H080-630 ERIC REPORT RESUME

ED 010 042 9-02-66 24 (REV)

ASSESSMENT OF THE WRITTEN LANGUAGE OF DEAF STUDENTS.

STUCKLESS, E. ROSS * MARKS, CLAUDE H.

TGZ65134 UNIVERSITY OF PITTSBURGH, SCHOOL OF EDUCATION
CRP-2544
-MAY-66 OEC-5-10-123
EDRS PRICE MF-$0.27 HC-$5.68 142P.

*DEAF CHILDREN, *DEAF RESEARCH, *COMPOSITION (LITERARY), *COMPARATIVE ANALYSIS, *LANGUAGE TEACHERS, WRITING, REMEDIAL TEACHERS, PITTSBURGH, PENNSYLVANIA

THIS INVESTIGATION WAS CONDUCTED TO DESCRIBE THE WRITTEN LANGUAGE OF DEAF STUDENTS BETWEEN 10 AND 18 YEARS OF AGE IN TERMS OF SIX MEASURABLE VARIABLES, AND TO RELATE THESE VARIABLES TO TEACHER JUDGMENTS OF QUALITY OF LANGUAGE. TEN STRATIFIED RANDOM SAMPLES OF COMPOSITIONS BY DEAF STUDENTS WERE SELECTED FROM 14 RESIDENTIAL AND DAY EDUCATIONAL PROGRAMS FOR THE DEAF. COMPOSITIONS WERE WRITTEN UNDER STANDARDIZED CONDITIONS, IN RESPONSE TO THE PRESENTATION OF A FOUR-PICTURE SEQUENCE. THREE MASTER TEACHERS OF LANGUAGE OF THE DEAF SUBJECTIVELY SCORED THE COMPOSITIONS IN EACH, OF THE 10 SAMPLES. THE MEAN CORRELATION OF THE THREE TEACHER-JUDGES SCORES WAS .744. ALL SIX MEASURABLE VARIABLES TENDED TO CORRELATE POSITIVELY WITH AGE, BUT CORRELATIONS WERE SMALL. EQUATIONS WERE DEVELOPED WHICH WERE CONSIDERED TO BE USEFUL INSTRUMENTS FOR EVALUATING THE WRITTEN LANGUAGE OF GROUPS OF DEAF STUDENTS RELATIVE TO THE NATIONAL POPULATION OF DEAF STUDENTS REPRESENTED BY THE SAMPLES. (GD)

ASSESSMENT OF THE WRITTEN LANGUAGE OF DEAF STUDENTS

E. Ross Stuckless, Ph.D.
Claude H. Marks, Ed.D.

Program in Special Education and Rehabilitation
School of Education
University of Pittsburgh

May, 1966

Cooperative Research Project 2544
Contract OE-5-10-123

The research reported herein was supported through the Cooperative Research Program of the Office of Education, U. S. Department of Health, Education, and Welfare.

Acknowledgments

It would be impossible to cite individually every colleague who contributed to this investigation. We wish first to thank the 359 teachers of the deaf who participated directly in the various phases of this project by administering the test to their 2,855 deaf students. Teachers of the deaf compete actively with the clock and the calendar to give their students the language skills so desperately needed. We appreciate their setting time aside for the administration of the test, and preparing records on each of their students for us.

Appreciation is extended to the superintendents, principals, and supervising teachers of the participating schools for making their schools and their students available to us, for altering schedules to accommodate the project, and not least, for sharing insights with us during the course of data collection.

Special thanks are expressed to the three teacher-judges who each carefully read, re-read, and scored 1350 compositions. Every teacher knows the time and decisions involved in such a task. We thank Mr. Philip Schmidt, Mrs. Barbara Wentworth Wilson, and Mrs. Margaret Wood LeDuc for generously participating in this critical phase. The fact that no specific criteria for evaluating the compositions were provided made their task particularly difficult.

We wish to thank also the several consultants who contributed to various aspects of the investigation, among whom were Dr. William Craig, Dr. Joseph Rosenstein, Dr. Jerome Schein, and Dr. Audrey Simmons.

The mechanics of collection and analysis of the compositions used in this investigation were formidable. Miss Judith Fallon, Graduate Student Assistant, was of valuable assistance in these tasks, and, indeed, made numerous suggestions which were incorporated into the design of the investigation. We would be remiss also if we did not mention Miss Toby Saville who, as secretary to the project, committed herself to it as though it were her reponsibility alone.

Finally, sincere appreciation is expressed to our colleagues in the Program in Special Education and Rehabilitation and throughout the School of Education of the University for generously giving their time to consultation on the project, and to the Administration of the University for its supportive policies in accommodating to the demands of this research in terms of facilities, teaching loads, and other necessary adjustments.

E. Ross Stuckless, Ph.D.
Claude H. Marks, Ed.D.

TABLE OF CONTENTS

List of Tables

List of Tables (Cont.)

List of Figures

I. PROBLEM

A. Introduction

The most immediate behavioral consequence of deafness in the young child is the repression of communication skills. Without auditory integrity, the deaf child is unable to assimilate and code sounds as language.

As a consequence, the young deaf child is unable to acquire the auditory-receptive or vocal-expressive language system with the spontaneity of the hearing child. Without this primary language base, the deaf student is severely restricted in the acquisition of secondary communication modes such as reading and written language, and one alternative to auditory language reception, speech reading.

The deaf child, if encouraged, may adopt a second alternative to auditory transmission in use of the manual alphabet through finger-spelling or he may adopt as his primary communication system, at least temporarily, the language of signs with finger-spelling as a supplement. The difficulty in acquiring and depending upon vocalization, auditory language reception, and speech reading as primary modes, is illustrated by the prevalence of manual communication among deaf adults when they communicate with others who possess the ability to communicate manually.

Recent investigations by Lunde and Bigman (1959), Rosenstein and Lerman (1963), and Boatner, Stuckless, and Moores (1964) have disclosed that employed deaf adults tend to rely heavily on written language for communicating with supervisors and foremen on the job. Lunde and Bigman found that 68 percent of approximately 8,000 deaf adults queried, indicated full or partial dependence on written language for direct interpersonal communication. Rosenstein and Lerman (1963) found that among 118 employed deaf women, 60 percent were fully or in part dependent on written language for direct communication regarding their work. Boatner et al. (1964) reported a corresponding figure of 62 percent among 101 young deaf employees.

Accordingly, there is considerable evidence that in certain settings, considerably over one-half of the deaf population is dependent upon written language in direct interpersonal communication with hearing persons. Speech remains an important asset to the deaf student as he emerges from the school setting into the adult social milieu, and indeed for many deaf adults it remains a primary communication mode. However, the heavy reliance

of most deaf adults on written communication is evident.

It is generally acknowledged that social, educational, and intellectual judgments about persons are at least in part based on their language usage. Social acceptance, job seeking, and job advancement, for example, are influenced by language facility. Accordingly, it is important that the written as well as the spoken language of the deaf be as polished as possible. A poorly informed hearing public may draw unwarranted inferences about the deaf person directly from his written language production.

Finally, unlike the hearing child whose written and reading systems tend to evolve from his basic vocal communication system, the deaf child is expected to acquire the language which comprises this vocal communication system largely through transfer from written language. Refined written language, then, is supportive of his spoken language.

It may be seen that skill in producing written language is crucial to the general social and economic adjustment of the deaf. Cognizant of this fact, educators of the deaf have traditionally given major attention to the written language development of deaf students.

Considerable research energy has been expended on the study of written language of the deaf. However, research is handicapped by the absence of a

valid instrument for evaluative purposes. Such an instrument should (1) be objective in nature, (2) be based on actual language production, (3) possess construct validity based on strong external criteria, and minimum error of estimate, (4) yield norms which are representative. The major objective of this investigation was to attempt to develop such an instrument.

B. Review of the Literature

1. Concepts of spoken and written language

Most research on the expressive language of hearing children has been based on spoken rather than on written output. Since the primary communication mode of expression for hearing children, and a desired outcome for deaf children, is oral, it follows that spoken language should receive primary attention. Carroll (1955) has defined language as:

> "--a structured system of arbitrary vocal sounds and sequences of sounds which is used, or can be used, in interpersonal communication by an aggregation of human beings, and which rather exhaustively catalogs the things, events, and processes in the human environment." (p.10)

It is notable that this definition restricts language to its vocalized form. Carroll would not speak of written language, but rather of "the system of writing language X" (1955, p. 11). McCarthy's (1954) classical review of the literature on language development in children reflects the fact that major attention has been given to oral communication skills in the developing child. Since for hearing children and for many deaf children language has both its origin and its major outlet in speech, major attention deservedly remains in this area.

Fries (1963) has attached higher status to written language than have Carroll and others. Fries has stated:

> "All 'writing' is the substituting of patterns of graphic shapes to represent the language signals of a code for the patterns of sound waves that have been learned as representing the same language signals." (p.119)

Fries suggests a congruency between the patterns of "graphic shapes" of writing and of "sound waves" of speaking, both having direct

reference to the same language signal. If indeed this is not so, Joos (1964) suggests this is only because teachers of English inadvisedly inhibit students from "feeling free to write by ear". It is notable that the teacher of English at the secondary level does not teach "written language" but rather "composition". Teachers of English are presently giving major attention to the problems of how to evaluate or assess English (Meckel, 1963), how to profitably teach English (Frances, 1958), and even to the constituents of English (Roberts, 1958).

2. Assessment procedures

Several methods of evaluating written language have been developed. One traditional method has been that of the standardized language test. Typically this test leans heavily on formal knowledge of grammar and spelling. The validity of this type of instrument with respect to tapping general skill in language usage depends upon the assumption that there exists a strong relationship between formal knowledge of grammar and spelling, and written language expression. However, Meckel (1963) after an exhaustive review of the literature has stated, "There is no research evidence that grammar as traditionally taught in the schools has any appreciable effect on the improvement of writing skill" (p. 981). Meckel's conclusion, of course, has reference to hearing students.

A second general approach to the evaluation of written language has been through direct subjective evaluation of written compositions, a grade or score being assigned on the basis of an overall appraisal. While this system probably has considerable validity, several problems present themselves. French (1962) has suggested error in assessment may have four sources: (1) student error, attributable to varying performance on the student's part from day to day, (2) test error, the single composition being comparable to a one-item test, (3) scale error, some markers being "easy", others "tough", and (4) reader disagreement, disagreement among readers as to what constitutes good and poor writing.

French (1962)[1] studied inter-judge agreement, and found that when a group of 53 readers representing several professions independently graded 300 student essays, the mean correlation among readers was only .31. When ten English teachers' grades were compared the mean correlation rose to .41. English teachers revealed greater but still low agreement.

Considerably higher inter-judge agreement may follow if standards are presented to the judges, scoring criteria carefully presented, and considerable training offered to the judges to produce maximum

---

[1] Diederich, French, and Carlton (1961) factor-analyzed scores, and found five predominating factors which they identified as ideas, form, flavor, mechanics, and wording.

agreement. When readers were thus prepared before scoring compositions for the College Board Examinations, inter-judge reliability rose to .70 (French, 1962). While the increased inter-judge reliability is highly desirable, a problem is introduced. Predetermined standards and criteria for scoring in turn assume predetermined judgments of quality of language. Validity of scores becomes in part dependent on the validity of the predetermined criteria, eg., relative weight to be attached to style and syntax.

Finally, assessment of written language has been attempted by means of rating scales. One such scale has been developed by Rosner (1963). This rating depends upon both the reader's general impression and five general variables established through factor analysis. Fourteen scales were developed around these five variables, four related to ideas, four to mechanics, two to wording, two to form, and two to flavor. Notably, "general impression" continued to receive weight.

3. Analytic procedures

Expressive language lends itself to description. Some descriptive procedures suggest evaluative power. For example, many cross-sectional studies have been conducted on the language of various ontogenetic levels of children. Sentence length and

complexity increases with age. Frequency of usage of various parts of speech varies with age, as do other language variables. However, one also finds major differences among individuals in writing style as noted by examinations of styles of noted authors. Several techniques used for describing written language are often incorporated into "readability formulae", used to estimate appropriate grade level of reading materials (Klare, 1963). Sentence length in particular is often used in these formulae of which Klare lists 41.

Among the more frequently employed analytic and implicity evaluative procedures are the following:

(a) Sentence length

(b) composition length

(c) sentence complexity

(d) frequency of usage of various parts of speech

(e) grammatic correctness

(f) spelling

(g) diversity of vocabulary

McCarthy (1954) has reviewed the literature, citing 14 studies of the length of sentences used by hearing children. Heider and Heider (1940), Myklebust (1960), Goda (1959), and Gunderson (1965) have studied the length of sentences in the written language of the

deaf. While some disparity is present among the findings of the individual investigators with regard to the mean sentence length at particular age levels, there is general agreement that (1) mean sentence length tends to increase at least through middle adolescence, and that (2) according to the studies which compared deaf and hearing students, deaf students consistently produce shorter sentences than do hearing students.

Unfortunately, the results of most studies of composition length cannot be directly compared. Assigned topics or other stimuli have differed as have the amounts of time provided for writing. Among those who have investigated the length of compositions written by deaf students are Heider and Heider (1940), Myklebust (1960), Goda (1959), and Simmons (1963). In general, composition length was seen to increase with age. However, a plateau was suggested as students reached late adolescence. In those studies comparing deaf and hearing subjects on the same topic, the hearing student was seen to produce greater length at mid and upper ages. Myklebust found, however, that at age seven, deaf students produced greater length than did hearing students of a similar age.

Among those who have examined the complexity of sentences written by deaf students are Walter (1959), Heider and Heider (1940), and Goda (1959). These investigations have in general adapted a classification system developed by McCarthy (1930) by which sentences are ranked in complexity, varying from structurally incomplete sentences to elaborate sentence constructions. In general, the sentence constructions of the deaf are concluded to be simpler than those of the hearing. Goda found that among deaf students, complexity of sentences correlated .57 with composition length and .69 with mean sentence length. Williams (1937) had similarly found that among hearing children, sentence complexity and sentence length correlated .80. Sentence complexity and sentence length apparently share considerable variance.

Frequency of usage of various parts of speech has been examined with varying populations and under varying conditions. Most analyses have been conducted on adult populations, eg., college freshmen (Mann, 1944), noted writers (Yule, 1944), adult telephone conversations (Zipf, 1935). Others have been concerned with developmental characteristics of the language of children (Templin, 1957; Davis, 1937).

Myklebust (1960), Simmons (1963) and Goda (1964) have studied the relative frequency of usage of parts

of speech in the language of deaf students. Myklebust used a traditional classification system, while Simmons and Goda have used a structural classification system based on the work of Fries (1952).

Goda found deaf adolescents to employ relatively less Class III and Class IV words (adjectives and adverbs), and less function words (prepositions, conjunctions, articles, etc.) than hearing adolescents. Myklebust found pronouns, prepositions, adjectives adverbs and conjunctions to be used less frequently by deaf than by hearing students. The literature in general suggests that relative frequency of usage of adjectives, adverbs, and function words tends to increase as the child matures, while relative frequency of nouns and verbs tends to decrease.

Numerous investigators have studied grammatic errors in the written language of the deaf. Thompson (1936) classified errors in syntax as errors of substitution, omission, addition, and word order. Myklebust (1960) adapted and added to Thompson's system in such a way as to permit assignment of a syntax score to language productions. Similarly, Birch and Stuckless (1963) developed a system for scoring language from direct inspection of grammatic errors in compositions. Most recently, Gunderson (1965) has studied compositions

written by hearing and deaf students, deriving a scoring procedure whereby different weights are attached to different syntactical errors.[1]

Spelling has received relatively little attention in investigations of the written language of deaf students. Gunderson (1965) found that deaf students make few errors in spelling. If so, this is possibly attributable to the fact that deaf children tend to be less dependent on phonetic aspects of words for correct spelling, tending more toward learning to spell words directly from their graphic representations.

With the development of the type-token ratio by Johnson (1944), a useful instrument was provided for the study of diversity of vocabulary. The ratio (TTR) is expressed as number of different words divided by total number of words in a sample (or a predetermined number of words such as the first 50 or 100 words). Simmons (1963) compared the TTR of language samples written by deaf and by hearing students, and found the deaf students to be more redundant in use of words (lower TTR) than hearing students. Simmons extended her investigation to a determination of type-token ratios for different parts of speech. She found hearing

---

[1] The reader is referred to Gunderson (1965, pp. 7-50) for an excellent review of the literature related to grammatic errors produced by deaf students.

students to be more versatile than deaf students in their use of all parts of speech except class III words (adjectives).

As indicated earlier, most of the above techniques for describing language have an evaluative connotation. However, few have been employed with an explicit outside criterion. This investigation is in part concerned with the relationships among various analyses, but also with their individual relationships with outside judgments of language made by teachers of the deaf.

Although subjective evaluations of language have been seen to be deficient in reliability, it is speculated that inter-judge reliability among teachers of the deaf is relatively high due to the fact that syntax and intelligibility receive more attention than "flavor" and ideational aspects from teachers of the deaf. If this is not so, correlations between objective analyses and teacher-judgments are expected to be weak.

C. Objectives

Objectives of the investigation were:

1. to describe compositions written by deaf students 10 through 18 years according to the following measures:

a. composition length

b. sentence length

c. frequency of usage of Class III words, Class IV words, and function words, relative to total words.[1]

d. variety of vocabulary usage[2]

e. grammatic correctness

f. spelling correctness

2. to determine the relationships between each of the above measures and the criterion of teacher-judgment.

3. to develop scoring procedures[3] for assessing compositions written by deaf students 10 through 18 years.

---

[1] Class I words are typified traditionally by nouns, Class II words by main verbs, Class III words by adjectives, Class IV words by adverbs, and function words by prepositions, articles, conjunctions, etc. (Fries, 1952).

[2] Measured by the type-token ratio

[3] Multiple-regression equations, with teacher-judgment serving as the criterion and measurable aspects of compositions as the predictor variables (one for each level 10-18, and one for the entire age distribution, with national norms

II. PROCEDURE

A. Pilot Study

As reported by Birch and Stuckless (1963), the selection of appropriate stimuli (pictures and verbal topics) for the elicitation of written language from deaf students is critical. Styles of writing, grammatical correctness, and length of a composition will vary within the same student as a function of the stimuli presented to him.

Accordingly, a pilot phase of this investigation served to guide the investigation in:

(a) selecting a stimulus which would elicit maximum length of written response,

(b) testing several objective means of describing the compositions,

(c) determining the optimum conditions for collecting composition samples in terms of instructions and length of time to be permitted each student for writing,

(d) providing an indication of the basal age at which deaf students might be expected to produce a composition of sufficient length to be scorable with reasonable reliability.

Seven classes totaling 57 students were selected from the population of the Western Pennsylvania

School for the Deaf, Pittsburgh, to participate in the pilot phase. These classes were selected on the basis of being considered by supervising teachers as representative of deaf students varying from 7 through 18 years of age in that school. A meeting was held with the seven teachers of the above classes, information presented about the investigation, and oral and written instructions given to each teacher.

Five stimuli, four pictorial and one verbal, had been selected for presentation to the students. An artist was employed to draw the pictures, and these pictures were reproduced to provide each student with a copy at his desk. The stimuli and specific directions were as follows:

(a) A sequence of four related pictures. Teachers were instructed to write on the chalkboard and say, "Write a long story about the pictures. You have 45 minutes."[1]

(b) Picture of jet airliner in flight. Teachers were instructed to follow the above procedure, and to write and say, "Write a long story about the picture. You have 45 minutes."

(c) Picture of family shopping. Teachers were instructed to write and say, "Write about the picture. You have 45 minutes."

---

[1]Subsequently selected, with revisions in the pictures and directions.

(d) Playground scene. Teachers were instructed to follow the instructions as in "c".

(e) Letter. Teachers were instructed to write and say, "Write a long letter to your mother and father. You have 45 minutes." (Students could write to a friend if unable to write to parents).

Each class was presented with each of the five stimuli, at the rate of two stimuli per week. Students were given 45 minutes in which to write their compositions. An observer was present in each class to take notes on time given by each student to the composition, relevant questions asked, etc.

Total words in each composition were counted, mean composition length determined for each age, 7 through 18 years, and the mean composition length regardless of age determined for each stimulus. Table 1 indicates mean composition length for each stimulus, without regard to age.

Table 1. Mean composition lengths from five stimuli for 57 students 7-18 years

	Mean composition length
Picture sequence[1]	205.0 words
Jet in flight[1]	158.0 words
Family shopping[1]	167.7 words
Playground scene[1]	158.9 words
Letter (body of letter only)	187.2 words

[1] Pictorial

The picture sequence, as noted in Table 1, tended to produce the greatest composition length. Also, observations suggested greater interest on the part of students to the sequence than to other stimuli. Therefore, the picture sequence was concluded to be the best suited to the purposes of the investigation.

A second point of interest concerned the ability of the younger students to write compositions of reasonable length (tentatively set at 50 words or more for purpose of reliability of assessment). Table 2 reveals the percentage of students in four age groups who wrote compositions of at least 50 words in response to each of the five stimuli.

Table 2. Percentage of students who wrote compositions of 50 words or more

	Age in years			
	7-9 (n=17) per cent	10-12 (n=16 per cent	13-15 (n=15) per cent	16-18 (n=9) per cent
Picture sequence	53	87	100	100
Jet in flight	47	93	100	100
Family shopping	71	100	93	100
Playground scene	53	93	100	100
Letter	65	100	100	100

Inspection of Table 2 indicates that a relatively high proportion of the students under ten years tended to write compositions of less than 50 words, while most students ten years and older wrote compositions of 50 words or more. This finding was important in terms of the establishment of a basal age cutoff for the investigation.

The type-token ratio (TTR) is generally based on a minimum of 100 words. However, inspection of the compositions revealed that if students were required to write a composition of 100 words or more, the applicability of the type-token ratio as an assessment instrument would be considerably restricted.

As a check of the sensitivity of the TTR based on 50 and 100 words, the "picture sequence" compositions were retyped, identification was removed, and two experienced teachers of the deaf were asked to independently assign a subjective score to each of the 57 compositions. Type-token ratios first were calculated on the first 50 words (n=47), then on the first 100 words (n=42). A Pearson product moment correlation coefficient was calculated between a combined T score of the two teacher-judges and each of the two methods of calculating the TTR. The

TTR based on 50 words correlated .70 with teacher-judgments, and the TTR based on 100 words correlated .69 with teacher-judgments. Both coefficients appeared reasonably high when considered as validity coefficients, and that based on 50 words as high as that based on 100 words.

Several additional objective descriptive techniques were applied to the compositions, and refinements in these techniques were made. Others were discarded as being unreliable or adding no additional information to that available from other techniques (eg., "number of structural units per sentence" correlated so high with "sentence length" as to appear to measure virtually the same variable).

The picture sequence was shown to several colleagues who drew attention to the fact that a cultural loading was present. The picture sequence was accordingly redrawn to reduce possible cultural loading.

The directions to students were revised from "Write a long story about the pictures," to "Write about the pictures."

In summary, the pilot phase led to:

(a) selection of a picture sequence for presentation,

(b) the decision to establish the C.A. limits at 10 years, 0 months through 18 years, 11 months,

(c) the decision to restrict assessment to compositions of 50 words or more,

(d) modification of the picture sequence to reduce cultural loading,

(e) revision of directions to students,

(f) acceptance of general directions to teachers (Appendix A),

(g) selection and refinement of objective descriptive techniques (II. D. 3. Objective analyses).

B. Sampling Procedures

1. Population parameters

The parameters of the population from which the final sample was to be drawn, and of which norms would be considered to be representative, were:

(a) Public and private residential and day schools for the deaf in the United States with student populations of 100 or more as listed in American Annals of the Deaf, January, 1964, and extending through a minimum of seven year levels of instruction.

(b) Male and female students, age 10-0 through 18-11 with 70 decibel hearing losses or more in the better ear as measured by pure tone audiometry (mean loss - 500, 1000, 2000 cps).

(c) Students who write compositions of 50 words or more under conditions to be described.

For the purpose of census, the U.S. Census Bureau has divided the continental United States into nine regions, each encompassing several states. Table 3 is a summary of numbers of schools for the deaf in the United States and student populations as defined by (a) above.

Table 3. Schools for the deaf and student populations in the U.S.[1]

Region	Residential			Day		
	No.	Pop.	Per cent[2]	No.	Pop.	Per cent[2]
1. New England	6	1201	6.2	-	-	-
2. Mid-Atlantic	12	3077	15.8	3	730	3.7
3. E.N. Central	8	2183	11.2	7	1218	6.2
4. W.N. Central	8	1840	9.4	2	323	1.6
5. S. Atlantic	11	2873	14.8	-	-	-
6. E.S. Central	6	1377	7.1	-	-	-
7. W.S. Central	7	1586	8.1	1	235	1.2
8. Mountain	4	745	3.8	-	-	-
9. Pacific	4	1559	8.0	3	548	2.8
TOTAL	66	16,441	84.4[3]	16	3,054	15.5[3]

Total, all schools - 82
Total population, all schools - 19,945

[1]Meeting criteria of 100 students or more, and seven levels of instruction (Am. Ann. Deaf, January, 1964)
[2]Per cent of total population, residential and day (19,495)
[3]Minor discrepancy due to rounding

It is noted that the residential schools comprising the population from which the sample was selected numbered 66, and the student population numbered 16,441. Similarly, day schools numbered 16, with a student population of 3,054. Residential schools contributed to 84.4% of the total student population (19,495) and day schools contributed to 15.5% of the total (minor discrepancy of .1% due to rounding).

2. Schools from which samples were drawn

From each geographical region (Table 3) a residential school and, where represented, a day school, were randomly selected.[1] From Table 3, it is noted that this step involved the selection of nine residential schools[2] and five day schools.

Schools invited and subsequently volunteering to participate as representatives of the population are listed in Table 4.

Table 4. Programs participating in investigation

Region	Residential	Pop.	Day	Pop.
1. New England	Gov. Baxter Portland, Maine	135	- -	- -
2. Mid-Atlantic	St. Mary's Buffalo, N.Y.	303	P.S.47, 158 New York,N.Y.	379
3. E.N. Central	Illinois S.D. Jacksonville, Ill.	445	Columbus Day Columbus, Ohio	124
4. W.N. Central	Nebraska S.D. Omaha, Neb.	155	Minneapolis Day Minneapolis,Minn.	219
5. S. Atlantic	Florida S.D. St.Augustine, Fla.	428	- -	- -
6. E.S. Central	Tennessee S.D. Knoxville, Tenn.	389	- -	- -
7. W.S. Central	Texas S.D. Texas S.for D.B.O. Austin, Texas	605	Houston Indep. S.Dist. Houston, Texas	235
8. Mountain	Utah S.D. Ogden, Utah	201	- -	- -
9. Pacific	California S.D. Berkeley, Cal.	480	Univ. Hts. S. Seattle, Wash.	161
TOTAL		3,141		1,118

[1] Throughout this study, a table of random numbers was used for all random selections.

[2] Three exceptions were made. (1) It was necessary to add a second school to each of two regions to assure numerical representativeness.

3. General procedures

Prior to the collection of compositions, the chief administrative officer of each of the 14 above programs was contacted by mail. Included in the correspondence was an abstract outlining the purposes and procedures of the investigation. Follow-up telephone calls were then made, remaining questions answered, and dates established for the collection of compositions.

Upon approval of the administrators of the various schools and school districts, directions for the test administration were sent to participating teachers (Appendix A). Actual test materials were not sent at this time.

In order to assure that students of 10 years and above would be adequately represented, compositions were collected from classes whose mean ages were eight years plus if they included a student of 10 years or older.

A member of the research team visited each school immediately prior to and during each test administration in order to (a) deliver materials, (b) discuss the investigation and specific directions with teachers and (c) collect compositions immediately after they were written. All testing was completed within a two-month period (January-February, 1965).

4. Compositions collected

Compositions were written in response to the picture sequence by a total of 352 classes of deaf students, numbering 2798 students. Of this total, 2181 and 617 were students in residential and day programs respectively.

Schools were also asked to report on birth-date, sex, and hearing loss of each student (db loss, better ear, 500, 1000, 2000 cps., pure tone, most recent test).

5. Conformity to population parameters

Of the 2798 students, 431 were deleted by the investigators because their chronological ages were less than 10 years, 0 months, or more than 18 years, 11 months (rounded off to lowest month at date of testing). A total of 2367 students fell within these C.A. limits.

Since only those students with hearing losses of 70 decibels or greater in the speech range were to be included in the final sample, additional students were deleted. Table 5 indicates the number of students 10 years, 0 months through 18 years, 11 months whose hearing loss was less than 70 db. and 70 db. or greater.

Table 5. Age, sex, and hearing loss of 2,367 students of 14 programs

C.A. (yrs.)[1]		Residential		Day		Total	
		<70 db.	≥70 db.	<70 db.	≥70 db.	<70 db.	≥70 db.
10	M	20	92 (82%)	10	31 (76%)	30	123 (80%)
	F	14	69 (83%)	7	26 (79%)	21	95 (82%)
11	M	13	72 (85%)	13	22 (63%)	36	94 (78%)
	F	12	62 (84%)	9	25 (74%)	21	87 (80%)
12	M	17	90 (84%)	13	30 (70%)	30	120 (80%)
	F	15	89 (86%)	8	26 (76%)	23	115 (83%)
13	M	20	103 (84%)	5	29 (85%)	25	132 (84%)
	F	13	88 (87%)	9	29 (76%)	22	117 (84%)
14	M	23	93 (80%)	6	20 (77%)	29	113 (80%)
	F	11	88 (89%)	3	20 (87%)	14	108 (89%)
15	M	20	132 (87%)	7	19 (73%)	27	151 (85%)
	F	23	93 (80%)	1	17 (94%)	24	110 (82%)
16	M	27	108 (80%)	3	20 (87%)	30	128 (81%)
	F	16	95 (86%)	5	18 (78%)	21	113 (84%)
17	M	16	95 (86%)	3	11 (79%)	19	106 (85%)
	F	13	74 (85%)	4	17 (81%)	17	91 (84%)
18	M	13	74 (85%)	5	14 (74%)	18	88 (83%)
	F	4	63 (94%)	1	11 (91%)	5	74 (94%)
TOTAL, (male)		169	859 (83.6%)	65	196 (75.1%)	234	1,055 (81.8%)
TOTAL, (female)		121	721 (85.6%)	47	189 (80.1%)	168	910 (84.4%)
GRAND TOTAL		290	1,580 (84.5%)	112	385 (77.5%)	402	1,965 (83.0%)

[1] in years (0-11 months)

The samples of the population which would subsequently be subject to analysis, would be drawn from 1,965 students in the 14 programs, all of whom met the following criteria: chronological age 10-0 through 18-11, and 70 db. hearing loss or greater in the speech range.

From this number, a total of 900 subjects would be drawn:

(a) 50 at each age level in years between 10 and 18 years inclusive (450 subjects), for variable age sample,

(b) 100 at each age level in years between 10 and 18 years inclusive (900 subjects, 450 of whom would be the same subjects as selected for (a), for nine constant age samples.

In order that each geographical region would be accurately represented in the samples with regard to students in residential and day programs, the percentage figures in Table 3 were used. For example, residential schools in the Pacific region (California, Oregon and Washington) contribute 8.0 per cent of the total population of residential and day programs meeting the criteria discussed earlier. Accordingly, 8 per cent of the subjects in each sample would be drawn from the residential school representing that region.

Although males outnumbered females in the 1,965 students meeting the basic criteria (1,055 males, 910 females) it was decided that males and females would each represent 50 per cent of each sample to simplify statistical analysis.

Table 6 provides an enumeration of subjects by region, type of program and sex selected to represent samples of 50 and 100. Percentages indicated in Table 3 served as the basis for selecting different numbers of subjects from different regions and programs.

Table 6. Selection of subjects for sub-samples of 50 [1,2]

Region	Residential Number[2]	Day Number[2]
1. New England	3 (2m,1f)[3]	--
2. Mid-Atlantic	8 (4m,4f)	2 (1m,1f)
3. E. N. Central	6 (3m,3f)	2 (1m,1f)
4. W. N. Central	4 (2m,2f)	1 (1m)
5. S. Atlantic	8 (4m,4f)	--
6. E. S. Central	3 (1m,2f)	--
7. W. S. Central	4 (2m,2f)	1 (1f)
8. Mountain	2 (1m,1f)	--
9. Pacific	4 (2m,2f)	2 (1m,1f)
TOTAL	42 (21m,21f)	8 (4m,4f)

[1] See Table 3 for calculation of percentage.
[2] Multiply by two for samples of 100 (constant age samples); multiply by 9 for sample of 450 (variable age sample).
[3] m = male, f = female

The pilot study led to the decision to restrict the parameters of the population to students who wrote compositions of 50 words or more. It was evident that some students over 10 years would not meet the latter criterion. In addition, some would write lists of words without any attempt to write in sentence form.

As a general check on the findings of the pilot study, the investigators randomly selected 50 subjects at each age level, 10-18, as stratified in Table 6. In addition, since 326 compositions written by 8 and 9 year old students had also been collected, it was decided to sample 8 and 9 year old students in order to confirm (or reject) the decision to delete these two age levels.

Table 7 indicates the length and several characteristics of the productions of 550 students selected in accordance with ratios indicated in Table 6.

Table 7. Productions of 550 students (8-18 yrs.) in response to picture sequence

CA[1]	Mean Comp.[2] Length	No. in Sent. Form	No. Word Lists	No. Drawings	No. Blanks	No. Comp.[3] 50 words plus	No. Comp.[3] 100 words plus
8[4]	32.38	33	12	3	2	11	2
9[4]	52.40	38	9	3	-	22	8
10	131.34	45	4	-	1	35	18
11	138.76	47	3	-	-	38	28
12	137.64	46	2	1	1	40	27
13	164.08	48	1	-	1	46	37
14	195.80	50	-	-	-	48	44
15	201.60	48	2	-	-	46	39
16	217.58	50	-	-	-	46	43
17	227.36	50	-	-	-	49	48
18	235.70	50	-	-	-	50	44

[1]50 subjects at each age level (0-11 months)

[2]Including compositions of less than 50 words and word lists.

[3]Included to confirm decision to include compositions of 50 words or more in sample.

[4]Included to confirm decision to establish basal CA at 10 years, 0 months.

Inspection of Table 7 reveals that the population parameters would be greatly restricted if they included only students who wrote 100 words or more (eg., only 36 per cent of the 10 year olds produced 100 words or more, while 70 per cent produced 50 words or more).

Restriction of the population parameters to students of 10 years or older is also supported by reason of the fact that only 22 per cent and 44 per cent of the 8 and 9 year olds respectively wrote 50 words or more, a minimum number of words acceptable for reasonably reliable evaluation.

6. Selection of final samples

(a) Variable age sample (10-18 yrs.)

In order to select the samples described in Table 7, compositions written by all students (students with less than 70 decibel hearing losses deleted) were placed in rank order by chronological age in months, and males separated from females. Programs were kept separate. Where two schools represented one program (eg., P.S. 47 and 158 in New York City) these were drawn together as one school.

Subjects were randomly selected, with stratification based on program, age, sex, and number as described in Table 6. This yielded the sample of 450 subjects (50 at each age level 10-18 yrs.).

Subjects who wrote compositions of less than 50 words were deleted and replaced by substitutions representing the same program, age, and sex. A total of 52 substitutions were made within the sample of 450.

In 20 instances, it was necessary to select a student from a school other than that from which it was designated that he be drawn, because no subject of the required age and sex, and meeting hearing loss and composition length criteria was available from that school. In these instances, alternate selections were made from the same type of school (residential or day).

A general description of the variable age sample is indicated in Table 8.

Table 8. Subjects comprising variable age sample

C.A. (yrs.)	No. (25m,25f)		Mean C.A. (mos.)	Mean db. Loss[1]
10	50		125.74	89.1
11	50		137.68	88.8
12	50		149.34	88.9
13	50		161.64	85.0
14	50		173.34	90.2
15	50		186.46	91.0
16	50		196.98	90.9
17	50		209.02	89.8
18	50		220.18	90.9
TOTAL	450 (225m, 225f)	Grand Mean	173.38	89.6

[1]Better ear, 500, 1000, 2,000 cps., puretone

(b) Nine constant age samples

Sampling procedures were identical to those for the variable age sample, except that 100 subjects were selected for each of the nine constant age samples (10-18 years inclusive). To the 50 subjects at each age level within the variable age sample (Table 8) were added 50 additional subjects, under the same stratified random selection procedures (Table 6). This resulted in 100 subjects for each of the nine constant age samples.

As with the selection of the variable age sample, it was necessary to make substitutions for subjects who had written less than 50 words, and to substitute a subject from a second school when no selection of an appropriate age and sex could be made from the first school.

Table 9 generally describes the nine constant age samples

Table 9. Subjects comprising nine constant age samples

C.A. Sample (Years)	No. (50m 50f)	Mean C.A. (mos.)	Mean db. Loss[1]
10	100	125.90	90.1
11	100	137.57	88.9
12	100	148.78	87.6
13	100	161.29	87.9
14	100	173.22	89.3
15	100	186.02	90.0
16	100	196.96	90.6
17	100	209.17	88.5
18	100	219.84	88.7

[1]Better ear, 500, 1000, 2000 cps., puretone

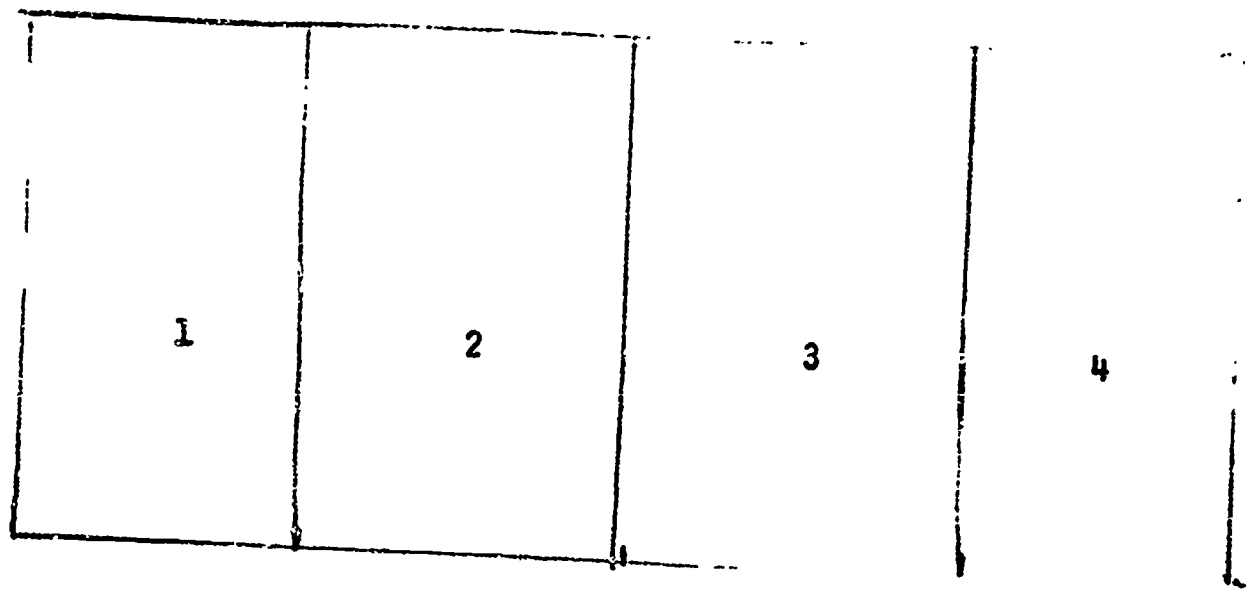
C. Administration of Written Language Test

Earlier in this chapter, it was indicated that directions for the administration of the Written Language Test were sent to all participating teachers (Appendix A). A total of 352 teachers received these directions. A member of the investigating team personally visited each school immediately prior to the test administration to answer any remaining questions and reiterate the directions. In most cases, it was possible to speak directly to the

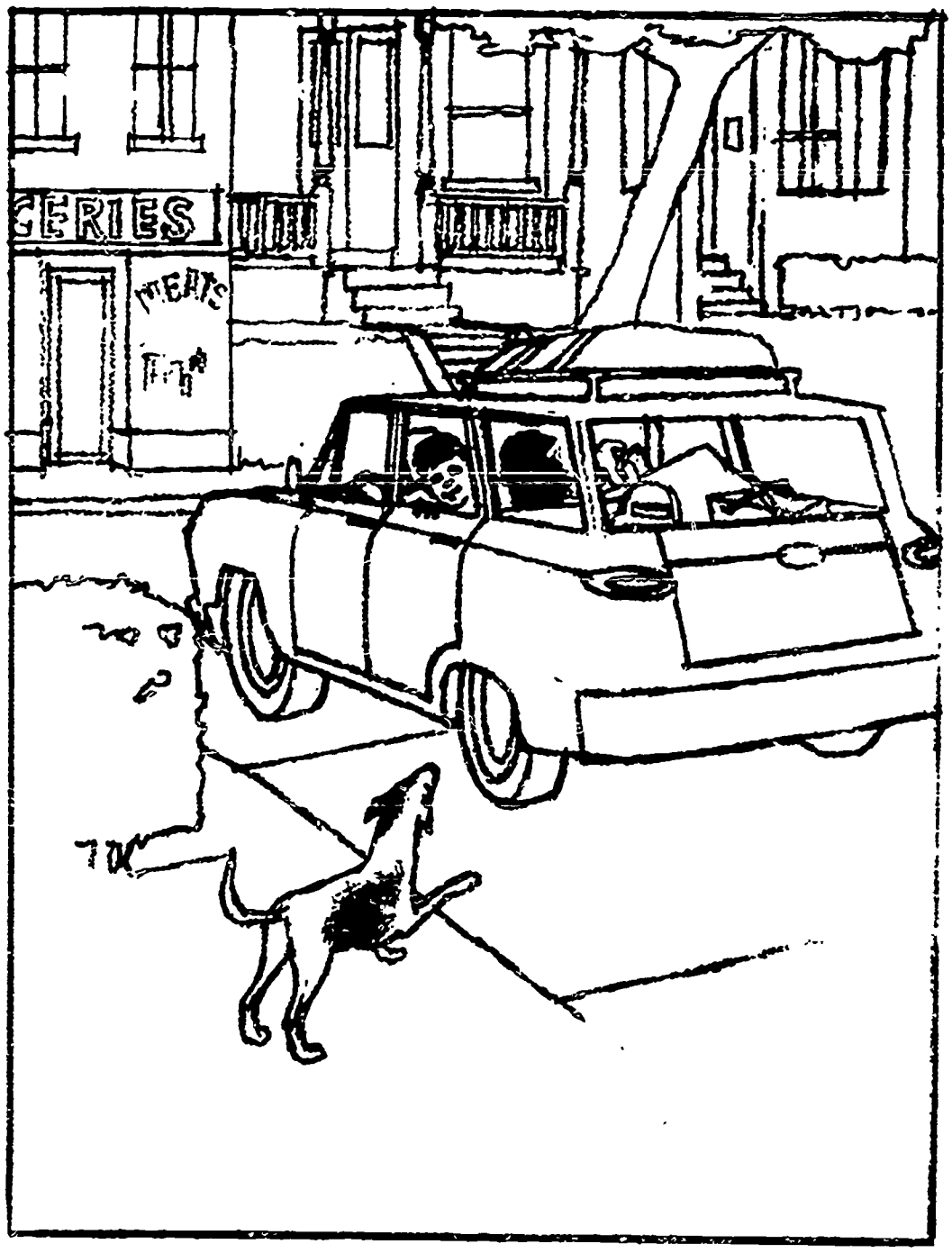
teachers in groups. Where teaching schedules did not permit this, a meeting was held with the supervisory staff to discuss the directions in detail. They then transmitted the directions to the teachers. As indicated earlier, compositions were written under the following conditions by 2798 students.

1. Picture sequence

The picture sequence consisted of four related pictures. These pictures were drawn for the project in black and white, and multilithographed on a single sheet of paper 20" by 6 1/2", in linear sequence, as shown below.



The sequence portrays a family of four, two parents, a boy and a girl, in the kitchen preparing for a picnic. The family's dog is present in the scene (picture 1).



In the second picture, the family is driving onto the street, the dog following. The son glances out of the station wagon window toward the dog (picture 2).

In the third picture, the father has stopped the car while the boy retrieves the dog (picture 3).

In the fourth picture, the family has arrived at the picnic site. The mother and daughter are preparing the lunch while the father, son, and dog play baseball in the background (picture 4).

2. Directions to teachers

The written directions to teachers are indicated in Appendix A.. Teachers were urged to exercise objectivity in the administration of the test. They were asked not to open the test packet (containing copies of the picture sequence for each student, legal size lined note paper,[1] duplicate directions, and a class record form[2]) until immediately before the test. Teachers did not see the picture sequence until this time.

Upon collection of compositions, teachers were asked to place these compositions back in the packet for collection by a member of the supervisory staff of the school or the investigator.

Specific directions to teachers are as follows (Appendix A):

1. Write on chalkboard, "Write about the pictures. You have 45 minutes. You may use two sides of the paper."

2. Distribute to each student a sheet of paper and a copy of the PICTURE SEQUENCE. Please ensure that each student has a pencil.

3. Check that each student writes his full name on top right of writing paper. Indicate that they should then begin.

---

[1]Of this size in order to implicitly suggest length to students, and to permit most students to complete their compositions on a single sheet, both sides.

[2]To enter name, sex, birthdate, and db. hearing loss in better ear at 500, 1000 and 2000 cps. of each student in class.

4. DO NOT aid the children in any way. For example, some students may ask questions about the pictures or seek aid in spelling or selecting appropriate words. It is important that no help be given and no corrections be made for the child.

5. Collect the papers 45 minutes after the students have been told to begin. (After they have begun, if none of your students has written during a 5 minute period, collect all papers. This may occur before 45 minutes have elapsed. Enter the time elapsed on the Class Record Form).

6. Remember to return all materials except the Class Record Form to the package. A pr ject staff member will collect the package.

D. Descriptive and Evaluative Techniques

The compositions written by the 450 subjects comprising the variable age sample (Table 8), and the compositions written by the 900[1] subjects comprising the nine (10-18 years) constant age samples (Table 9), were accurately transferred (original content, paragraphing, errors, etc. being retained) to typewritten form and mimeographed (see Appendices D, E, F). This served several purposes: (a) all identifying information on the original compositions could be, and was, coded to provide anonymity in terms of the identification of the subject's sex, age, hearing loss, and the program he represented, (b) handwriting was eliminated as a possible variable, (c) several copies of each composition became

[1]Inclusive of the 450 subjects within the variable age sample also.

available for simultaneous description and evaluation by different persons. Notations on copies also became permissible.

Two basic techniques were employed, teacher-judgments and objective descriptions, to describe and evaluate the compositions written by the subjects represented in Table 8 and Table 9.

1. Teacher-judges

Because of the relatively low agreement generally expressed by English teachers in ranking compositions written by normal students (French, 1962), it was imperative that more than one judgment be obtained on the relative evaluations of compositions written by the subjects in the ten samples. In the interest of increasing reliability of judgment and broad coverage of criteria used by teachers of the deaf to assess language, independent teacher-judgments were obtained on the compositions of each sample (Guilford, 1954, pp. 394-397). An estimate of the inter-judge reliability is presented in the following chapter.

Considerable care was taken in the selection of the three teacher-judges. Numerous eminent educators of the deaf were contacted and a list of possible teacher-judges was compiled. The three

judges who were invited to participate, and who joined the project, had been consistently highly recommended as skilled teachers of the deaf with particular competencies in teaching language to deaf students. One had graduated from the University of Manchester, an internationally recognized center for the training of teachers of the deaf, had subsequently taught at the Clarke School for the Deaf, Massachusetts, and has more recently been engaged in the preparation of teachers of the deaf. Another judge had trained as a teacher of the deaf and until recently taught at the Lexington School for the Deaf. Another judge had trained to teach the deaf at the University of Wisconsin at Milwaukee, taught the deaf in a public day school setting and has more recently been engaged in the preparation of teachers of the deaf in a University setting.

While all possess acknowledged competencies in language instruction to deaf students, their training and sources of experience have differed, (eg., the Clarke and Lexington Schools are both recognized for their national influence on language instruction, but certain aspects of their techniques differ considerably). Their judgments should, therefore, reflect differing criteria for evaluation of

written language of the deaf, if indeed true differences exist.

None of the three judges had taught in any of the schools from which the samples were drawn.

2. Assignments to teacher-judges

The three teacher-judges evaluated the compositions in their homes. They were not asked to convene, lest they might indirectly influence each other's judgment criteria. They were free to contact the investigators at any time if questions pertaining to the evaluation of the compositions arose.

A standard set of written directions for scoring compositions was sent to each (Appendix B).

A total of ten packets of compositions (mimeographed and uncorrected) was sent to each judge. The mailing of these packets was staggered at intervals of several days. The judges were asked to complete scoring and return each packet before scoring the succeeding packet. Judges were asked to score on the basis of criteria which they would ordinarily apply in judging a composition, using a numerical scoring system with a ceiling of 100.

The first sample to be sent to judges consisted of the 450 compositions comprising the

variable age sample (Table 8). Compositions had been coded and placed in random sequence. As indicated in Appendix B, judges were informed that the compositions were written by deaf students varying from 10 to 18 years of age.

Nine succeeding samples were then sent to the judges, one for each age level 10-18 inclusive (Table 9). Judges were told the C.A. range of each of these samples. They were instructed to score each sample independently. They were also informed that they had already scored 50 of the 100 compositions in each packet in the first sample of 450. It is highly improbable that the score on the first sample would influence their second score due to the large number which had been scored and the fact that the first sample had already been returned. Teacher-judges each devoted approximately one month full time to this demanding task.

Each of the three teacher-judges assigned a numerical score to each of 1350 compositions (450 being scored twice). Accordingly, each composition in each sample was given three independent numerical scores. These were subsequently converted to T scores, and a single normalized T score for each composition in each sample was calculated. This conversion will be

discussed in detail under II. E Statistical Procedures.

3. Objective analyses

Six objective analyses were made of each composition.[1] These analyses were made by members of the investigating team. A single method of analysis was applied to each composition, then a second method, etc., until all six analyses had been completed.

The six objective analyses conducted on each composition consisted of (a) composition length in words, (b) mean sentence length, (c) a description of parts of speech based on Fries' structural classification of words (Fries, 1952), (d) type-token ratio (number of different words comprising the first 50 words of the composition), (e) a count of grammatic errors within the first 50 words of the composition, and (f) a count of spelling errors, again with the first 50 words. A description of these six methods of analysis follows:

(a) Composition length This analysis consists of counting the total number of words in the composition. The following rules guide the determination of the word count:

---

[1]Three compositions and their scores appear in Appendices D, E, and F.

(1) Where not intelligible as a word, the word may be identified by a preceding and succeeding space or punctuation. The beginning letter of a word may sometimes be identified with capitalization.

(2) Titles assigned to a composition are disregarded as are closings such as "The end."

(3) Hyphenated words are treated as single words (eg., bow-wow).

(4) Compound identification of persons (eg., Mr. Jones) or location (eg., New York) is treated as two words.

(5) Contractions (eg., didn't, won't) are considered as two words, an exception to rule 1.

(b) Sentence length Following the total word count (composition length), the number of sentences within the compositions is counted. The following rules guide the determination of mean sentence length:

(1) The termination of a sentence is identified by a period, question mark or exclamation mark.

(2) The beginning word of a sentence is identified by capitalization. At least one of the above cues must be present to signify the introduction or completion of a sentence.

(3) The total number of words in the composition (composition length) is divided by the number of sentences within the composition to yield the mean sentence length (two decimal places).

(c) Word Structure ratio

The basic rules provided by Fries (1952) for classification of words into five categories are followed in this analysis. The reader is referred to Fries' The Structure of English (1952) for a detailed description.

The Class I word is generally seen as bearing resemblance to the noun, the Class II word to the verb, the Class III word to the adjective, and the Class IV word to the adverb. Function words comprise the remainder. Function words include words traditionally known as articles, prepositions, and conjunctions. Words may be identified by formal characteristics. The structural classification of a word may vary in accord with its position in the sentence. The use of the word "is" may serve to illustrate this point:

"Diane is ill. The physician is coming."

The formal significance of "is" in these two sentences differs. By structural classification, "is" as in "is ill" is considered a Class II word, whereas in

"is coming," this word would be considered a function word.

The word structure ratio is derived as follows:

(1) The first 50 words of the composition are rewritten under five columns representing the four classes and function words. Identical words may appear in two or more columns, if they represent different classes or function in different contexts.

(2) Where sentences are unintelligible, revert to classifying a word by lexical rather than structural meaning.

(3) The total words (words sometimes recurring) appearing in the Class III, Class IV and function word columns are added.

(4) The word structure ratio is calculated as follows:

$$\frac{\text{Class III, Class IV, Function words}}{50} \times 100, \text{ or (III, IV, F)} \times 2$$

(d) Type-token ratio

The type-token ratio is derived as follows:

(1) Count the number of different words in the first 50 words in the composition, giving the type count.

(2) The first 50 words are derived in accordance with the rules established to determine composition length.

(3) Variation in the spelling, suffix, etc. of two words otherwise similar, remains a single word, eg., if "boy" and "boys" both appear in the first 50 words, they constitute a single word.

(4) The TTR is derived as follows:

$\frac{\text{No. different words(types)}}{50} \times 100$, or (types) x 2.

(e) Grammatic correctness ratio

The grammatic correctness ratio (GCR) is intended to determine the syntactical and morphological correctness of the grammar of the subjects. Errors are counted but not classified.

The following rules dictate counting of grammatic errors in the first 50 words of the composition:

(1) Each sentence should be scored discretely. Verb tense, for example, is free to vary from sentence to sentence.

Eg., The family made lunch. The family will go in a car. The family played baseball (no error).

(2) If error is one of wrong word but correct part of speech, this is counted as one error; however, if wrong word and wrong part of speech, this is counted as two errors.

Eg., The boy threw the girl.[1] (1 error)

The boy threw the around. (2 errors)

---

[1]This error might be that the direct object was not added to the sentence. However, the sentence is structurally correct as it stands. Score in such a way as to give the student a minimum penalty.

(3) Disregard error within a word if the word is identifiable, except for suffixes indicating wrong person, tense, or number. A dash in place of a word should be counted one error.

Eg., The family go last Monday. (1 error)

The two child played. (1 error)

(4) If uncertain whether the grammar is correct, and if idiomatically acceptable, consider correct.

(5) If the first 50 words end in the middle of the sentence, disregard errors in the remainder of the sentence.

(6) No total errors in a given sentence should exceed half the number of words in the sentence. If the sentence is unintelligible, errors should equal one-half the sentence length, rounding off to the smaller whole number in case of an odd number of words in the sentence.

Eg., Girls boy the jump is the up. (3 errors)

Even though the entire composition appears unintelligible, repeat this operation for each sentence through the fiftieth word.

(7) If the fiftieth word should appear in the middle of the sentence, inspect the entire

sentence for intelligibility, then follow above procedure, basing sentence length on the number of words in the sentence up to and including the fiftieth word.

	Word	46	47	48	49	50	51	52	
Eg.,	Girls	boy	the	jump	is	/	~~the~~	~~up~~	- (2 errors)

(8) Total errors in the first 50 words are subtracted from fifty.

(9) The GCR is derived as follows:

$\frac{50\text{-no.grammatic errors}}{50} \times 100$, or (50-no.gram.errors) x 2.

Because of rule 6, no GCR will be less than 50.

(f) Spelling correctness ratio

Rules for establishing the spelling correctness ratio (SCR) are as follows:

(1) Count all words spelled incorrectly in the first 50 words, excluding suffixes on nouns and verbs. (eg., in "the boy is run," "run" would not be considered a spelling error but a grammatic error).

(2) Disregard blanks indicated by student.

(3) Do not count repeated spelling error more than once in the first 50 words.

(4) The SCR is derived as follows:

$\frac{50\text{-no. spelling errors}}{50} \times 100$, or (50-no. spelling errors)x2.

E. Statistical Procedures

Except where otherwise indicated, the following procedures were followed for examination of each of the ten samples.[1]

(1) The mean and standard deviation of the raw scores assigned by each teacher-judge to the sample were calculated, yielding three raw score means and three standard deviations for each sample.

(2) Intercorrelations of the raw scores assigned by the three teacher-judges to the variable age sample were calculated to provide an estimate of inter-judge agreement.

(3) Raw scores assigned by each teacher-judge were converted to T scores with a mean of 50 and a standard deviation of 15. The three raw scores on each composition were thereby converted to three T scores. The mean T score for each composition was then determined. These mean T scores were then normalized, again with a mean of 50 and a standard deviation of 15.[2]

---

[1]Extensive calculations were facilitated through the use of IBM 7070 and 7090 computer systems and appropriate programs.

[2]After normalization, standard deviations within the ten samples actually varied between 13.8 and 14.5 due to the properties of the T score and normal distribution.

(4) Normalized T Scores of males and females were tested for significant differences by means of the t test of significance of difference.

(5) The mean normalized T scores for the 50 compositions at each age level (age 10-18 years inclusive) within the variable age sample were calculated to determine the presence or absence of a developmental trend.

(6) After all objective analyses had been conducted, means and standard deviations for each were calculated for males, females, and the combined sexes. Male and female performances were compared by means of the t test of significance of difference. Comparisons were made on each variable within each sample.

(7) A seven variable intercorrelation matrix was developed for each sample. The seven variables consisted of the six objective variables and the normalized T based on teacher-judgments.

(8) Multiple correlations of the six objective variables with the normalized T were derived.

(9) Multiple-regression equations (Guilford, 1956, pp. 390-432), based on the six objective predictor variables and the normalized T score criterion

(teacher-judgments) were derived. Standard errors of estimate were calculated for each of ten equations.

(10) The three objective variables which collectively contributed to minimum standard error of estimate of the criterion were isolated, and ten multiple-regression equations (one variable age equation and nine constant age equations) derived. Each equation was based on the same three predictor variables and the normalized T criterion.

III. Results

A. Teacher-judgments of Quality of Language

1. Inter-judge agreement

As was expected, the means and standard deviations of raw scores assigned to each of the ten samples differed considerably among the three teacher-judges. Although these differences were not critical because of the statistical conversions of scores to T scores, they nevertheless illustrate the absolute differences among scores assigned to compositions by different teachers scoring compositions subjectively.

Table 10 reveals the means and standard deviations of raw scores of the compositions of the variable age sample as assigned by the three judges.

Table 10. Means and standard deviations of raw scores assigned by three judges to variable age sample[1]

		Judge A		Judge B		Judge C	
Variable age sample	n	M	S.D.	M	S.D.	M	S.D.
	450	37.16	26.06	50.05	26.33	52.49	17.85

[1]No inferences should be drawn about quality of language from this table.

Inspection reveals that judge A assigned lower scores than judges B and C. The dispersion of scores for judge C, as indicated by standard deviation, was less than that of the other two judges. These findings were consistent for scoring on all ten samples.

Pearson product moment correlation coefficients of raw scores assigned by each of the three judges to the variable age sample are indicated in Table 11.

Table 11. Correlations among scores assigned by three teacher-judges to the variable age sample.

Sample size	Correlations between raw scores of judges Judges A and B	A and C	B and C
450 subjects	.708	.653	.872

Judges B and C revealed substantially greater agreement in scoring compositions than did judges A and B, or A and C. When a normalized T score based on the three independent judgments was derived, the correlations of the raw scores of the three judges A, B, and C with the single normalized T became .855, .943, and .922 respectively. These correlations are higher than those indicated in Table 11, since raw scores of each of the judges were reflected in the normalized T scores.

Higher inter-judge reliability (and consequent lower standard errors of estimate in the multiple-

regression equations) would have been obtained if the scores assigned by judge A had been deleted. However, high inter-judge reliability could not be assumed to reflect high external validity.

2. Sex differences among subjects and judgments of quality

As stated in the preceding chapter, the raw scores assigned by each judge to each sample were converted to T scores, a mean for each sample being established at 50, with the standard deviation of the distribution, 15. Each composition in each sample received three T scores, one score derived from the raw score of each judge. The three T scores for each composition in each sample were combined and normalized, again around a mean of 50 and a standard deviation of 15. Accordingly, each composition in each sample received a single normalized T score.

The normalized T scores of males and females in each sample were tested for significance of difference by means of a t test for correlated means. Table 12 reveals the mean T scores and standard deviations for the males and females in each of the ten samples, with the results of administration of the t test.

Table 12. Differences between normalized T scores of males and females in ten samples

Sample	Male Mean	Male S.D.	Female Mean	Female S.D.	t
Variable age[1]	47.45	(13.36)	52.67	(13.75)	4.07[3]
10 years[2]	47.00	(13.85)	53.32	(13.67)	2.27[4]
11 years[2]	49.26	(15.57)	51.12	(12.29)	n.s.
12 years[2]	47.88	(14.81)	52.40	(13.06)	n.s.
13 years[2]	45.72	(14.04)	54.60	(12.73)	3.28[3]
14 years[2]	47.82	(14.72)	53.40	(12.95)	1.99[4]
15 years[2]	45.67	(14.54)	54.96	(12.67)	3.35[3]
16 years[2]	46.69	(13.67)	53.70	(13.54)	2.56[4]
17 years[2]	48.74	(12.64)	51.58	(15.32)	n.s.
18 years[2]	50.48	(13.37)	49.86	(14.85)	n.s.

[1] (225 males, 225 females)

[2] (50 males, 50 females)

[3] $p < .01$

[4] $p < .05$

Females in the variable age distribution received significantly higher composition scores than did males (significant at the 1 per cent level of confidence). Similarly, the females within the 10, 13, 14, 15 and 16 constant age distributions received significantly higher composition scores from judges (significant at the 1 or 5 per cent levels of confidence).

3. Age differences and quality of language

Differences in quality of compositions as reflected in teacher-judgments were examined through an inspection of normalized T scores of the 50 subjects at each age (10-18 yrs.) in the variable age distribution. The reader is reminded that when judges scored compositions in this sample, they were unaware of the ages of the 450 subjects except that they had been informed of the general constitution of the sample. Figure 1 indicates the mean normalized T score of the 50 subjects (25 male, 25 female, at each age level between 10 and 18 years inclusive.)

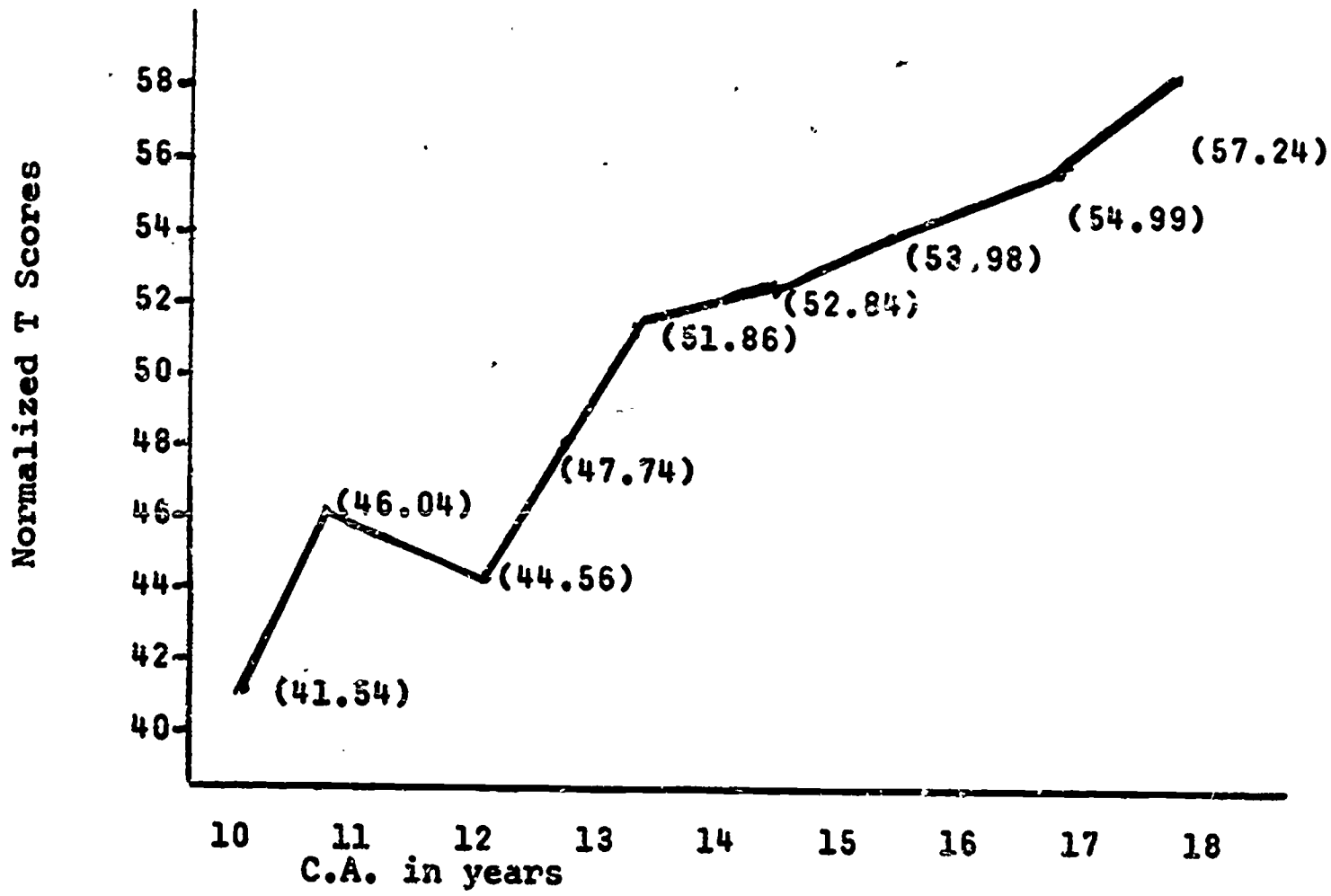


Figure 1. Mean normalized T scores at each age level 10-18 years inclusive

It is noted that mean normalized T scores increased consistently with age except at age 12 where the mean was lower than the mean at age 11. Increase in mean score was less apparent between ages 14 and 17 than at other ages. The correlation of age with T score was .367.

B. Objective Descriptions

Unlike the normalized T distributions reflecting teacher judgments, the constant age samples may be compared directly with regard to the six objective variables.

1. Composition length

The reader is referred back to "Objective analyses" for derivation of composition length.

Table 13 reveals the mean composition length within the ten samples, and differences in length between the compositions of males and females.

Table 13. Mean length of compositions of males and females 10 through 18 years of age.

Sample	Male Mean	Male S.D.	Female Mean	Female S.D.	Total Mean	Total S.D.
Variable age[1]	195.53	(115.31)	208.08	(99.59)	201.80	(108.04)
10 years[2]	130.22	(85.49)	183.04	(106.64)	156.63	(100.69)
11 years[2]	159.54	(97.95)	198.24	(110.07)	178.89	(106.50)
12 years[2]	160.18	(97.62)	187.90	(106.48)	174.04	(103.60)
13 years[2]	164.72	(97.83)	207.10	(82.57)	185.91	(93.43)
14 years[2]	182.34	(97.89)	209.66	(100.59)	196.00	(100.69)
15 years[2]	198.04	(99.30)	232.14	(119.29)	215.26	(111.73)
16 years[2]	210.75	(91.64)	238.48	(115.64)	224.47	(105.66)
17 years[2]	234.18	(108.27)	228.60	(98.11)	231.39	(103.87)
18 years[2]	256.52	(110.41)	208.66	(102.53)	232.59	(109.74)

[1] (225 males, 225 females)

[2] (50 males, 50 females)

Although the trend is evident, the difference between composition lengths of males and females within the variable age sample was not statistically significant. Within the age 10 sample, the females wrote significantly longer compositions than males ($t = 2.71$, significant at the 1 per cent confidence level) and similarly at age 13 ($t = 2.32$, significant

at the 5 per cent confidence level). However, at age 18, males wrote significantly longer compositions than females (t=2.22, significant at the 5 per cent level of confidence).

The increase in composition length with age is shown graphically in Figure 2, reflecting the progressive increase in composition length from a mean of 156.63 words at age 10 to 232.59 words at age 18. It is notable that a small reduction in composition length occurred between the age 11 year sample and age 12 year sample. This was noted also in Figure 1, teacher-judgments.

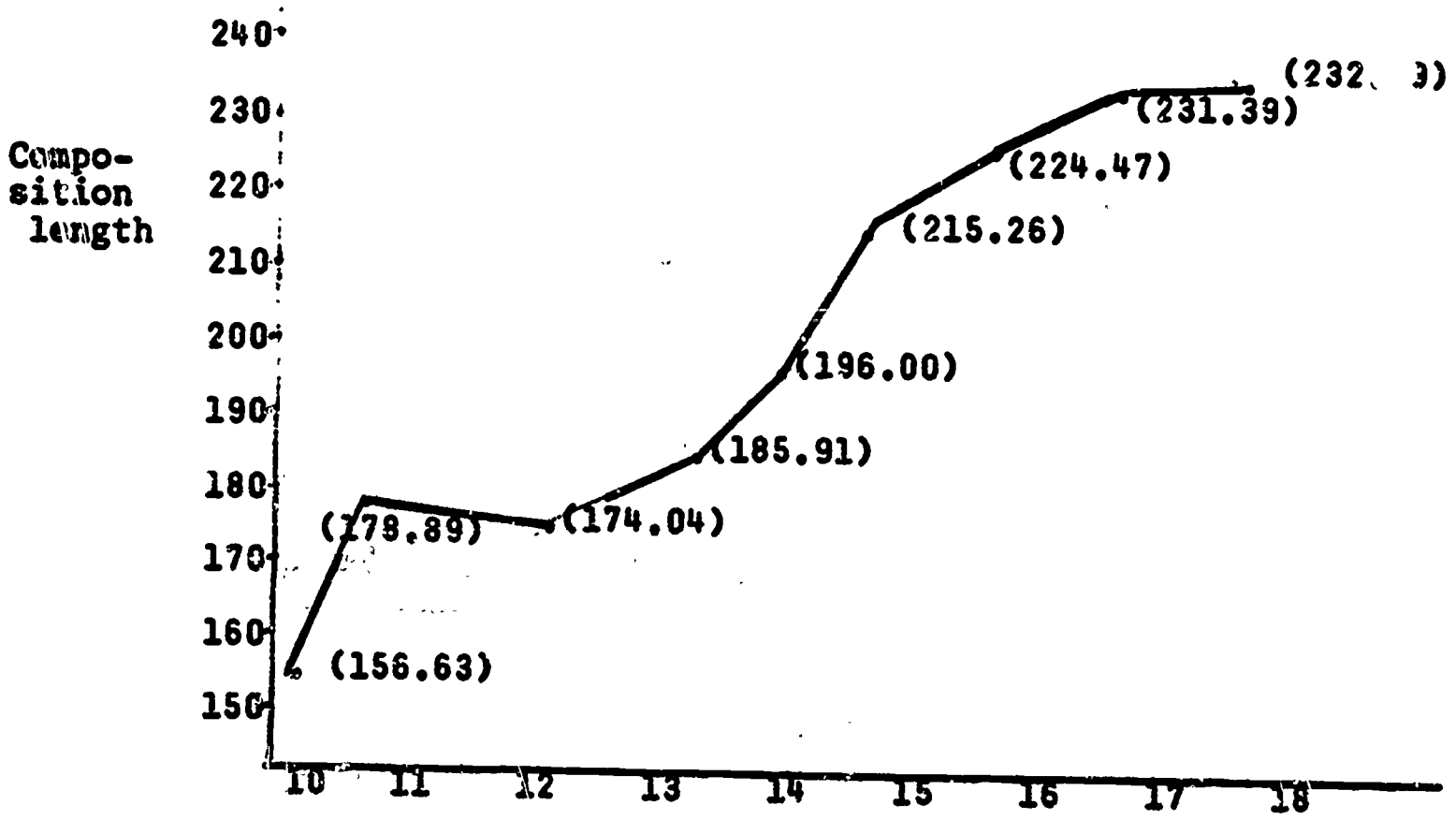


Figure 2. Mean composition length at each age level 10-18 years inclusive.

The correlation of composition length and age within the variable age sample was .237.

2. Sentence length

The reader is referred back to "Objective Analyses" for derivation of sentence length.

Males and females in each of the ten samples were compared with regard to the mean length of sentences in their compositions. No statistically significant differences were found between these two subgroups in any of the ten samples upon administration of the t test.

Table 14. Mean sentence length of males and females 10 through 18 years of age.

Sample	Male		Female		Total	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
Variable age[1]	8.20	(3.45)	8.40	(3.13)	8.30	(3.30)
10 years[2]	7.06	(2.46)	7.64	(3.15)	7.35	(2.86)
11 years[2]	6.99	(3.42)	6.73	(1.38)	6.86	(2.63)
12 years[2]	7.48	(2.88)	6.95	(2.40)	7.21	(2.68)
13 years[2]	7.30	(2.91)	7.87	(2.22)	7.58	(2.61)
14 years[2]	8.11	(4.90)	8.23	(2.57)	8.17	(3.94)
15 years[2]	8.05	(2.66)	8.57	(2.71)	8.31	(2.71)
16 years[2]	8.81	(3.78)	9.41	(3.52)	9.10	(3.68)
17 years[2]	9.80	(3.39)	10.54	(3.81)	10.17	(3.64)
18 years[2]	10.77	(3.84)	10.54	(3.42)	10.66	(3.66)

[1](225 males, 225 females

[2](50 males, 50 females

The 10 year sample revealed a greater mean sentence length than either the 11 year or the 12 year samples as shown in Figure 3. However, a developmental trend was noticed.

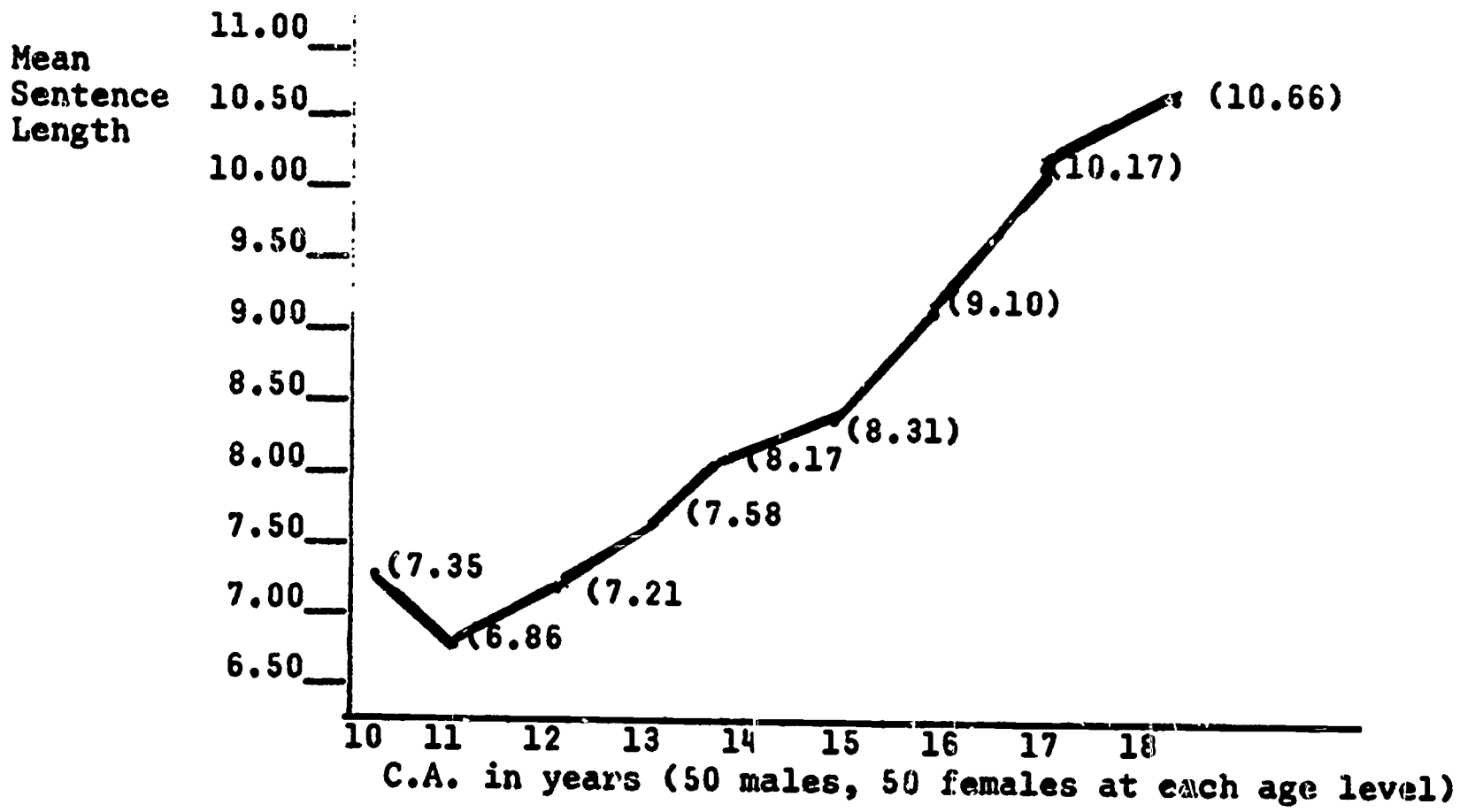


Figure 3. Mean sentence length at each age level 10-18 years inclusive

The correlation of mean sentence length and age within the variable age sample was .342.

3. Word structure ratio

The reader is referred back to "Objective analyses" for derivation of word structure ratio.

Males and females in each sample were compared with regard to the means of word structure ratios. Females within the variable age sample received significantly higher word structure ratios than males (t = 3.23, significant at the 1 per cent level of confidence). Statistically significant differences favoring females also were found within the 10 and 15 year samples (significant at the 5 per cent level of confidence, with t = 2.31 and 2.19 for the two respective samples).

Table 15. Mean word structure ratios of males and females 10 through 18 years of age

Sample	Males		Females		Total	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
Variable age[1]	38.40	(8.99)	41.05	(8.34)	39.72	(8.78)
10 years[2]	35.12	(9.16)	39.32	(8.87)	37.22	(9.30)
11 years[2]	36.76	(10.24)	37.80	(10.12)	37.28	(10.25)
12 years[2]	38.04	(9.97)	40.04	(8.15)	39.04	(9.21)
13 years[2]	39.48	(12.29)	41.36	(7.95)	40.42	(10.44)
14 years[2]	39.56	(9.02)	42.28	(6.68	40.92	(8.09)
15 years[2]	39.31	(7.25)	42.24	(5.89)	40.79	(6.79)
16 years[2]	39.73	(8.05)	41.36	(7.71)	40.53	(7.96)
17 years[2]	39.56	(7.23)	41.26	(8.60)	40.41	(8.03)
18 years[2]	42.80	(6.20)	43.32	(6.91)	43.06	(6.60)

[1](225 males, 225 females)

[2](50 males, 50 females)

As shown in Figure 4, a small but upward trend is noticed with increasing age, suggesting increased use of Class III, Class IV, and function words as students advance in age between 10 and 18 years of age. Little change was found between 13 years and 17 years.

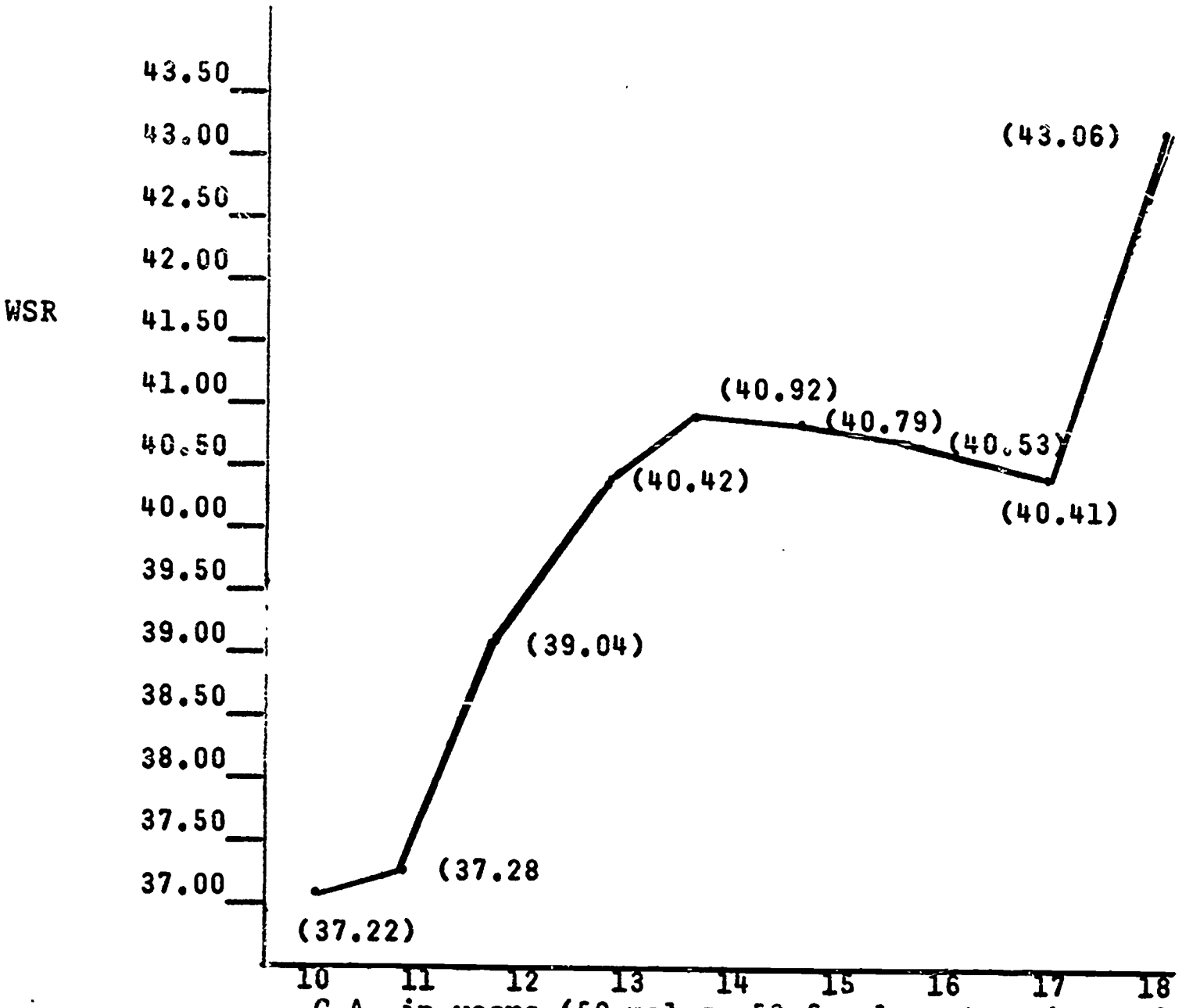


Figure 4. Mean Word structure ratio at each age level 10-18 years inclusive

The correlation of ~~mean sentence length~~ word structure ratio and age within the variable age sample was .199.

4. Type-token ratio

The reader is referred back to "Objective Analyses" for derivation of type-token ratio.

Table 16 indicates the means and standard deviations of type-token ratios for males and females in each of the ten samples. Within the variable age

sample, females yielded a significantly higher type-token ratio than males (t=2.52, statistically significant at the 5 per cent level of confidence). At only one specific age level, the age 15 sample, was this tendency also observed (t=2.37, statistically significant at the 5 per cent level of confidence).

Table 16. Mean type-token ratios of males and females 10 through 18 years of age

Sample	Male Mean	Male S.D.	Female Mean	Female S.D.	Total Mean	Total S.D.
Variable age[1]	62.68	(10.92)	65.34	(11.34)	64.01	(11.22)
10 years[2]	55.48	(10.07)	59.36	(10.05)	57.42	(10.30)
11 years[2]	58.12	(12.87)	58.28	(11.21)	58.20	(12.13)
12 years[2]	60.84	(9.24)	59.60	(10.96)	60.22	(10.21)
13 years[2]	59.00	(11.19)	63.60	(12.55)	61.30	(12.17)
14 years[2]	64.64	(9.73)	66.20	(10.30)	65.42	(10.10)
15 years[2]	64.49	(8.80)	68.82	(9.71)	66.68	(9.31)
16 years[2]	66.31	(8.46)	68.76	(8.10)	67.52	(8.41)
17 years[2]	68.64	(6.55)	70.48	(8.07)	69.56	(7.44)
18 years[2]	70.48	(8.96)	70.24	(8.92)	70.36	(8.99)

[1] (225 males, 225 females)

[2] (50 males, 50 females)

Inspection of the mean type-token ratios for the full 100 subjects in each constant age sample reveals a consistent increase in type-token ratio from 10 through 18 years. This is depicted in Figure 5.

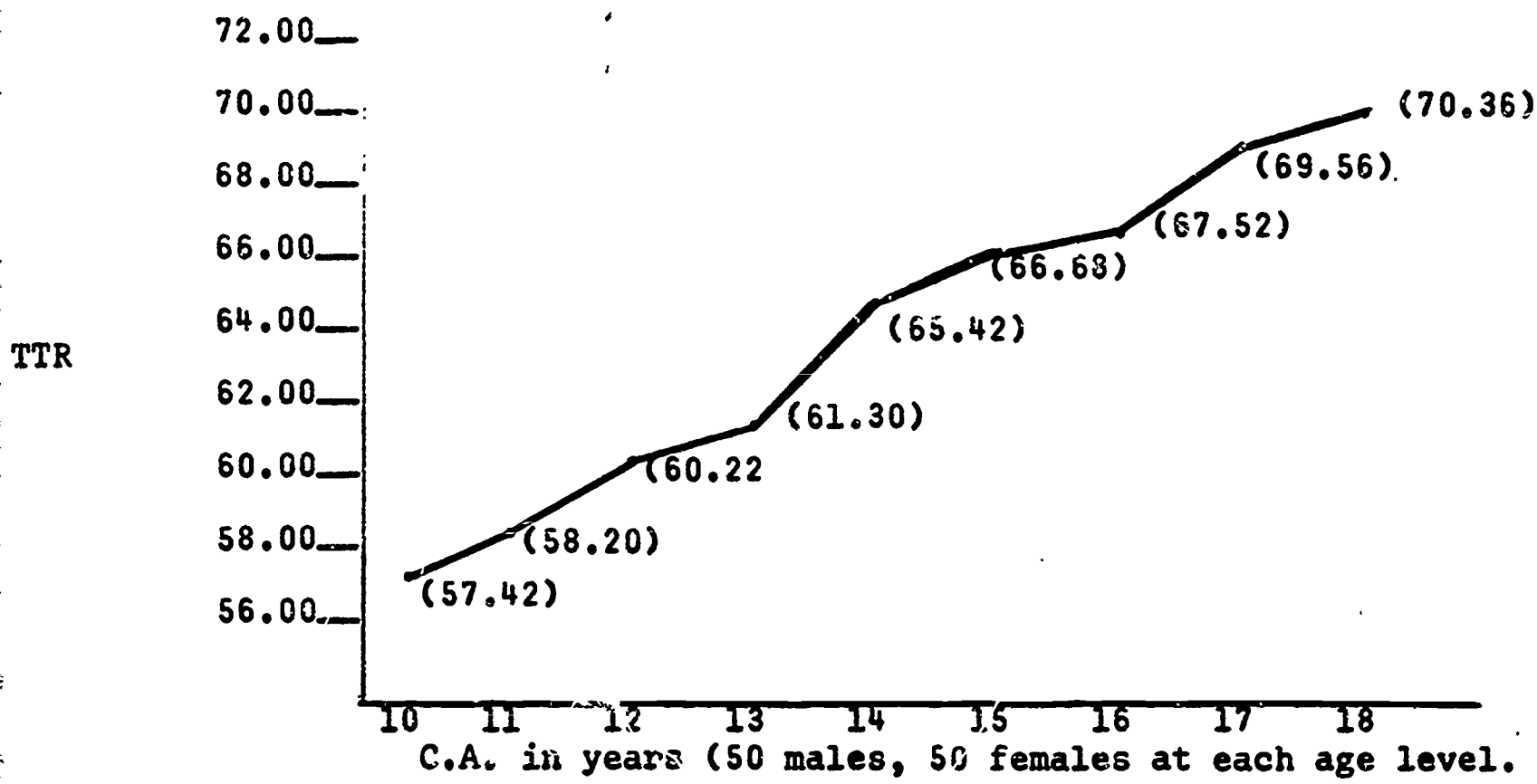


Figure 5. Mean type-token ratio at each age level 10-18 years inclusive.

The correlation of type-token ratio and age within the variable age sample was ,399.

5. Grammatic correctness ratio

The reader is referred back to "Objective analyses" for derivation of the grammatic correctness ratio.

Table 17 reveals the means and standard deviations of grammatic correctness ratios for males and females in the ten samples. Females received significantly higher ratios than males within the variable age sample (t=2.45, statistically significant at the 5 per cent level of confidence). This was found also within the age 15 sample (t=2.50, statistically

significant at the 5 per cent level of confidence). However, at the 18 year level, males received significantly higher ratios than females (t=2.18, statistically significant at the 5 per cent level of confidence).

Table 17. Mean grammatic correctness ratios of males and females 10 through 18 years of age.

Sample	Male Mean	S.D.	Female Mean	S.D.	Total Mean	S.D.
Variable age[1]	78.01	(12.18)	80.81	(11.95)	79.41	(12.16)
10 years[2]	77.72	(12.71)	80.12	(13.53)	78.92	(13.25)
11 years[2]	77.08	(14.04)	75.22	(13.36)	76.15	(13.81)
12 years[2]	75.24	(12.92)	78.48	(10.77)	76.86	(12.07)
13 years[2]	75.76	(13.25)	79.52	(10.76)	77.64	(12.27)
14 years[2]	78.12	(11.17)	82.36	(11.97)	80.24	(11.83)
15 years[2]	78.98	(10.29)	84.28	(10.57)	81.66	(10.82)
16 years[2]	82.63	(10.48)	85.40	(11.00)	84.00	(10.88)
17 years[2]	80.90	(12.52)	84.68	(12.25)	82.79	(12.59)
18 years[2]	86.60	(10.36)	81.56	(12.45)	84.08	(11.79)

[1] (225 males, 225 females)

[2] (50 males, 50 females)

Figure 6 reveals a trend for grammatic correctness ratio to increase with the age of the subjects comprising the nine constant age samples. It is notable that there is a relatively minor change

in grammatic correctness ratio from 10 through 18 years, 10 year old subjects tending to make approximately 11 grammatic errors in 50 words (21.02 per cent) compared with approximately 8 errors per 50 words (15.92 per cent) for 18 year old subjects.

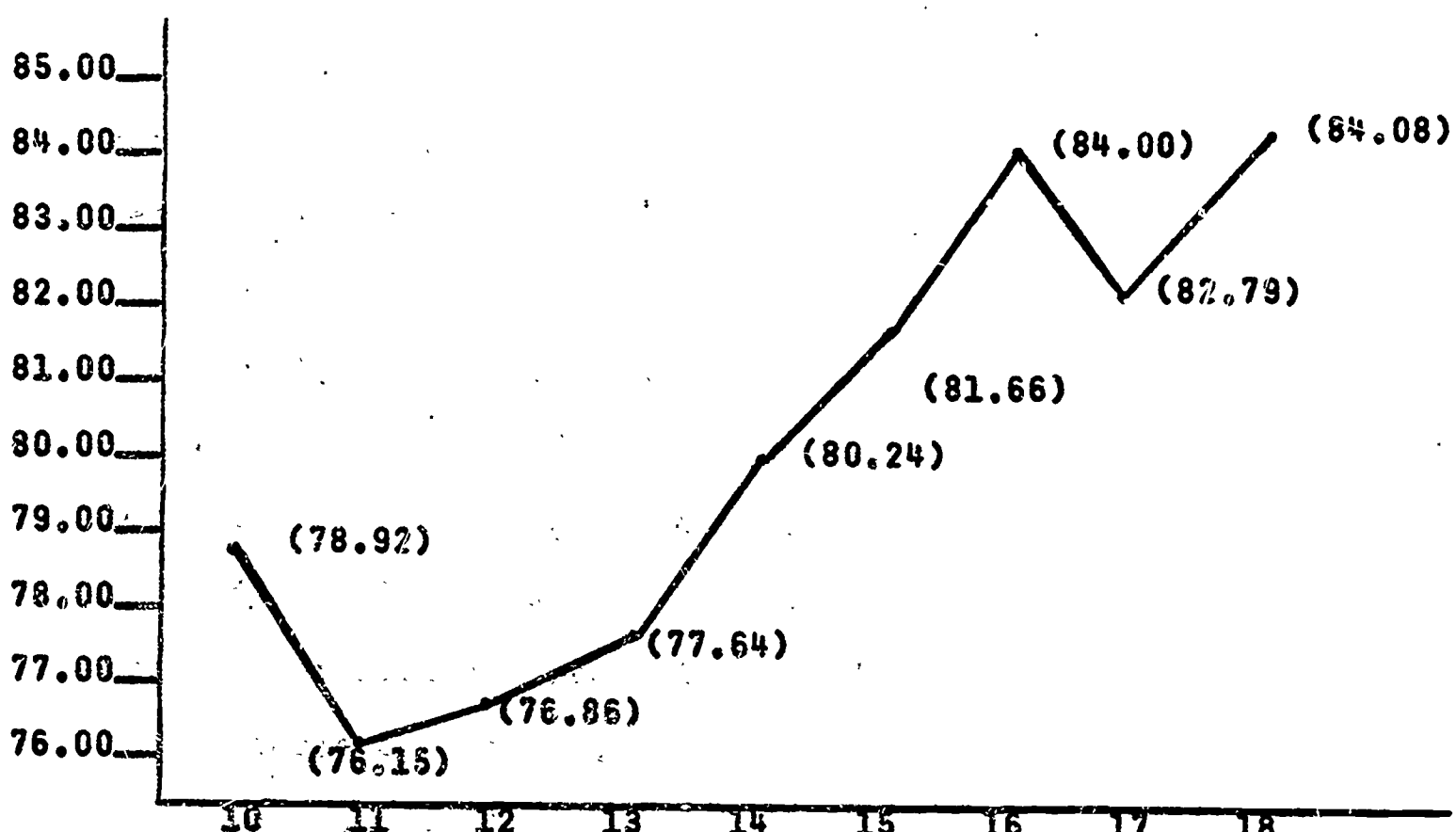


Figure 6. Mean grammatic correctness ratio at each age level 10-18 years inclusive

The correlation of grammatic correctness ratio and age within the variable age sample was .178.

6. Spelling correctness ratio

The reader is again referred back to "Objective analyses" for derivation of this ratio.

Approximately three per cent of the first 50 words expressed by subjects were misspelled. Table 18 reveals the means and standard deviations of SCR for males and females in 10 samples. Within the variable age sample, females received higher spelling correctness scores (t=2.03, statistically significant at the 5 per cent level of confidence). Higher scores were received by females also within the 13 year and 14 year samples, (t=2.07 and 2.13 for the respective samples, both statistically significant at the 5 per cent level of confidence). It should be added, however, that true differences are probably negligible (e.g., within the variable age sample, the mean for males was 97.63, and for females, 98.13).

Table 18. Mean spelling correctness ratios of males and females 10 through 18 years of age

Sample	Male Mean	Male S.D.	Female Mean	Female S.D.	Total Mean	Total S.D.
Variable age[1]	97.63	(2.78)	98.13	(2.49)	97.88	(2.65)
10 years[2]	96.96	(3.03)	97.70	(3.12)	97.08	(3.09)
11 years[2]	97.16	(3.78)	97.84	(2.91)	97.50	(3.40)
12 years[2]	97.88	(2.24)	98.20	(2.47)	98.04	(2.38)
13 years[2]	96.96	(3.13)	98.04	(1.90)	97.50	(2.66)
14 years[2]	98.16	(2.15)	98.96	(1.51)	98.56	(1.91)
15 years[2]	97.96	(2.27)	98.40	(2.33)	98.18	(2.32)
16 years[2]	98.24	(2.30)	98.40	(1.92)	98.32	(2.13)
17 years[2]	97.44	(3.42)	98.32	(2.75)	97.88	(3.15)
18 years[2]	98.20	(1.89)	98.36	(1.91)	98.28	(1.91)

[1] (225 males, 225 females)

[2] (50 males, 50 females)

Inspection of means in Table 18 suggests little trend in spelling correctness as a function of chronological age. The correlation of age and spelling correctness within the variable age sample was .190, statistically significant but slight.

C. Relationships among Analyses and Teacher-judgments

1. Intercorrelations among objective variables

Table 19 and Table 20 report on six variables. Their identification is as follows:

Variable 1 - composition length
Variable 2 - mean sentence length
Variable 3 - word structure ratio
Variable 4 - type-token ratio
Variable 5 - grammatic correctness ratio
Variable 6 - spelling correctness ratio

Means and standard deviations of scores on each variable within the variable age sample are reported in Table 19.

Table 19. Means and standard deviations of scores on six objective variables within variable age sample

Variable	Mean	S.D.
1	201.80	108.04
2	8.30	3.30
3	39.72	8.78
4	64.01	11.22
5	79.41	12.16
6	97.88	2.65

Table 20 indicates intercorrelations among the six objectively derived descriptions of composition of each subject in the variable age sample of 450 subjects.

Table 20. Intercorrelation matrix based on six objective variables within variable age sample[1],[2]

Variable	2	3	4	5	6
1 (CL)	.333	.279	.335	.191	.135
2 (MSL)		.356	.307	.174	n.s.
3 (WSR)			.148	.386	.127
4 (TTR)				.248	n.s.
5 (GCR)					.190
6 (SCR)					—

[1] (225 males, 225 females)

[2] with df=448, r (.05)=.093, r(.01)=.122

With the exception of mean sentence length and spelling correctness, and type-token ratio and spelling correctness, all variables reveal statistically significant positive relationships (statistically significant at the 1 per cent level of confidence).

2. Correlations of objective variables and teacher judgment

Scores on the six objective variables for each subject were tested for correlation with the

normalized T score derived from teacher-judgments. Table 21 indicates correlations based upon scores within the variable age sample of 450 subjects from 10 through 18 years.

Table 21. Correlations between six objective variables and teacher-judgment within variable age sample.

Variable	Teacher-judgment[1]
1. (CL)	.389[2]
2. (MSL)	.299[2]
3. (WSR)	.408[2]
4. (TTR)	.537[2]
5. (GCR)	.715[2]
6. (SCR)	.227[2]

[1] Normalized T score

[2] $p < .01$

All objective scores are seen to correlate positively with the combined subjective evaluations of the three teacher-judges.

The correlation of .715 between variable 5 and teacher-judgments may be interpreted as indicating that approximately 55 per cent of the

variance in teacher-judgment is predictable grammatic correctness. Approximately 29 per cent of the teacher-judgments variance is predictable from type-token ratio. Approximately 17 per cent is predictable from the word structure ratio, and 15 per cent from composition length. The reader is reminded that these percentages are not directly additive.

Table 22 indicates the correlation of the objective variables with teacher-judgment within each of the nine constant age samples.

Table 22. Correlations between six objective variables and teacher-judgment within each of nine constant age samples.

Sample	Objective variable[1] 1	2	3	4	5	6
10 years	.457[2]	.152	.479	.505	.596	.437
11 years	.429	.202	.446	.636	.759	.336
12 years	.364	.113	.469	.463	.686	.330
13 years	.501	.304	.126	.511	.646	.167
14 years	.430	.296	.440	.514	.763	.168
15 years	.489	.414	.378	.636	.758	.365
16 years	.478	.345	.334	.381	.723	.110
17 years	.218	.381	.355	.444	.648	.277
18 years	.209	.337	.387	.322	.791	.406

[1] See identification immediately preceding Table 19

[2] with df=98, r(.05)=.196, r(.01)=.255

D. Estimation of Teacher-judgment from Objective Description

A major objective of this investigation was the development of a system for estimating quality of written language as reflected in teacher-judgments, from objective description of this language.

In the derivation of a multiple-regression equation which yields this estimate, optimum prediction must be balanced against efficiency in the use of the instrument.

From an intercorrelation matrix for each of the ten samples, multiple-regression equations were developed to yield maximum multiple correlation and minimum standard error of estimate of the criterion variable, teacher-judgment.

Table 23 is a summary indicating for each sample the order in which descriptive variables contributed to the multiple correlation of the predictor variables with the criterion, teacher-judgment. When the addition of certain variables no longer served to reduce the standard error of estimate, this is indicated by the absence of that variable in the table. The maximum multiple correlation is indicated, as is the standard error of estimate based on all descriptive variables employed in the equation.

Variables as indicated by number are:

Variable 1 - composition lehgth
Variable 2 - mean sentence length
Variable 3 - word structure ratio
Variable 4 - type-token ratio
Variable 5 - grammatic correctness ratio
Variable 6 - spelling correctness ratio

Table 23. Objective variables contributing to maximum multiple correlation with teacher-judgment.

Sample	Variables	Multiple correlation with criterion	Standard error[1] of estimate
Variable age	5,4,1,3,6,5	.827	7.820
10 year	5,4,6,1,3,2	.856	7.576
11 year	5,4,6,2	.830	8.034
12 year	5,4,3,6,1,2	.821	8.382
13 year	5,4,1,6,2,3	.806	8.659
14 year	5,4,1,3,6,2	.845	7.832
15 year	5,4,6,2,1	.860	7.577
16 year	5,1,4,6,3,2	.811	8.520
17 year	5,4,2,1,3,6	.739	9.870
18 year	5,4,6,3,1,2	.827	8.238

[1] Standard deviation of criteria, 15

Inspection of Table 23 indicates that variable 5, the grammatic correctness ratio, without exception is the primary variable contributing to prediction of the criterion, followed by variable 4, the type-token ratio.

With variables 5 and 4 partialled out, variable 1, composition length, tended to contribute more heavily to most of the multiple-regression equations than did other variables. This variable also lends itself to easy calculation.

Accordingly, with variables 2, 3 and 6 suppressed, multiple-regression equations were developed for each sample, incorporating composition length, type-token ratio, and grammatic correctness ratio. Since multiple-regression equations on the variable age sample, the 13 year age sample, the 14 year age sample, and the 16 year age sample already reflected these three variables as the primary contributors to the multiple correlation (see Table 23), it was not necessary to recompute multiple-regression equations on these four samples.

Table 24 indicates the sample means and standard deviations for the three selected predictors (indicated earlier in Tables 13, 16, and 17).

Table 24. Means and standard deviations for three predictor variables.

Sample	C.L. Mean	C.L. S.D.	TTR Mean	TTR S.D.	GCR Mean	GCR S.D.
Variable age[1]	201.80	(108.04)	64.01	(11.22)	79.41	(12.16)
10 years[2]	156.63	(100.69)	57.42	(10.30)	78.92	(13.25)
11 years[2]	178.89	(106.50)	58.20	(12.13)	76.15	(13.81)
12 years[2]	174.04	(103.60)	60.22	(10.21)	76.86	(12.07)
13 years[2]	185.91	(93.43)	61.30	(12.17)	77.64	(12.27)
14 years[2]	196.00	(100.69)	65.42	(10.10)	80.24	(11.83)
15 years[2]	215.26	(111.73)	66.68	(9.31)	81.66	(10.82)
16 years[2]	224.47	(105.66)	67.52	(8.41)	84.00	(10.88)
17 years[2]	231.39	(103.87)	69.56	(7.44)	82.79	(12.59)
18 years[2]	232.59	(109.74)	70.36	(8.99)	84.08	(11.79)

[1] n=450

[2] n=100

Table 25 indicates the multiple correlation of the three predictor variables with the criterion (teacher-judgment), and the standard errors of estimate of the normalized T score (M=50, S.D.=15), for each of ten samples.

Table 25. Multiple correlations of three predictors with the criterion, and standard errors of estimate.

Sample	Multiple r	Standard error of estimate
Variable age	.819	7.952
10 years	.794	8.783
11 years	.824	8.093
12 years	.780	9.025
13 years	.792	8.801
14 years	.837	7.887
15 years	.842	7.932
16 years	.809	8.429
17 years	.725	9.921
18 years	.814	8.371

Table 26 indicates the beta weights and constant to be attached to the predictors to arrive at the multiple-regression equations indicated in Table 27.

Table 26. Beta weights and constants for estimates of teacher-judgment.

Sample	Beta weight Attached to variable CL	TTR	GCR	Constant
Variable age	.021	.410	.683	-34.621
10 years	.036	.538	.560	-30.806
11 years	.000	.420	.605	-20.317
12 years	.020	.445	.700	-33.984
13 years	.039	.382	.594	-26.552
14 years	.031	.363	.747	-39.168
15 years	.020	.513	.747	-49.169
16 years	.035	.359	.799	-49.191
17 years	.027	.471	.648	-42.580
18 years	.007	.287	.905	-47.675

Beta weights and constants for each distribution as indicated in Table 26 lead directly to the multiple-regression equations indicated in Table 27.

Table 27. Multiple-regression equations with three predictors leading to estimated normalized T score [1,2,3]

Distribution	Multiple-regression equation
Variable age	.021(CL)+.410(TTR)+.683(GCR)-34.621=T
10 years	.036(CL)+.538(TTR)+.560(GCR)-30.806=T
11 years	.0(CL)+.420(TTR)+.605(GCR)-20.317=T
12 years	.02(CL)+.445(TTR)+.7(GCR)-33.984=T
13 years	.039(CL)+.382(TTR)+.594(GCR)-26.552=T
14 years	.031(CL)+.363(TTR)+.747(GCR)-39.168=T
15 years	.020(CL)+.513(TTR)+.747(GCR)-49.169=T
16 years	.035(CL)+.359(TTR)+.799(GCR)-49.191=T
17 years	.027(CL)+.471(TTR)+.648(GCR)-42.580=T
18 years	.007(CL)+.287 (TTR)+.905(GCR)-47.675=T

[1] Reflecting teacher-judgment of quality of written language

[2] See Appendix C for standard score and percentile equivalents of normalized T scores

In calculating the normalized T score from these equations, each sub-calculation should be rounded at no less than three decimal places. After the constant is subtracted, the normalized T score may be rounded to the nearest whole number.

A table has been provided in Appendix C for convenience in converting T scores based on a mean of 50 and a standard deviation of 15 into standard scores and percentiles.

[3] See Appendices D, E, F, for samples

While the table of standard score and percentile equivalents of normalized T scores provided in Appendix C should be satisfactory for most purposes, it is based on standard scores carried to only two decimal places. Furthermore, the table is based on a normalized T distribution of 50, and a standard deviation of 15.

For reasons indicated earlier, the mean normalized T in the ten samples varied from 50.06 to 50.61, and the true standard deviations varied from 13.82 to 14.47.

Table 28 is provided for the reader who wishes to calculate standard scores of maximum accuracy from the T scores estimated from the equations.

Table 28. Mean T scores and standard deviations in ten samples.

Sample	T score	
	Mean	S.D.
Variable age	50.06	13.82
10 years	50.16	14.19
11 years	50.19	14.13
12 years	50.14	14.21
13 years	50.16	14.19
14 years	50.61	14.21
15 years	50.36	14.47
16 years	50.16	14.12
17 years	50.16	14.19
18 years	50.17	14.21

An alternative to use of the table in Appendix C is as follows:

(1) Derive estimated T score (rounded at two decimal places) from the appropriate multiple-regression equation (Table 27).

(2) Subtract this figure from the mean T score for the appropriate distribution indicated in Table 28, or subtract the mean T score from this figure if the figure is greater than the distribution's mean T score.

(3) Divide the resulting figure by the appropriate standard deviation provided in Table 28.

(4) Retain the sign (positive or negative). The resulting figure is the standard score of the composition.

IV. Discussion

A. Teacher-judgments of Quality of Language

1. Reliability

As expected, the means and spread of raw scores assigned by the three teacher-judges varied considerably. The mean raw scores assigned to compositions in the variable age sample by the three teacher-judges varied between 37.16 and 52.49, and standard deviations varied between 17.85 and 26.33. However, the correlations between the scores assigned by the judges were .708, .653, and .872, reflecting substantial agreement, particularly between two of the three judges. In retrospect, it becomes apparent that greater reliability of teacher-judgment would have been obtained if the judgments only of judges B and C, whose raw scores correlated so high, had been used to establish the criterion. However, reliability was sacrificed in order to retain judgmental criteria reflected by judge A which may have received little weight on the part of the remaining two judges.

The mean correlation among the three judgments was .744, substantially higher than the correlation of .41 reported by French (1962) among English teachers without predetermined standard criteria. This may be due in part to the age range

of nine years within the variable age sample, the effect being to spread quality of composition and to facilitate comparisons of compositions. Nevertheless, it is suggested that the relatively high teacher-judge agreement was due also to readily apparent standards such as grammatic correctness. This contention is supported by the fact that grammatic correctness correlated quite high with teacher-judgments (.715 within the variable age distribution).

Reliability of teacher-judgment (the criterion) should be reflected in the multiple correlation of the predictor variables with the criterion. The multiple correlation of three predictors with the criterion within the variable age sample was .819, and in fact increased within three of the 9 constant age samples (.824, .837, .842). This is surprising in view of the fact that within the restricted age ranges of the constant age samples, it might be anticipated that differences in grammar would be restricted, leading to lower correlations.

2. Sex differences and quality of language

When the normalized T scores of compositions written by the 225 males and the 225 females in the variable age sample were compared (Table 12), the scores obtained by females were observed to be

significantly higher than those of the males (1 per cent confidence level). The mean normalized T score of females was significantly higher than that of males in five of the nine constant age samples. The extent of the difference between males and females within the variable age sample is illustrated by the fact that the Z score equivalent of the mean male T score (47.45) and of the mean female T score (52.67), represented a difference of approximately 14 percentile units.

It is notable that no statistically significant differences between males and females were found within the 11, 12, 17 and 18 year constant age samples. No explanation suggests itself for the 11 and 12 year samples, but it is quite possible that better female students tend to leave schools for the deaf at a younger age than male students of similar ability, thereby eliminating the superiority of females over males at the 17 and 18 year levels.

The observation of superiority of female students over male students in this linguistic area adds support to general research findings among non-handicapped children. It is remarkable

that differences favoring females should be found in spite of severe auditory deprivation.

3. Age and quality of language

An examination of Figure 1 (mean normalized T scores at nine age levels) discloses a continuous improvement in quality of written language through 18 years. A positive correlation of .367 was found between normalized T scores and age within the variable age distribution.

An exception is noted between 11 and 12 years. It may be that this dip is caused by major curricular modifications which take place around this period of development (eg., introduction of social studies and concomitant reduction in formal language instruction).

Of major importance is the fact that subjects continued to show gains at 17 and 18 years. This finding runs contrary to speculation that quality of written language tends to plateau at mid-adolescence, speculation which was formerly shared by the investigators (see "Implications").

The mean normalized T score (41.54) for 10 year old subjects fell at approximately the 30th percentile for the entire distribution while the mean normalized T score (57.24)

for 18 year old subjects fell at the 68th percentile (see Appendix C).

This information may be interpreted as indicating major differences in central tendency at ages 10 and 18. However, it may also be inferred statistically that 37 per cent of the compositions written by 10 year old subjects might be expected to be superior to the "average" compositions written by 11 year old subjects, 23 per cent of "average" compositions written by 14 year old subjects, and indeed 13 per cent of "average" compositions written by 18 year old subjects. This finding has implications for the class placement of deaf students.

B. Objective Descriptions

1. Composition length

Differences in mean composition length between the productions of males and females were not statistically significant within the variable age sample. Females wrote significantly longer compositions than males within the 10 and 13 year constant age samples, while males wrote significantly longer compositions at the 18 year level.

Comparison of the length of compositions

written by the 100 subjects in each constant age sample reveals a continuing increase in composition length with age between 10 and 18 years, except between ages 11 and 12. The finding is similar to that of the normalized T scores derived from teacher-judgments. It should be added that findings relative to composition length are based on a biased sample, particularly at the 10 and 11 year levels, resulting from the elimination of any compositions of less than 50 words.

Notable within all ten samples is the wide range in length, as revealed by the large standard deviations. The mean composition length within the variable age sample, for example, was 201.8 words, with a standard deviation of 108 words.

2. Sentence length

Mean sentence length within compositions written by males and females did not differ within any of the ten samples. As depicted in Figure 3, mean sentence length increased with age, except that 10 year old subjects wrote sentences of greater length than 11 and 12 year old subjects. This may reflect the bias introduced in excluding more compositions of less than 50 words at the

PRECEDING PAGE ...ssing

The type-token ratio increased consistently with age (r=.399); 57 per cent of the first fifty words written by 10 year old subjects were different words, compared with 70 per cent for 18 year old subjects.

5. Grammatic correctness ratio

Females within the variable age sample made significantly fewer grammatic errors than males. Compositions written by males within the variable age sample were 78 per cent correct (as measured by the GCR), and by females, 81 per cent correct. The difference remains small. It is noteworthy that the female superiority within the variable age sample was reversed within the 18 year constant age sample, where males made 5 per cent fewer errors than females. This finding is similar to that for composition length, males writing significantly longer compositions than females at 18 years.

There tended to be a minor reduction of grammatic errors with increasing age (r between GCR and age, .176). However, there were inconsistencies. Subjects at age 10 obtained a mean GCR of 78.9, greater than that of 11, 12, and 13 year old subjects. It may be that students around

age 11 begin to use more complex grammatic structures than at age 10, exposing their language to more grammatic errors. Grammatic correctness did not noticeably increase between 16 and 18 years, and indeed a dip was noted at age 17. This suggests that grammar undergoes no improvement beyond age 16, grammatic errors resisting correction by this age.

Gains in grammatic correctness between ages 10 and 18 appear slight (21 per cent error at age 10, 16 per cent at age 18). However, it should be stated again that as grammatic complexity increases (eg., reflected in increasing sentence length), likelihood of error is also increased.

6. Spelling correctness ratio

Few spelling errors were found in compositions within any of the samples. Within the variable age sample, only 2.1 per cent of words were misspelled. The female subjects within this sample made significantly less errors than males, but the difference between the mean ratios was only .5 per cent.

Improvement in spelling with age was significant ($r=.190$) but slight, 18 year old subjects making 1.2 per cent less errors than 10 year olds.

It is apparent that spelling (aside from morphemic considerations) constitutes no major problem to deaf students, probably because the deaf student depends less on phonetics than on rote learning to learn to spell a given word. It may be, however, that deaf students refrain from writing a word unless reasonably certain of its correct spelling.

7. Relationships among objective variables

Intercorrelations among the six objective variables were derived from the variable age sample, and indicated in Table 20. Of the 15 correlations derived, 13 were significant, and in a positive direction. No relationship was present between spelling correctness and sentence length. Similarly, no relationship was present between spelling correctness and type-token ratio. The latter lack of relationship is of interest insofar as one might speculate that increased variety of word usage might lead to increased spelling error. This apparently is not so.

The highest correlation was that between the word structure ratio and the grammatic correctness ratio (r=.386), followed by mean sentence

length and word structure ratio (r=.356). The correlation between composition length and type-token ratio was .335, and between composition length and sentence length, .333.

It is of interest to note that the highest correlation among any two of the six objective variables is that of word structure ratio and grammatic correctness ratio. Students who tend toward relatively complex sentence structures as suggested by a high word structure ratio (eg., free use of adjectives, adverbs, conjunctions, prepositions, etc.) tend also to use these parts of speech correctly.

As anticipated, sentence length is related in part to extensive use of words other than Class I and Class II words (nouns and verbs). As the child begins to introduce adjectives, adverbs, prepositions, conjunctions, etc., the length of his sentence also increases. This is an expected finding (eg., with the addition of conjunctions, simple sentences become expanded).

Among the lower correlations were those of the word structure ratio and type-token ratio (r=.148), mean sentence length and the grammatic correctness ratio (r=.174), and composition length

and the grammatic correctness ratio (r = .191).

It would be expected that high word structure ratio would be closely related to a high type-token ratio, yet there is only a remote relationship. Apparently a high type-token ratio may be achieved largely through the use of varied nouns (Class I words) and main verbs (Class II words).

8. Dispersion of scores relative to age

It will be recalled that scores on all six objective variables were positively related to age. Because the variable age sample consisted of compositions written by subjects 10 through 18 years, it might be expected that the dispersion of scores on each of the six variables for this sample would be greater than the dispersion of scores within each of the constant age samples. For example, the standard deviation of grammatic correctness ratio scores within the variable age distribution should be consistently greater than the standard deviation of these scores within each of the constant age distributions. Examination of the ten sample standard deviations for each of the six objective variables reveals this not to be consistently the case.

The standard deviation of grammatic correctness ratio scores within the variable age

sample was 12.16 yet the standard deviations of scores of four of the nine constant age samples exceeded this standard deviation. The same held for sentence length, the word structure ratio, and the spelling correctness ratio.

This observation suggests that there is as much variation in performance of students on these variables within a one year age range as across a nine year age range.

This finding adds support to the need for class grouping on variables other than age.

C. Relationship Between Objective Variables and Teacher-judgment

Examination of the variable age sample reveals that the single objective variable which bore the strongest relationship to teacher-judgments of quality was the grammatic correctness ratio (r=.715). This observation supports the earlier discussion of inter-judge agreement. Apparently teachers' judgments of language quality are most strongly influenced by judgments of grammatic correctness.

Teachers' judgments also bore a close relationship to the type-token ratio (r=.537). Diversity (a minimum of redundancy) in use of words is also reflected in teacher-judgments of quality of written language.

The third strongest relationship to teacher-judgments was the word structure ratio (r=.408), follow closely by composition length (r=.389).

Within the nine constant age samples, correlations of each of the objective variables with teacher-judgments varied considerably. However, among the six objective variables, the grammatic correctness ratio consistently correlated highest with teacher-judgments (Table 22). The type-token ratio ranked second in its correlation with teacher-judgments in six of the nine samples, while within the three remaining samples, the word structure ratio, composition length, and sentence length each ranked second in their correlation with teacher-judgments.

The maximum multiple correlation of up to six objective variables with teacher-judgments varied between .860 (15 year constant age sample) and .739 (17 year constant age sample). The multiple correlation of the six objective variables with teacher-judgments within the variable age sample was .827, notably lower than the coefficients within four of the constant age samples.

D. Estimating Quality of Language from Objective Variables

A primary objective of this investigation was the development of a method of scoring composi-

tions of deaf students based on objective procedures, and yielding a score which reflects teacher-judgments of quality.

Multiple-regression equations were developed, the six objective variables serving as predictor variables, and normalized T scores serving as criterion scores.

Multiple-regression equations were developed for each of the ten samples. For each, variables were added one by one, in the order in which they contributed to the maximum multiple correlation with the criterion.[1] As may be noted by comparison of the multiple correlations based on six variables (Table 23) and multiple correlations based on three variables (Table 25), little was added to the multiple correlation by the addition of the final three variables.

As noted for each sample, variables 5 and 4 (grammatic correctness ratio and type-token ratio) were the two first variables to be entered into the equation, followed in three samples by variable 1, composition length, and four samples by variable 6, spelling correctness ratio. Inspection of the

---

[1]A program written for the Health Services Facility, UCLA, was selected for this operation.

relative contribution of variable 1 and 6 to the ten equations led variable 1 to be selected as the third predictor variable in each equation.

Variables 2, 3, and 6 were systematically suppressed to provide ten multiple-regression equations in which variables 5, 4, and 1 became the predictor variables.

Table 25 indicates the standard error of estimate of the criterion for each of the ten resulting equations. Standard errors vary between 7.95 and 9.92, representing standard errors of .53 standard deviation and .66 standard deviation respectively.[1]

E. Use and Limitations of the Equations

By deriving the grammatic correctness ratio, the type-token ratio, and composition length (see"Objective Analyses"), and assigning these scores to the appropriate equation (Table 27), teachers and researchers may objectively calculate an estimate of the position of a given student relative to other deaf students nationally with regard to his quality of written language (see Appendix C, or description following Table 28).

---

[1] Based on standard deviation of 15.

Several limitations exist in the use of these equations.

1. Students who are administered the test should be between 10 years, 0 months, and 18 years, 11 months.

2. Directions and the stimulus (pictorial sequence) must conform precisely to those used in this investigation (See "Administration of Language Test").

3. Norms are based on students whose mean hearing loss is 70 decibels or greater in the better ear at 500, 1000, 2000 cps.

4. Norms are based on residential and day programs consisting of 100 or more students.

Because the standard errors of estimate of teacher-judgments are substantial, the equations should not be considered sensitive instruments for determining the quality of the language of individual students, nor for estimating short-term gains in language quality (eg., see Appendices D, E, F).

The equations may be used for estimating class levels, general levels of language quality of groups of students within a particular school relative to the national population of deaf students upon which the norms are based. Similarly, these equations may be used

as research instruments where relatively large groups of students are tested.

F. Selection of Particular Equation to be Used

The equation based on the variable age sample is as follows:

.021 (CL) + .410 (TTR) + .683 (GCR) - 34.621 = f

This equation is appropriate under the following condition:

1. Research purposes for study of two or more groups which are equated by age.

Except under the above condition, constant age equations should be used. These equations are useful to estimate the following:

1. The mean position of a group of students of a constant age (eg., 14 years, 0 months through 14 years, 11 months) relative to norms for deaf students of the same age. From the mean normalized T score of such a group, the mean standard score and mean percentile may be calculated from Appendix C or Table 28.

2. The mean position of a group of students of differing ages (using a particular constant age equation for each student of a particular chronological age) relative to norms for deaf students of the same age distribution. For example, a group may consist of ten 11 year old students, eight 12 year old students, and five 13 year old students. The

11 year constant age equation (Table 27) may be used to establish the T score for each of the 11 year old subjects, the 12 year constant age equation for each of the 12 year old subjects, and the 13 year constant age equation for each of the 13 year old students. The mean T score may then be calculated, and the group tendency determined, eg., if the mean T score equals 65, we have estimated that the group as a whole is performing in language at the 84th percentile, (Appendix C), a point below which 84 per cent of deaf students whose ages are distributed similarly might be expected to fall.

This procedure may be followed for a class, a department, or a school population between ages 10 and 18. For purposes of accuracy, Table 28 can be used for calculating standard scores directly.

Again, the constant age equations as presented in Table 27 for each age level 10 - 18, are:

10 years	.036(CL)+.538(TTR)+.560(GCR)-30.806=T
11 years	.0(CL)+.420(TTR)+.605(GCR)-20.317=T
12 years	.02(CL)+.445(TTR)+.7(GCR)-33.984=T
13 years	.039(CL)+.382(TTR)+.594(GCR)-26.552=T
14 years	.031(CL)+.363(TTR)+.747(GCR)-39.168=T
15 years	.02(CL)+.513(TTR)+.747(GCR)-49.169=T
16 years	.035(CL)+.359(TTR)+.799(GCR)-49.191=T
17 years	.027(CL)+.471(TTR)+.648(GCR)-42.580=T
18 years	.007(CL)+.287(TTR)+.905(GCR)-47.675=T

G. Implications

Several implications are suggested from the results of this investigation. Some pertain directly to the language instructional process. Others pertain to further avenues of research.

1. The instructional process

The quality of written language of deaf students tends to continue to improve through 18 years. This finding lends support to continuing emphasis on language instruction through the secondary level in schools for the deaf. While the value of continued language instruction through the post-secondary years is not directly determined

from this finding, it may be stated that deaf students have not plateaued in their written language skills before 18 years, and extended language instruction would seem desirable for students who continue their studies at Gallaudet College, at the National Technical Institute for the Deaf, other post-secondary programs, and in adult education.

Teachers may with justification regard increasing composition and sentence length, increasing use of adjectives, adverbs, and function words, increasing variety in vocabulary, and increasing grammatic correctness, as reflecting quality in written language. Some of these attributes, however, are superior to others as reflections of quality. The best single index of quality is grammatic correctness, followed by variety of vocabulary usage.

In spite of these facts, none of these variables is closely related to age. The spread within a one year interval appears to be almost as great as that over nine years. Gains in quality of language as a function of age are relatively small. Accordingly, class grouping on the basis of age cannot

be supported on the basis of homogeneity of language proficiency. Class grouping for language instruction can best be achieved through the direct inspection of the language of individual deaf students. For example, more than 10 per cent of ten year old deaf students appear to be superior in written language to the mean level of 18 year old students.

It should now be possible to evaluate the mean level of written language within a class of deaf students relative to national norms. The reliability of the Written Language Test is further increased when a given school population is compared with the national population of deaf students (see "Use and Limitations of the Equations").

2. Research

The primary implication of this investigation for research is the availability of the multiple-regression equations and the national norms on deaf students. These equations should not be used with small groups, in view of the reported standard errors of estimate. The objectivity in the scoring procedure and its predictive validity in terms of quality of language make it a useful instrument for large groups.

The three objective variables used in the 10 equations share approximately 65 per cent of the variance (variable age distribution) of teacher-judgments. Thirty-five per cent remains untapped. It may be that additional variables not considered in this investigation would increase the multiple correlation of the objective variables with the criterion.[1]

The significance of differences between male and female students with respect to their normalized T scores and several objective variables highlights the importance of sex differences in sampling for research. Considerable care should be taken that samples contain males and females in proportion to the populations which they purport to represent.

---

[1]Miss Judy Fallon, Graduate assistant to this investigation and presently a teacher at the California School for the Deaf, Berkeley, has developed a seventh objective scoring procedure which correlates approximately .62 with teacher-judgments.

H. Conclusions

1. In the written language of deaf students aged 10 through 18, a positive relationship exists between age and

(a) extensiveness of writing, reflected in composition length,

(b) complexity of sentence, reflected in sentence length,

(c) use of Class III, Class IV, and function words,

(d) variety of vocabulary usage,

(e) freedom from grammatic errors,

(f) freedom from spelling errors.

2. While quality of language tends to be related to the age of the deaf student, differences within one year age intervals are major, suggesting that more attention should be given to quality of language than to age in grouping students for language instruction.

3. Female students tend to be superior to male students in the quality of their written language.

4. Variables in the written language of deaf students which relate most strongly with teacher-judgments of written language quality are grammatic

correctness and type-token ratio (variety of vocabulary usage). In addition, composition length, sentence length, frequency of usage of Class III, Class IV, and function words, and spelling correctness relate positively to teacher-judgments of quality.

5. Multiple-regression equations employing measures of grammatic correctness, variety of vocabulary usage, and composition length as predictor variables are useful in estimating teacher-judgments of quality with limits imposed by standard errors of estimate varying between approximately .53 and .66 standard deviations for different age levels.

V. Summary

This investigation was conducted for the purpose of describing the written language of deaf students varying between 10 and 18 years of age in terms of six measurable variables, and relating these variables to teacher-judgments of quality of language.

Ten stratified random samples of compositions written by deaf students were selected from 14 residential and day educational programs for the deaf throughout nine regions of the country. One sample consisted of 450 compositions, 50 of which were written by ten year old students, 50 by eleven year old students, etc., through eighteen years. Males and females were equally represented. A second sample consisted of 100 compositions written by ten year old subjects, 50 males and 50 females. Similar samples were selected for each age, through eighteen years. Compositions were written under standardized conditions, in response to the presentation of a four-picture sequence.

Three master teachers of language of the deaf subjectively scored the compositions in each of the ten samples. They were not informed of the ages of subjects who wrote compositions for the

variable age sample except in terms of the age range. They were informed of the age range (1 year) represented in each of the nine constant age samples. Each sample was scored independently. The mean correlation of the three teacher-judges' scores of the 450 compositions in the variable age sample was .744. A normalized T distribution of scores for each sample was developed, with a mean T score of 50, and a standard deviation of 15.

Each composition was scored as to composition length, sentence length, ratio of Class III, Class IV, and function words to all words, type-token ratio (variety of word usage), grammatic correctness, and spelling accuracy.

All six objective variables tended to correlate positively with age, but correlations were small. Teacher-judgments of quality showed a positive correlation of .367 with age.

Within the variable age sample of 450 compositions, females received significantly superior scores to males on teacher-judgments (normalized T scores), frequency of use of Class III, Class IV,

and function words, type-token ratio, grammatic correctness, and spelling accuracy.

Within the variable age sample, all objective variables correlated positively with teacher-judgments. Grammatic correctness correlated highest (r=.715), type-token ratio ranked second in its correlation with teacher-judgments (r=.537), frequency of usage of Class III, Class IV, and function words ranked third (r=.408), and composition length fourth (r=.389), while mean sentence length and spelling accuracy ranked fifth and sixth respectively.

Maximum multiple correlation of all objective variables with teacher-judgments was .827 within the variable age sample, and varied between .739 and .860 among the nine constant age samples. Multiple correlations of grammatic correctness, type-token ratio, and composition length with teacher-judgments were .819 within the variable age sample, and varied between .725 and .842 among the constant age samples.

Multiple-regression equations were developed with these three objective variables serving as predictors, and normalized T scores (teacher-judgments) as the criterion. The standard errors

of estimate for the ten equations varied between .53 and .66 standard deviations.

These equations are considered to be useful instruments for evaluating the written language of groups of deaf students relative to the national population of deaf students represented by the samples.

References

Birch, J. W. & Stuckless, E. R. Programed instruction and the correction of written language of adolescent deaf students. Research Report submitted to the United States Office of Education, December, 1963.

Boatner, E. B., Stuckless, E. R., & Moores, D. F. Occupational status of the young deaf of New England. Hartford, Conn.: American School for the Deaf, 1964.

Carroll, J. B. The study of language. Cambridge: Harvard Univ. Press, 1955.

Davis, Edith A. The development of linguistic skill in twins, singletons with siblings, and only children from age five to ten years. Minneapolis: Univ. of Minn. Press, 1937.

Diederich, P. B., French, J. W., & Warlton, S. T. Factors in judgments of writing ability. Res. bull. Princeton Educ. Testing Service, 1961 (Mimeographed).

Francis, W. N. Revolution in grammar. In H. B. Allen (Ed.), Applied English linguistics. New York: Appleton-Century-Crofts, 1958. Pp. 46-62.

French, J. W. Schools of thought in judging excellence of English themes. Reprint. Princeton: Educ. Testing Service, 1962.

Fries, C. C. Linguistics and reading. New York: Holt, Rinehart, Winston, 1963.

Fries, C. C. The structure of English. N. Y.: Harcourt, Brace, 1952.

Goda, S. Language skills of profoundly deaf adolescent children. J. Speech & Hearing Dis. 2, 1959.

Goda, S. Spoken syntax of normal, deaf, and retarded adolescents. J. Verb. Learn. Verb. Behav., 3, 1964.

Guilford, J. P. Fundamental statistics in psychology and education. N. Y.: McGraw-Hill, 1956.

Guilford, J. P. Psychometric methods. N. Y.: McGraw-Hill, 1954.

Gunderson, A. N. A linguistic analysis of the written language of the deaf. Unpublished doctoral dissertation, Northwestern Univ., 1965.

Heider, F. K., & Heider, Grace M. Studies in the psychology of the deaf: I A comparison of sentence structure of deaf and hearing children. Psychol. Monog., 52, 1940.

Johnson, W. Studies in language behavior: I A program of research. Psychol. Monogr., 56, 1944.

Joos, M. Language and the school child. In Harvard Educ. Rev., (Special issue on language and learning), 2, 34, 1964. Pp. 203-210.

Klare, G. R. The measurement of readability. Ames, Iowa: Iowa State Univ. Press, 1963.

Lunde, A. S. & Bigman, S. K. Occupational conditions among the deaf. Washington, D. C.: Gallaudet College, 1959.

McCarthy, Dorothea. Language development in children. In L. Carmichael (Ed.), Manual of child psychology. (2nd ed.) New York: Wiley, 1954. Pp. 492-630.

McCarthy, Dorothea. The language development of the preschool child. Univ. of Minn. Inst. of Child Welfare Monog., 4, 1930.

Mann, Mary B. Studies in language behavior: III. The quantitative differentiation of samples of written language. Psychol. Monogr., 56, 1944.

Meckel, H. C. Research on teaching composition and literature. In N. L. Gage (Ed.), Handbook of research on teaching. Chicago: Rand McNally, 1963. Pp. 966-1005.

Myklebust, H. R. The psychology of deafness. N. Y.: Grune & Stratton, 1960.

Roberts, P. Pronominal "this": a quantitative analysis. In H. B. Allen (Ed.), Applied English linguistics. New York: Appleton-Century-Crofts, 1958. Pp. 267-274.

Rosner, B. Improving competence in English expression. New York: Brooklyn College, 1963 (Mimeo.)

Rosenstein, J. & Lerman, A. Vocational status and adjustment of deaf women. Lexington School for the Deaf Res. Series 1, 1963.

Simmons, Audrey A. Comparison of written and spoken language from deaf and hearing children. Unpublished doctoral dissertation, Washington Univ., 1963.

Templin, Mildred C. Certain language skills in children, their development and interrelationships. Inst. of Child Welfare Monogr. 26 Minneapolis: Univ. of Minn. Press, 1957.

Thompson, W. H. An analysis of errors in written composition by deaf children. Amer. Ann. Deaf, 81, 2, 1936.

Walter, J. Some further observations on the written sentence construction of profoundly deaf children. Amer. Ann. Deaf, 104, 3, 1959.

Williams, H. An analytical study of language achievement in preschool children. Univ. of Iowa Studies in Child Welfare Monogr., 13, 1937.

Yule, G. The statistical study of literary vocabulary. Cambridge: The University Press, 1944.

Zipf, G. K. The Psycho-biology of language. Boston: Houghton-Mifflin, 1935.

APPENDIX A

THE ASSESSMENT OF THE WRITTEN LANGUAGE OF DEAF STUDENTS

Directions for administration of the Written Language Test

TO THE TEACHER:

The students in your class have been selected to participate in this research project by a random sampling procedure. The major purpose of this research is to develop national norms on the written language of deaf children between ages ten and eighteen. For this reason it is imperative that every teacher in the study maintain complete objectivity when administering the Written Language Test. We earnestly solicit your cooperation in this respect.

On the morning of the day of the test administration you will receive a package containing test materials and a Class Record Form. Please do not open the package until you are ready to give the test. When finished, return all materials except the Class Record Form to the package. A member of the project staff will collect the package. Please complete the Class Record Form and return it to the appropriate person for forwarding to us.

A member of the project staff will be at your school on the day of testing to answer any questions you may have and to assist in any way toward the successful completion of the testing.

WRITTEN LANGUAGE TEST

1. Write on chalkboard, "Write about the pictures. You have 45 minutes. You may use two sides of the paper."

   (Give the same directions to the children orally. Point to the pictures in sequence. Run your finger down both sides of writing paper. You may repeat the directions, but not after materials have been distributed.)

2. Distribute to each student a sheet of paper and a copy of the PICTURE SEQUENCE. Please ensure that each student has a pencil.

3. Check that each student writes his full name on top right of writing paper. Indicate that they should then begin.

4. DO NOT aid the children in any way. For example, some students may ask questions about the pictures or seek aid in spelling or selecting appropriate words. It is important that no help be given and no corrections be made for the child.

5. Collect the papers 45 minutes after the students have been told to begin. (After they have begun, if none of your students has written during a 5 minute period, collect all papers. This may occur before 45 minutes have elapsed. Enter the time elapsed on the Class Record Form.)

6. Remember to return all materials except the Class Record Form to the package. A project staff member will collect the package.

We sincerely appreciate your cooperation in this project. We anticipate that the results of this project will later assist you in teaching language to deaf students.

E. Ross Stuckless
Project Director

Appendix B

THE ASSESSMENT OF WRITTEN LANGUAGE OF THE DEAF

Directions to Judges for Scoring Compositions Without Respect to Age of Subjects

It is our wish to present a minimum number of restrictions to your scoring of the 450 accompanying compositions. These compositions were randomly selected from among compositions written by approximately 3,000 deaf students in 14 residential and day programs for the deaf. Day schools and day classes are both represented at all age levels.

For your general information, this sample consists of the following:

1. The compositions were written by boys and girls between the ages of 10 and 18 inclusive, 50 at each age level. They have been scrambled. We ask you not to attempt to infer the age of each student or to use suspected age in any way to weight your scores.

2. The compositions were written in response to the picture-sequence stimulus which is attached. The only directions given were, "Write about the pictures. You have 45 minutes." No additional assistance was given.

In scoring, please observe the following:

1. assign numerical scores.

2. utilize a numerical continuum that reflects the excellence of the compositions.

3. use a broad range of scores, with as few compositions being assigned the same score as possible; do not assign scores greater than 100.

4. the criteria for assignment of scores should include what you would ordinarily apply when judging a composition.

5. remember that the score should reflect the excellence of the composition in terms of your concept of absolute "goodness"; do not attempt to infer an age-excellence interaction.

6. in preparation for assignment of scores to individual compositions, briefly run through all the compositions to acquaint yourself with the varying quality of the compositions.

7. in scoring, place the assigned number in the upper left hand corner of the composition; there is no need to make any other notation on the compositions.

8. you may rearrange the compositions in any manner you wish since they have been sent to you in a random order.

Appendix C

Standard score (Z), and percentiles (%ile) equivalents of T scores[1]

T score	Z	%ile[2]	T score	Z	%ile[2]	T score	Z	%ile[2]
15	-2.33	1	41	-.60	27	67	1.13	87
16	-2.27	1	42	-.53	30	68	1.20	88
17	-2.20	1	43	-.47	32	69	1.27	90
18	-2.13	2	44	-.40	34	70	1.33	91
19	-2.07	2	45	-.33	37	71	1.40	92
20	-2.00	2	46	-.27	39	72	1.47	93
21	-1.93	3	47	-.20	42	73	1.53	94
22	-1.87	3	48	-.13	45	74	1.60	95
23	-1.80	4	49	-.07	47	75	1.67	95
24	-1.73	4	50	0	50	76	1.73	96
25	-1.67	5	51	.07	53	77	1.80	96
26	-1.60	5	52	.13	55	78	1.87	97
27	-1.53	6	53	.20	58	79	1.93	97
28	-1.47	7	54	.27	61	80	2.00	98
29	-1.40	8	55	.33	63	81	2.07	98
30	-1.33	9	56	.40	66	82	2.13	98
31	-1.27	10	57	.47	68	83	2.20	99
32	-1.20	12	58	.53	70	84	2.27	99
33	-1.13	13	59	.60	73	85	2.33	99
34	-1.07	14	60	.67	75			
35	-1.00	16	61	.73	77			
36	- .93	18	62	.80	79			
37	- .87	19	63	.87	81			
38	- .80	21	64	.93	82			
39	- .73	23	65	1.00	84			
40	- .67	25	66	1.07	86			

[1] Mean = 50, standard deviation = 15

[2] To closest percentile

Appendix D

Composition of 10 year old judged "average" for that age by teacher-judgment (normalized T = 50)

The Family will go to the pinic. A little girl gave bread to* dog. Mother see* eat* a basket on the table. Father play* bat and ball with* boy. A little dog stand* up see* to eat with* girl. A little boy play* bat and ball with Father outside. The Family will go to the pinic with car. A little dog bark to see with pinic. A little boy come to see with dog in car. The Family good-bye see to dog. A little dog was sad on the sidewalk. A little dog dead to careful because the family dream to keep on the street. The Family are happy laugh to love dog with a little boy. A little dog lick to boy. A little girl see to dog her arm on the seat. Mother was happy laugh to dog with the family. Father was happy laugh to dog with the family. A little girl was happy laugh to dog with the family. Father learned to help in a car on the street. A little boy ran to dog on the grass. Mother made cook to meat on the chimney fire the wood. Mother help to girl put a basket.

---

* grammatic error (see Objective analyses)

Appendix D (Cont.)

Multiple-regression equation for 10 year constant age sample .036(CL) + .538(TTR) + .560(GCR)-30.806 = T.

CL = 200 words

TTR = 56

GCR = 80

.036(200) + .538 (56) + .560(80) - 30.806 = 51.322

From Appendix C, estimated Z = .07, estimated percentile = 53.

Appendix E

Composition of 14 year old judged "average" for that age by teacher-judgment (normalized T=50)

They packed for going to the pinic outside. The little boy wanting* to bring our* dog. But her* mother said to him, "Yes, you can bring our dog* in* pinic. Then they went with the dog. The dog bark* about 3 or 4 time*. he* excited to go with us./ They went out the car. They play baseball. Then they are ready to eat their pinic. They ate sandwishes. They drink their tea or peinch. Her Mother were cooked their hammers with the breads. They have fun to ate their food. They enjoy to have fun to ate their pinic. They like to have a wonderful pinic. Their dog were exicted to have fun to play baseball. The dog try to catch softball. The little girl named is Nancy. The little boy name is Jerry. His Father name is Dale Her Mother name is Freda.

Multiple-regression for 14 year constant age sample
.031 (CL) + .363(TTR) + .747(GCR) -39.168 = T

CL = 145

TTR = 70

GCR = 84

.031(145) + .363(70) + .747(84) - 39.168 = 53.485

From Appendix C, estimated Z = .20, estimated percentile=58

Appendix F

Composition of 18 year old judged "average" for that age by teacher-judgment (normalized T=50)

Going to the Picnic

Who* planned going to the picnic. They were excited to go having a wonderful time. A woman put some sandwiches in the box* Little girl handed* to* dog. He was hungry. and ate a ham.* Little boy was happy playing* the playground. A man handed* a bat. He put some/thing in his car.

[1]A man said "Ready." They drove in the car. The boy was sad to dog and the dog was disappointed to them. He didn't like staying at home. The boy said "The dog went with them." The man was "OK. They went back home. They took a dog with them. He were happy to see a boy. A boy handed a dog. and he faced the dog. They went to San Franscio. They found the place was very good.

The man carried the lunch box, etc. Little girl cleaned a table set. The woman brought some little woods in the fireplace. She put some barbunce with the fireplace and started puting some hamburger and hotdog. The boy throw a ball to the man. The man bated and hit it. The dog jumbed to a boy. They played some games and went hitting some chair and got some sandwiches, etc. They were very happy and had a wonderful time. A boy, girl, woman washed some dishes. They played some playground. They left from the picnic to home about night. They had a wonderful time and were very tired. They arrived at home and went to bed.

---

[1]Single spaced in Appendix to conserve space
Multiple-regression for 18 year old constant age sample
.007(CL) + .287(TTR) + .905 (GCR)-47.675 = T
CL = 250
TTR = 68
GCR = 86
.007(250) + .287(68) + .905(86) - 47.675 = 52.121
From Appendix C, estimated z = .13, percentile = 55