	315	096	00	00	50	42	✓
48	7	C	F	2	2	440	133	03	00	38	38	✓
49	7	C	F	2	2	307	111	01	00	49	42	✓
50	8	C	F	2	2	272	106	02	00	49	42	✓
51	10	C	M	2	3	400	117	03	00	49	42	✓
52	8	C	M	2	3	225	081	00	00	50	42	✓
53	9	C	M	2	3	235	083	01	00	50	42	✓
54	9	C	M	2	3	250	109	00	00	49	42	✓
55	8	C	M	2	3	220	083	01	00	50	42	✓
56	9	C	M	2	3	305	078	01	00	50	42	✓
57	8	C	M	2	3	455	104	02	00	47	42	✓
58	9	C	F	2	3	265	082	00	00	50	42	✓
59	9	C	F	2	3	215	101	00	00	50	42	✓
60	8	C	F	2	3	255	072	00	00	50	42	✓
												80

Subject	Age	Race	Sex	School	Grade	Time in Seconds		Number of Cued Words		Test Score		Test Preferred
						Templin-Darley	WATS	Templin-Darley	WATS	Templin-Darley	WATS	
61	9	N	M	3	1	535	317	21	03	44	39	✓
62	6	M-A	N	3	1	900	488	29	09	40	33	✓
63	8	M-A	M	3	1	750	358	25	05	36	33	✓
64	7	M-A	M	3	1	900	433	38	11	43	36	✓
65	6	M-A	F	3	1	495	213	15	00	46	39	✓
66	6	M-A	F	3	1	515	269	12	01	45	41	✓
67	8	N	F	3	1	405	229	09	00	44	36	✓
68	6	M-A	F	3	1	570	295	20	02	42	37	✓
69	7	N	F	3	1	355	141	07	01	49	42	*
70	7	M-A	F	3	1	485	182	09	02	47	41	✓
71	9	M-A	M	3	2	480	154	08	00	42	39	✓
72	10	M-A	M	3	2	595	153	10	00	41	40	✓
73	10	M-A	M	3	2	555	157	09	00	42	35	✓
74	8	N	F	3	2	330	162	06	01	48	40	✓
75	8	C	F	3	2	305	114	03	00	48	41	81 ✓

Subject	Age	Race	Sex	School	Grade	Time in Seconds		Number of Cued Words		Test Score		Test Preferred
						Templin-Darley	WATS	Templin-Darley	WATS	Templin-Darley	WATS	
76	8	C	F	3	2	305	078	02	00	50	42	✓
77	7	M-A	F	3	2	550	218	16	05	48	40	✓
78	8	M-A	F	3	2	315	127	02	00	49	41	✓
79	7	N	F	3	2	380	105	05	00	46	41	✓
80	9	M-A	F	3	2	490	147	06	01	45	40	*
81	9	M-A	M	3	3	340	119	03	00	47	42	✓
82	9	M-A	M	3	3	375	124	04	00	49	42	✓
83	10	M-A	M	3	3	385	192	04	02	46	41	✓
84	10	M-A	M	3	3	375	095	02	00	50	42	✓
85	10	M-A	M	3	3	420	157	05	00	46	39	✓
86	8	M-A	F	3	3	420	196	10	00	50	42	✓
87	9	C	F	3	3	350	112	02	00	50	42	✓
88	8	M-A	F	3	3	310	096	01	00	47	41	✓
89	10	N	F	3	3	430	105	03	00	48	41	✓
90	8	C	F	3	3	240	111	01	00	50	42	82 ✓

*Indicates subject had moved.