


1965

# An Investigation into the Relationship of Grammatical Judgments to Written Composition

Farrell Edward Taylor  
*Central Washington University*

Follow this and additional works at: <http://digitalcommons.cwu.edu/etd>

 Part of the [Educational Assessment, Evaluation, and Research Commons](#), [Educational Methods Commons](#), and the [Scholarship of Teaching and Learning Commons](#)

---

## Recommended Citation

Taylor, Farrell Edward, "An Investigation into the Relationship of Grammatical Judgments to Written Composition" (1965). *All Master's Theses*. 520.  
<http://digitalcommons.cwu.edu/etd/520>

This Thesis is brought to you for free and open access by the Master's Theses at ScholarWorks@CWU. It has been accepted for inclusion in All Master's Theses by an authorized administrator of ScholarWorks@CWU.

AN INVESTIGATION INTO THE RELATIONSHIP  
OF GRAMMATICAL JUDGMENTS TO  
WRITTEN COMPOSITION

---

A Thesis  
Presented to  
the Graduate Faculty  
Central Washington State College

---

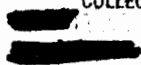
In Partial Fulfillment  
of the Requirements of the Degree  
Master of Education

---

by  
Farrel Edward Taylor  
August 1965

LD  
5771.3  
T241i

SPECIAL  
COLLECTION



1948  
of his  
of the  
of the  
of the

1948  
of the  
of the  
of the

ya

to the

of the

130729

APPROVED FOR THE GRADUATE FACULTY

---

John Herum, COMMITTEE CHAIRMAN

---

William D. Floyd

---

Keith Rinehart

## TABLE OF CONTENTS

CHAPTER	PAGE
I. INTRODUCTION . . . . .	1
The Purpose . . . . .	1
General considerations . . . . .	1
The postulate . . . . .	3
The purpose . . . . .	3
Definitions of Terms . . . . .	4
Acceptance . . . . .	4
Grammaticality . . . . .	4
Semantic relationships . . . . .	4
Sense of grammaticality . . . . .	4
Sequence . . . . .	5
Syntactic relationships . . . . .	5
Organization of the Thesis . . . . .	5
II. A REVIEW OF THE LITERATURE . . . . .	6
Views Concerning Grammatical Judgments . . . . .	6
A Pertinent Study . . . . .	10
The Need for Further Investigation . . . . .	12
III. THE SUBJECTS . . . . .	14
Their Backgrounds . . . . .	14
Their Achievement . . . . .	15
IV. DETERMINING ACCEPTANCE: THE MEASURING INSTRUMENT . . . . .	18
Factors Influencing the Design . . . . .	18

CHAPTER	PAGE
The procedure . . . . .	18
The materials . . . . .	19
The Measuring Instrument . . . . .	20
Form A . . . . .	20
Form B . . . . .	21
Controls . . . . .	21
Scoring . . . . .	25
Administering the Test . . . . .	26
Directions . . . . .	26
Differences . . . . .	27
Results . . . . .	29
V. EVALUATING COMPOSITION: ASSIGNMENTS, CRITERIA, AND VALIDITY . . . . .	31
The Assignments . . . . .	31
The Criteria . . . . .	32
Expression . . . . .	33
Grammar . . . . .	33
Thought . . . . .	36
The Validity . . . . .	38
VI. CORRELATIONS OF THE INDIVIDUAL TESTS AND GENERAL RESULTS . . . . .	42
The Individual Correlations . . . . .	42
Methods . . . . .	42
Two acceptance forms . . . . .	42

CHAPTER	PAGE
Composition and form A . . . . .	44
Composition and form B . . . . .	46
Composition and form AB . . . . .	46
The Iowa and form AB . . . . .	48
Some General Conclusions . . . . .	51
VII. SUMMARY AND CONCLUSIONS . . . . .	52
Summary . . . . .	52
Conclusions . . . . .	55
BIBLIOGRAPHY . . . . .	61

## LIST OF TABLES

TABLE	PAGE
I. Expression Scores of Subjects on Iowa Test of Educational Development . . . . .	15
II. Composite Scores of Subjects on Iowa Test of Educational Development . . . . .	17
III. Linguistic Materials (Form A) Listed in Order of Presentation . . . . .	22
IV. Linguistic Materials (Form B) Listed in Order of Presentation . . . . .	23
V. Frequency Distributions of Acceptance Test Scores . . . . .	29
VI. Composition Evaluating Criteria Part I: Expression and Purpose . . . . .	34
VII. Composition Evaluating Criteria Part II: Grammar and Structure . . . . .	35
VIII. Composition Evaluating Criteria Part III: Thought . . . . .	37
IX. Frequency Distribution of Composite Composition Scores . . . . .	39



LIST OF FIGURES

FIGURE	PAGE
1. Correlation of Composition Scores and Scores on Iowa Test of Educational Development Test Number Three, Correctness of Expression . . .	40
2. Correlation of Acceptance Scores, on Form A and Form B . . . . .	43
3. Correlation of Composition Scores and Scores on Acceptance Test, Form A . . . . .	45
4. Correlation of Composition Scores and Scores on Acceptance Test, Form B . . . . .	47
5. Correlation of Composition Scores and Scores on Acceptance Test, Form AB . . . . .	49
6. Correlation of Acceptance Scores Form AB and Composite Scores on Iowa Test of Educational Development . . . . .	50

## CHAPTER I

### INTRODUCTION

In view of the widespread agreement of research studies based upon many types of students and teachers, the conclusion can be stated in strong and unqualified terms: the teaching of formal grammar has a negligible or, because it usually displaces some instruction and practice in actual composition, even a harmful effect on the improvement of writing (1:37-38).

The vast accumulation of data which prompted this absolute statement by Braddock, Lloyd-Jones and Schoer is impressive, and it leaves little room for doubt about the value of formal grammar study to the public schools. However, this is a purely negative statement, one which derives its own value to education primarily from its ability to stimulate further research which may eventually result in wider knowledge about the act of writing, itself. At least one question is raised by the above statement: if writers do not rely upon their acquaintance with formal grammar to guide their writing, what do they employ?

#### I. THE PURPOSE

General considerations. It is quite apparent that writers are guided by some factor which is related to formal grammar but not derived from it. What that factor may be, and whether it can be isolated are two additional questions which have not, as yet, been answered. But Lloyd's view

that "accepted written usage is tied more closely to our divergent speech ways than we have been led to think" suggests the direction in which to look for the answers (8:601).

An interesting concept postulated by several prominent linguists may provide the link between spoken and written language; and this link may prove to be the factor by which writers are guided. Lenneburg describes the postulate well:

It is usually assumed by linguists--and there are compelling yet intuitive reasons for this--that there must be a finite set of rules which defines all grammatical operations for any given language, and that any native speaker will produce sentences that conform to these grammatical rules, and are recognized as being grammatical by any speaker of that community (7:876).

Ghomsky also mentions these ". . .rules that the native speaker must somehow have internalized when he has achieved the ability to produce and understand new sentences" (2:179). Both men contend that the facility to form one's own utterances grammatically, and to recognize the grammaticality of others' utterances is common to native speakers of a language. While their statements were made in regard to spoken rather than written language, it is not illogical to assume for the moment that they may also pertain to written composition--that this "set of rules" may be the guide for writing, as well.

One way to test this assumption would be to correlate written composition with another factor derived from those

rules. Two language responses were suggested to be connected with this abstraction: the ability to form grammatical utterances and the ability to recognize the grammaticality of others' utterances.

The latter possesses two distinct advantages which make it appropriate for this study. First, recognition of grammaticality should be relatively easy to measure; if a number of subjects are asked to judge the grammaticality of certain utterances, their judgments will theoretically reflect the influence of the "internalized rules." Second, since judgments of grammaticality are pertinent to modern linguistic theory, any additional data pertaining to such judgments may have important theoretical implications.

The postulate. If a subject's writing achievement and his acceptance of the sample utterances as grammatical are both derived from the same source, a set of internalized rules, they should show marked similarities. The judgments of the group might then be expected to occur in a range of scores which parallels the subjects' scores in written composition in whatever degree the two factors are similar. Correlating them will show the extent of the relationship.

The purpose. The purposes of this study were four: (1) to measure the writing abilities in a group of subjects;

(2) to measure the willingness of those subjects to accept certain sentence-like utterances as grammatical--their grammatical judgments; (3) to determine whether the subjects' acceptance of those utterances will occur in a range of scores; and (4) to correlate the subjects' scores on the two factors to determine the extent to which they are related.

## II. DEFINITIONS OF TERMS

Acceptance. A subject's willingness to accept relatively meaningless utterances as grammatical, as reflected by the number of test utterances he accepts, will be termed his "acceptance." This should not be confused with his ability to discriminate between grammatical and ungrammatical utterances, which is beyond the scope of the present study (see Chapter IV).

Grammaticality. The order or structural form which characterizes an utterance as conforming to the norms of English will be termed "grammaticality."

Semantic relationships. "Semantic relationships" will be used to denote those relationships within an utterance which are based primarily upon meaning rather than structure.

Sense of grammaticality. That "internalized set of rules" by which native speakers pattern their utterances will be termed, synonymously, a "sense of grammaticality,"

"grammatical sense," or "intuitive grammatical sense" throughout the remainder of this thesis.

Sequence. The term "sequence" will refer to those sentence-like utterances which may, but need not, be grammatical. It will encompass a range of utterances from normal sentences to totally ungrammatical word sequences; it is similar to "utterance," but does not imply that the sequence will ever be naturally uttered.

Syntactic relationships. The term "syntactic relationship" will refer to the interrelations between words which result primarily from structural rather than semantic origins.

### III. ORGANIZATION OF THE THESIS

The remainder of the thesis will fall into three general sections. The first of these, consisting of chapters two and three, will review the available literature and describe the subjects. The next, comprising chapters four and five, describes the methods employed in measuring the subjects' acceptance and composition achievement, and includes the results of the separate tests. In the final portion, chapter six deals with the correlations between the individual test scores; chapter seven presents the conclusions drawn from the cumulative data, and terminates the thesis.

## CHAPTER II

### A REVIEW OF THE LITERATURE

Although many investigations into various aspects of written composition have been conducted, none has investigated the relationship of judgments of grammaticality to writing. Indeed, few investigations have been concerned with grammatical judgments in any respect. But, since these judgments will be employed in this study, this chapter will present the views of several prominent linguists in regard to the postulated intuitive sense of grammaticality and to grammatical judgments. The one pertinent study investigating their significance to the study of language will also be reviewed, and the need for further study established.

#### I. VIEWS CONCERNING GRAMMATICAL JUDGMENTS

The ability possessed by each native speaker of a language to structure his utterances into patterns consistent with the norms of his language is a phenomenon which has long interested linguists, for it is fundamental to the study of language. In attempts to formulate that basic grammatical sense into a grammar which will more precisely describe the language being studied, certain linguists have incorporated as a part of their methods the subjective

judgments of native speakers. The subjects are asked to judge the grammaticality of certain "generated" word sequences. Many sequences have been submitted for such judgment in order to add to that body of knowledge which may eventually lead to a theoretical generative grammar--one which will describe the set of rules by which speakers intuitively operate, and will itself generate all grammatical sequences and no ungrammatical ones.

Chomsky advocates the use of grammatical judgments. He states this clearly: "One way to test the adequacy of a grammar proposed for L [the particular language being investigated] is to determine whether or not the sequences that it generates are actually grammatical, i.e., acceptable to a native speaker, etc." (3:13). Chomsky also contends that the native speaker's ability to recognize the grammaticality of an utterance is independent of the meaning of the utterance being judged (3:15). He uses his now widely known sequence, "colorless green ideas sleep furiously," as an example of a word-sequence which exhibits little readily recognizable meaning, but which has immediately apparent grammatical structure. He states further, that any speaker of English will recognize the grammaticality (3:15), and that ". . .the basis for whatever meaningfulness we can assign to it [the sequence above] is its independently recognized grammatical structure" (2:184).



Lenneberg postulates this intuitive sense much as Chomsky does, but while seconding Chomsky, he states his views in more absolute terms:

We are dealing here with an extremely complex mechanism and one that has never been fully described yet in purely formal terms for any language (if it had, we could program real or theoretical computers that could speak grammatically) and yet, we know that the mechanism must exist for the simple reason that every speaker knows and agrees with fellow speakers about whether a sentence is grammatical or not. (This has nothing to do with familiarity or meaning of an utterance. Chomsky demonstrates this convincingly by comparing the two sentences "Colorless green ideas sleep furiously" and "Furiously sleep ideas green colorless" where both sentences are meaningless and have never been heard before. Yet one is recognized as grammatical and the other is not.)" (7:876).

Both linguists agree that a native speaker's recognition of the grammaticality of an utterance is independent of the meaningfulness or familiarity of the utterance. Since meaning is a variable which is difficult to adequately control, and since (according to Chomsky and Lenneberg) native speakers can recognize grammaticality independently of meaning, relatively meaningless sequences appeared to be the appropriate materials to present to the subjects for judgment in this study. The influence of the uncontrolled variable would theoretically be reduced without affecting the validity of the judgments.

Hill, however, disagrees with both Chomsky and Lenneberg. He questions the value of grammatical judgments, particularly those based upon sequences such as Chomsky's,

above. In doing so, Hill also questions whether subjects can, in fact, recognize grammaticality independently of meaning; he states that, given normal word order and intonation pattern, a sequence may be accepted by a native speaker, but that the listener's judgment will be based upon qualities other than Chomsky's "independently recognized grammatical structure." Hill invokes Joos' "law," contending that naive listeners tend to overlook conflicts in meaning by interpreting the conflicting word or words as those which will "do least violence to the context," that they glean from the utterance a semantic relationship that is not necessarily present (6:169-170). The informants, influenced by word order and intonation pattern, subconsciously alter the awkward or unmeaningful portion according to the context and accept or reject the utterance on a semantic basis, according to Hill; this link with contextual meaning causes an informant to judge isolated sequences on grounds too tenuous to be of linguistic significance (6:169-170).

Fries also questions the use of grammatical judgments as a test of a proposed grammar. He states reservedly that:

In the discussions of those who have tried to understand these new approaches a number of fundamental questions have been raised for which adequate answers do not seem to be available in the published materials. Valid criteria for the judgments of "grammaticality" as applied to sentences are essential for a generative grammar. The theoretical and practical principles upon which the criteria now used depend seem hard to find. It is also difficult to determine all the criteria to be used to judge the acceptability or permission of any particular type of "transformation." (5:91).

While Hill and Fries raise several interesting theoretical questions concerning the value of grammatical judgments, their doubts do not affect the structure of the present study. Whatever it may be in the sequences that stimulates the responses, whether patterns of meaningful units or structure independent of meaning, each subject's judgments of them are presumably governed by his individual sense of grammaticality; and they are, therefore, a tangible and measurable expression of that sense.

## II. A PERTINENT STUDY

The question raised by Hill concerning the reliability of individual judgments of grammaticality was responded to in a study conducted by Maclay and Sleator. Their purpose was expressly

. . . To investigate in some detail a fundamental assumption underlying the methods of linguistic analysis. This is the often implicit belief that native speakers of a language are able to make certain reliable and linguistically relevant decisions about their own language (9:275).

The study consisted of presenting thirty-six sequences of six different types to fifty-seven undergraduates enrolled in beginning rhetoric classes. The subjects were divided into three groups, and the same tape-recorded sequences were submitted to each group separately. The materials ranged from utterances that were ordinary, through utterances that were grammatical-but-meaningless (much like the Chomsky

example above), to those that were clearly ungrammatical. The experimental variable was the criterion by which each group was asked to judge the utterances. The criterion for one group was meaningfulness; the next was grammaticality; and the third group's was ordinariness.

The results of this study tended to verify Chomsky's view that unlikely and meaningless sequences can be grammatical--that naive native speakers can and do recognize the grammatical qualities of such sequences (9:279). Nevertheless, the judgments of the three groups were far from absolute; the informants were unable to discriminate with a high degree of reliability the particular quality by which they were judging. However, of the three groups, the one judging grammaticality was considerably more consistent and achieved a significant level of discrimination (9:280).

Although the Maclay-Sleator study suggests that Chomsky's earlier contention--that grammaticality and meaning are independent factors--may be correct, it does not negate Hill's statement that meaning is imposed by informants upon meaningless sequences, thereby making them grammatically acceptable. However, the informants' decisions--whether intuitive or reasoned, whether influenced by structure or meaning--were found by the experimenters to be linguistically significant: "We find that subjects are able to rank a variety of word sequences in a linguistically relevant way" (9:280).

### III. THE NEED FOR FURTHER INVESTIGATION

The results of the Maclay-Sleator study show that while many grammatical but relatively unmeaningful sequences were accepted as grammatical by the informants, none were accepted by all. Also, several ungrammatical sequences enjoyed the same level of acceptance as many grammatical ones. These results seriously question Lenneberg's statement that all native speakers agree about whether a sentence is grammatical. Chomsky, however, recognizes that absolute agreement will not occur (3:17). But he does not attempt to account for the divergent judgments. Maclay and Sleator mention the inconsistencies they encountered, but since they anticipated no absolute results, they emphasize the statistical properties of their data. Their results tend to raise doubts about the significance of all such judgments--except as a statistical value, as they point out. But statistics, even with an adequate sampling, cannot explain these inconsistencies, e.g. their sequence "Label break to calmed about and" was accepted by several subjects (9:281-282). Better controls and a more representative sampling are not likely to eliminate these judgments, but any theory which fails to explain them is inadequate.

Perhaps Hill is correct in stating that isolated sequences are not conducive to reliable judgments. But Hill poses another still more interesting question: that the

grammaticality of the sample utterances may not be the sole influence upon the subjects' judgments; that uncontrollable variables such as connotative meaning may also affect their decisions.

Two additional variables might also explain the conflicting judgments: the first, the subjects' interpretations of the criterion, "grammaticality," was mentioned by Maclay and Sleator, but not resolved; the second also lies with the subjects themselves--their abilities in language as reflected by their achievements in other areas of language behavior, specifically writing. The first variable, individual subject's interpretations of the criterion, is impossible to ascertain; but the latter variable provides a means by which more can be learned about both grammatical judgments and written composition. This study will set out to explore the relationships between these two factors by placing them in a perspective which may shed light upon both forms of language behavior.

## CHAPTER III

### THE SUBJECTS

Since the students provided both the grammatical judgments and the compositions--both bases for the correlations--they must be identified as accurately as possible.

#### I. THEIR BACKGROUNDS

The subjects for this study were a group of 140 senior students enrolled in three terminal and three college preparatory classes at Sumner Senior High School, Sumner, Washington. They represent a variety of backgrounds from suburban Sumner (population under 4,000) and the surrounding non-high school districts of primarily rural character, a fact which influences the school environment; the high school enrollment of 750 students belies the town's size.

This wide drawing area also broadened the range of the subjects' backgrounds, both economical and ethnic. Several students from military families who had traveled abroad were also present. Perhaps the most important factor which added contrast and sophistication to the group was Sumner's proximity to the urban centers of Tacoma and Seattle. This factor increased the variety of trades and professions in the area. All subjects were native speakers of English with no formal background in linguistics; they were appropriate

subjects for a study of this nature.

## II. THEIR ACHIEVEMENT

In addition to the wide variety of economic and cultural backgrounds, the subjects also represented a broad range of academic capabilities. One hundred twenty-seven of the subjects took the Iowa Test of Educational Development in their junior year. Their scores are shown in Table I.

TABLE I  
EXPRESSION SCORES OF SUBJECTS ON IOWA  
TEST OF EDUCATIONAL DEVELOPMENT

ITED Expression Score		
<u>Centile</u>		<u>Frequency</u>
91-100		8
81- 90		9
71- 80		5
61- 70		14
51- 60		13
41- 50		9
31- 40		31
21- 30		22
11- 20		10
1- 10		6
		Total 127
Mean	46	
Median	39	
Mode	40	

NOTE: Scores are expressed in centiles based upon ITED national norms.



The "correctness of expression" scores are particularly appropriate to the express purposes of this study. According to the publisher's note to the students, "your score on this test indicates your ability to write correctly, to use proper words in expressing your ideas, and to organize your writing sensibly" (11:1). The expression test scores range from the first through the tenth deciles, but the mean for the group in this column is the forty-sixth centile, beneath the test norms. This becomes significant when the median score (39) and the mode (40) are noted. A closer look at Table I reveals that the distribution is slightly bimodal. The frequency distribution indicates that 78 scores (61 per cent) are at or below the fiftieth centile, while only 49 are above it.

The composite scores for the group, shown in Table II, are more nearly normal; the bimodal tendency has disappeared, and the distribution has become nearly normal. The mean has risen to the fiftieth centile; and although the median (48) and mode (47) are still somewhat lower, they are much closer to the mean than were the same figures on the expression scores. Only 66 scores (52 per cent) are at or below the fiftieth centile.

Both expression and composite scores in Table II suggest that the subjects as a group, although including extremely high and similarly low scores, are somewhat below

the national norms established for the Iowa Test.

TABLE II  
COMPOSITE SCORES OF SUBJECTS ON IOWA TEST  
OF EDUCATIONAL DEVELOPMENT

ITED Composite Score		
<u>Centile</u>		<u>Frequency</u>
91-100		7
81- 90		9
71- 80		10
61- 70		16
51- 60		19
41- 50		21
31- 40		15
21- 30		16
11- 20		10
1- 10		4
		Total 127
Mean	50	
Median	48	
Mode	47	

NOTE: Scores are expressed in centiles based upon ITED national norms.

## CHAPTER IV

### DETERMINING ACCEPTANCE: THE MEASURING INSTRUMENT

The lack of a readily available test necessitated the fabrication of a group of utterances designed to measure an overt expression of an intuitive sense. This chapter provides a discussion of both the rationale which influenced its design, and the instrument itself.

#### I. FACTORS INFLUENCING THE DESIGN

The procedure. This study is based upon the fundamental assumption that a subject's judgments of grammaticality reflect his sense of grammaticality; this is tacitly accepted by Chomsky and others attempting to study these matters. However, in this study no attempt was made to determine the subjects' ability to discriminate between grammatical and ungrammatical sequences; nor were the judgments interpreted as a means of determining the relative grammaticality of the test sequences. The factor being measured was merely the number of items each subject would accept--his "acceptance." One advantage of this procedure is that it makes possible an instrument composed of grammatical items (in the judgment of the experimenter), thus this procedure simplifies the informant's decision. He only has to recognize a quality that is present, rather than determine its presence or absence.

The materials. The experimental sequences included in the final forms of the instrument were strongly influenced by the Chomsky example noted above: "colorless green ideas sleep furiously." Others were similar to those Maclay and Sleator categorized "grammatical, not meaningful, not ordinary" in their study (9:277). All were designed to provide a variety of utterances for judgment. The experimental sequences were bounded on one extreme by the three normal controls; on the other, by the two ungrammatical controls: from "a collision made the scene a shambles," to the ungrammatical "the quit self an very brings." Within these bounds the sequences varied from nearly normal to several reminiscent of Lewis Carroll's "Jabberwocky," for since all experimental sequences were to be grammatical, they must provide a variety that would encompass the acceptance of the most liberal subject.

A closer look at Chomsky's example reveals a high degree of what Francis terms "lexical incompatibility" (4:22). The word order, the affixes, and the stress and intonation patterns identify this sequence as grammatical, but the particular words don't normally occur together. This sequence served as a model for many of the sequences devised for this test. Several more obscure utterances were patterned after Carroll's mode. These were composed of nonsense words coined for this purpose. The coined words

exhibited normal affixes, and, when arranged in normal word order, linked with appropriate function words, and read with ordinary stress and intonation patterns, they formed grammatical (if nonsensical) sequences of English--albeit with little chance of natural occurrence.

## II. THE MEASURING INSTRUMENT

The test sequences were presented to the subjects in two forms, called for convenience form A and form B. Although similar in many respects, the forms are different in presentation.

Form A. Form A consists of thirty sequences--twenty-five experimental and five controls--shown in Table III in the order of presentation to the subjects. First were two practice items used to acquaint the subjects with the materials. They provide a clear contrast between the ungrammatical (P-1) and the grammatical (P-2) by employing the same words in differing ways. The test sequences were in random order, but the controls were roughly spaced throughout the test. Those marked with asterisks are the control sequences: numbers six, fourteen, and twenty-six are normal; numbers ten and twenty are ungrammatical controls.

The remaining sequences represent the culmination of the many ideas discussed above. The debt owed Chomsky is obvious in number seven; and Carroll's presence is apparent

in numbers thirteen, eighteen, twenty-one, and twenty-eight. The remainder are similar to Chomsky's but are tempered by Maclay and Sleator's more conventional influence.

Form B. The materials in form B were basically the same as those in the first, except that each sequence was paired with another--an inverted form of the original. Table IV enumerates both the original (as found in form A) and the inversion. Each inversion was similar to the original, but some structural changes resulted. For example, in altering number nine, "minutes" became a sentence adverbial--a fact that necessitated that the inversion be read with different intonation and stress patterns. Others changed from declarative to interrogative, etc., demanding similar modifications. Some became more obscure; however, each inversion was read with intonation and stress patterns corresponding to its formula as delineated by Sledd (10:153-163). Those which were originally grammatical remained so; the ungrammatical controls retained their original lack of well-formedness.

Controls. The five control sequences--three normal sentences and two clearly ungrammatical sequences as noted above--were included as a means of measuring the respondents' sincerity. Any subject who answered incorrectly four of the five--rejecting the normal sentences or accepting the

TABLE III

LINGUISTIC MATERIALS (FORM A) LISTED  
IN ORDER OF PRESENTATION

Number accepting	Sequence
P-1	Home dived cleanly stones fresh.
P-2	Fresh stones dived cleanly home.
(103)	1. Sprinkle words brightly over the moral tonnage.
( 51)	2. Mornings sleeves question the necessity.
(134)	3. Remorseful destiny forbids fresh interpretations.
( 33)	4. Reminisce the glass a wreath.
(117)	5. The grave granted them all decisions.
(118)	6. We usually have a battery of full voices.*
( 94)	7. Creativity moves the inert to doubt.
( 65)	8. Topical azures become uncrated panoramas.
( 77)	9. Frailty appears strolled inaccurately.
( 7)	10. The quit self an very brings.*
( 74)	11. Nights showers alibi existence.
( 35)	12. A size different smell blue liars.
( 86)	13. The nainies congoled several reps of unclathed wallers.
(118)	14. A collision made the scene a shambles.*
(109)	15. A shred made anarchy its choice of chaos.
( 96)	16. Necessity breathes coherent answers for sprucing play.
( 67)	17. Odorless mediocrity smells freshly engraved.
( 27)	18. Revolently lames the quivic nofter.
(120)	19. Ideals abscond minutes into obscurity.
( 14)	20. Lobby no straw recover in were keeps.*
( 46)	21. The robb throught the grobes a ravish of mandy worens.
(118)	22. Exercise can make the lamp ambitious.
( 85)	23. Fronted ice is eminently exertable.
( 41)	24. Bless the wind two pains.
( 62)	25. Wrinkly grass between laws revolves pedal nine.
(108)	26. Steel yourself to the water.*
( 70)	27. Quaintly handicapped are few visages.
( 40)	28. Three grouns niggled to the fren.
(116)	29. Reality is a fretful orange.
( 50)	30. Fell sent the weather a management.

NOTE: An asterisk indicates a control sequence.

TABLE IV  
LINGUISTIC MATERIALS (FORM B) LISTED  
IN ORDER OF PRESENTATION

Number accepting	Sequence pair
P-1	Home dived cleanly stones fresh. Cleanly stones home dived fresh.
P-2	Fresh stones dived cleanly home. Fresh stones dived home cleanly.
(124)	1. We usually have a battery of full voices.* Usually we have a full battery of voices.*
( 83)	2. Exercise can make the lamp ambitious. Exercise can make the ambitious lamp.
(116)	3. The grave granted them all decisions. All decisions granted them the grave.
( 27)	4. Nights showers alibi existence. Showers existence alibis nights.
(128)	5. A collision made the scene a shambles.* The collision made a shambles of the scene.*
( 74)	6. Steel yourself to the water.* Steel to the water yourself.*
( 77)	7. Remorseful destiny forbids fresh interpretations. Remorseful destiny fresh interpretations forbids.
( 34)	8. Necessity breathes coherent answers for sprucing play. Necessity breathes coherent answers for play sprucing.
( 92)	9. Ideals abscond minutes into obscurity. Minutes ideals abscond into obscurity.
( 24)	10. The quit self an very bring.* And very brings the self quit.*
( 99)	11. Topical azures become uncrated panoramas. Uncrated topicals become azure panoramas.



TABLE IV (continued)

- ( 86) 12. Reality is a fretful orange.  
Is an orange reality fretful.
- ( 58) 13. A size different smell blue liars.  
Blue liars smell a different size.
- ( 97) 14. Quaintly handicapped are few visages.  
Quaintly handicapped visages are few.
- ( 80) 15. Fronted ice is eminently exertable.  
Exerted ice is eminently frontable.
- (127) 16. Creativity moves the inert to doubt.  
Doubt moves the inert to creativity.
- ( 37) 17. Mornings sleeves question the necessity.  
Sleeves mornings question the necessity.
- ( 89) 18. Frailty appears strolled inaccurately.  
Inaccurately strolled frailty appears.
- ( 60) 19. Fell sent the weather a management.  
The management sent fell a weather.
- ( 17) 20. Lobby no straw recover in were keeps.\*  
Straw no recover in were lobby keeps.\*
- ( 73) 21. Wrinkly grass between laws revolves pedal nine.  
Wrinkly grass laws revolve between pedal nine.
- (119) 22. A shred made anarchy its choice of chaos.  
Anarchy made a shred a choice of its chaos.
- ( 56) 23. The nainies congoled several reps of unclathed  
wallers.  
Of several unclathed wallers the nainies reps  
congoled.
- ( 43) 24. Bless the wind two pains.  
Wind bless the two pains.
- ( 52) 25. Reminisce the glass a wreath.  
The glass reminisced a wreath.
- (123) 26. Sprinkle words brightly over the moral tonnage.  
Sprinkle words over the brightly moral tonnage.

TABLE IV (continued)

- (90) 27. Revolently lames the quivic nofter.  
The quivic nofter lames revolently.
- (86) 28. Odorless mediocrity smells freshly engraved.  
Engraved mediocrity smells freshly odorless.
- (83) 29. Three grouns niggled to the fren.  
To the fren niggled three grouns.
- (59) 30. The robb throught the grobes a ravish of mandy  
worens.  
The ravish of mandy worens throught the grobes a  
robb.

NOTE: An asterisk indicates a control sequence.

ungrammatical ones in any combination totaling four--was considered to have responded fraudulently, and his answer sheet was discarded. The reasoning involved was that a subject whose acceptance was so broad that he accepted an ungrammatical control would surely not reject a normal sentence. Even should this occur, it would be extremely unlikely that a sincere effort would result in four such responses.

Scoring. Since all of the sequences except two controls were grammatical, and those may have been legitimately accepted by an extremely liberal subject, the total score for each subject was simply the number of items accepted. To deduct those which were answered incorrectly would merely

introduce discrimination into the study. No measurement of the subjects' accuracy of discrimination was intended or implied (although this may be inherent in the method). The measurement sought was simply the individual subject's willingness to accept relatively meaningless word sequences to be grammatical, as reflected by his acceptance of a specific group of such utterances.

### III. ADMINISTERING THE TEST

The acceptance test was administered to the subjects in the two forms spaced a week apart. Both forms were tape-recorded to ensure that all subjects in the six class groups were judging the same materials read in the same way. This method also made certain that each item received approximately an equal amount of the subjects' time. Because the sequences varied slightly in length, only the time allotted each response (eight seconds) was measured. All sequences were read with normal intonation and stress patterns--even the two ungrammatical controls.

Directions. To provide an incentive while allowing freedom to judge intuitively, the subjects were told that the test results would not influence their grades, but that a good grade on the test could result in exemption from certain future class assignments. They were further instructed

to listen carefully to the recorded utterances and to judge, to the best of their abilities, the grammaticality of each sequence. The instructions included two practice sequences which were judged by the subjects. Each practice item was timed, just as were the sequences which followed: five seconds after each was read, the subjects were told to record their answers; three seconds later, the next sequence was begun.

The subjects recorded their answers by merely circling the more nearly correct answer--"is" or "is not"--in response to the question: "Is the following sequence of words a grammatical sentence?" After answering the practice sequences, the subjects were told the correct answers; the sequences were then repeated for additional clarification. Only then did the actual test begin.

Differences. Minor changes in administering the two forms were necessitated by the differences in the forms themselves. Each sequence was repeated once in the first form, but the separate versions in each pair were heard only once in form B. This kept the reading times nearly the same; the time allotted for responding was the same also.

Other differences were introduced into the forms deliberately to gain consistency without unduly influencing judgment. In form A the subjects were subtly encouraged to be liberal--to accept those sequences that they were uncer-

tain about. They were encouraged to use their own experience with language as the only basis upon which to judge, and to remember that there was no penalty for incorrect responses--to play hunches if undecided.

The opposite was true of the directions for form B. Conservatism was fostered in many ways: by implying that accuracy of discrimination was much more important in this phase of the test; by telling the subjects that while the new sequences were similar to the earlier ones, many had been changed slightly--a fact which could have influenced the sequences grammaticality; but most important of all, by pairing the sequences. The latter was probably more influential than the other more subtle means. The subjects were directed to listen closely to each half of the pair, and, if either were unacceptable, to reject the pair.

The purpose of the biased directions was primarily to reduce guessing should subjects be in doubt. The intent was to encourage the subject who may be vacillating to lean toward one extreme in form A and toward the other in form B. When both scores are totaled, the errors theoretically cancel each other and a more accurate score should probably result. The two separate forms were clearly not intended to measure reliability, nor was a reliability measure administered.

## IV. RESULTS

The distributions of scores attained on forms A and B are listed in Table V, together with the composite scores listed as form AB. Of particular interest is the similarity of mean, median, and mode for both forms. However, the differences in the two forms are prominently shown in the two ranges. Form A has a much wider range than B. This may

TABLE V  
FREQUENCY DISTRIBUTIONS OF ACCEPTANCE TEST SCORES

Intervals			Frequencies			
Form A	Form B	Form AB	Form A	Form B	Form AB	Standard deviations
25-26	25-26	47-49	1		2	3.0
24	24	44-46	3	1	2	2.5
22-23	22-23	41-43	1	5	6	2.0
20-21	20-21	38-40	10	13	12	1.5
18-19	18-19	35-37	21	20	15	1.0
17	17	32-34	15	13	21	.5
16	16	31	14	13	12	0
15	15	28-30	16	18	21	-.5
13-14	13-14	25-27	32	30	29	-1.0
11-12	11-12	22-24	20	17	15	-1.5
9-10	9-10	19-21	4	8	3	-2.0
8	8	16-18	1	2	2	-2.5
6- 7	6- 7	13-15	2			-3.0
Totals			140	140	140	
Mean	16	16	31			
Median	15	15	30			
Mode	14	14	26			
Sigma	3.45	3.50	6.09			

reflect the differences in design, or in directions; perhaps both. It also suggests that the emphasis upon accuracy curbed both extremes.

Notice the frequency distribution of the composite (AB) scores closely. The range appears to broaden again in the AB column, but referring to the interval portion of the table shows that the range has in fact shortened considerably, indicating that the extreme scores in form A have been moderated by form B even more than is superficially apparent in the distributions.

Generally, the results of the acceptance tests show a preponderance of scores falling just below the mean, with a broad distribution of higher scores above it. This pattern is a familiar one; it is similar to the distributions in both composition measurements discussed above--those attained by the criteria devised for this study, and by the Iowa Test of Educational Development. The correlations which follow will determine whether the similarities noted here are in fact meaningful.

## CHAPTER V

### EVALUATING COMPOSITION: ASSIGNMENTS, CRITERIA, AND VALIDITY

Evaluating the subjects' writing abilities was the second phase of the study. This task entailed two problems which were only partially resolved: equating the assignments among six class groups, and formulating a system to objectively measure a quality which challenges such methods of measurement.

#### I. THE ASSIGNMENTS

Three written compositions provided the basis for evaluating each subject's skill at writing. Each of the compositions was intended to be of nearly equal difficulty; however, different materials were used in the college preparatory and terminal English classes. The college preparatory classes wrote two analyses based upon essays read in class, and, in addition, a longer critical review of a novel chosen from a prescribed list. The review was written outside of class. The terminal classes wrote an analysis of a short story and two novel reviews; one of the latter outside of class. Each assignment was thoroughly discussed in class, and the subjects were provided with a general guide which urged the inclusion of certain factors in the paper, but no



specific format was suggested.

All of the in-class themes were prepared for in advance, but the specific topic was not assigned until the day of writing. The papers were written entirely within one two-hour class period, which allowed adequate time for both a rough and finished draft. These methods made more certain the authorship of each composition. The novel reviews were perhaps less certain in authorship.

Several means of acquiring objectivity in evaluating the compositions were contemplated, discussed, and discarded. Finally, the experimenter became the sole judge of the subjects' compositions, necessitating the formulation of a reasonably objective, if arbitrary, means of placing a score upon each.

## II. THE CRITERIA

In an effort to reduce the subjectivity of composition scoring to a point that it would not unduly influence the total grade, several steps were taken. First, the composition scores were recorded by number, not by name, to avoid the effect of personalities. Next, all compositions were scored as objectively as possible by criteria described below. The criteria devised place certain aspects of writing (grammar, thought, expression, purpose, etc.) into arbitrary rank order within three large categories.

Expression. In the first, expression and purpose were combined, as indicated in Table VI, and placed on a scale which tended to isolate these factors into four rather distinct levels of expression--accurate, adequate, inadequate, and poor. The presence of awkward and/or irrelevant passages reduces the score one point for either, two for both, at all but the lowest level. For example, if a theme were accurately expressed (nominally scored ten) but marred by the presence of infrequent awkward passages, the score it received would have been nine; if both awkward and irrelevant passages had occurred, only eight would have been awarded. This method was employed for each level. This manipulation of points allowed some flexibility within the sub-categories but discouraged arbitrary scoring based upon stylistic and other uncontrollable qualities.

Grammar. Structure and grammar were also evaluated using the criteria listed in Table VII. I scored the compositions according to the absence or presence of certain kinds of errors.

The kinds of errors were defined as follows. Minor punctuation errors included misuse of commas (excluding splices), quotation marks, underscoring, and similar conventions. Minor grammar errors were composed of: indefinite pronoun references; disagreement of number; shifts of person, tense, case, gender, voice, etc.; and subordination

TABLE VI  
COMPOSITION EVALUATING CRITERIA  
PART I: EXPRESSION AND PURPOSE

Score	Description
10	Accurate expression: good diction; clear progression of thought toward a purpose stated or implied; no noticeable irrelevancies or digressions; smooth transition.
9	As above, but with infrequent awkward passages.
8	As above, but with infrequent irrelevancies or digressions.
7	Adequate expression; adequate diction; clear progression of thought toward stated or implied purpose (somewhat less continuity); adequate transition.
6	As above (seven), but with occasional awkward or obscure passages.
5	As above (seven), but with occasional irrelevancies or digressions.
4	Inadequate expression: inadequate diction; apparent progression toward general purpose (much less continuity); harsh transition.
3	As above (four), but with frequent awkward and/or obscure passages.
2	As above (four), but with frequent irrelevancies and/or digressions.
1	Poor expression: inadequate diction; lack of central theme; little thought progression; frequent obscure or incoherent passages; frequent irrelevancies and/or digressions; poor transition.

NOTE: Scores two and three, five and six, and eight and nine are either-or-both quantities. If either one of them is present, the higher score is awarded; if both, the lower score applies.

TABLE VII  
 COMPOSITION EVALUATING CRITERIA  
 PART II: GRAMMAR AND STRUCTURE

Score	Description
9	No errors.
8	Infrequent punctuation errors only.
7	Minor grammar errors.
6	Minor structural errors.
5	Minor grammar <u>and</u> minor structural errors.
4	Gross structural errors.
3	Gross structural <u>and</u> minor grammar errors.
2	Gross structural <u>and</u> minor structural errors.
1	Gross and minor structural errors <u>and</u> minor grammar errors.

NOTE: The presence of errors described above was the basis for evaluation. For further definition of criteria, see the text.

errors. Minor structural errors consisted of nonparallel constructions, misplaced or dangling modifiers, and awkward structures that obscured the sense (other than stylistically). Gross structural errors included simply comma-splices and unintentional sentence fragments. Minor sentence types such as described in many rhetoric textbooks were not considered to be fragments.

Thought. Since written composition contains many intangibles, such as style, interest, etc., which cannot be placed upon an absolute scale of values, a less objective category was advisedly included: thought. This group of factors is presented simply in Table VIII, but it is in reality the most complex of the three. Each of the six qualities (choice of topic, development of topic, style, interest, originality, and logic) was scored solely upon the subjective opinion of the reader. This fact may weaken the objectivity of the scoring system; but to leave any of these qualities out would be to debase the total evaluation.

Each of the six sub-categories in section three was valued at a maximum of two points, but each was graded either zero, one, or two, making possible a range of zero through twelve. The score in this section, added to the other two scores, resulted in the score assigned the particular composition. Each subject's composite score of three individual theme scores was regarded as an indication of his

TABLE VIII  
COMPOSITION EVALUATING CRITERIA  
PART III: THOUGHT

Score*	Criteria
2	Writing style
2	Choice of topic
2	Development of topic
2	Interest
2	Originality
2	Logic
12	Total

\*Each sub-category is valued at a maximum of two points, totaling twelve for the division; a score of zero, one, or two was awarded each composition in each sub-category.

writing ability--for the purposes of this study.

### III. THE VALIDITY

Several questions must arise concerning the validity of this system, however. At least one is pertinent: how accurate is an evaluation which records the presence or absence of a factor, but not its frequency or intensity? The best available method of answering this question was to determine the distribution of scores and compare it with a relatively constant factor: the subjects' scores on the Iowa Test of Educational Development.

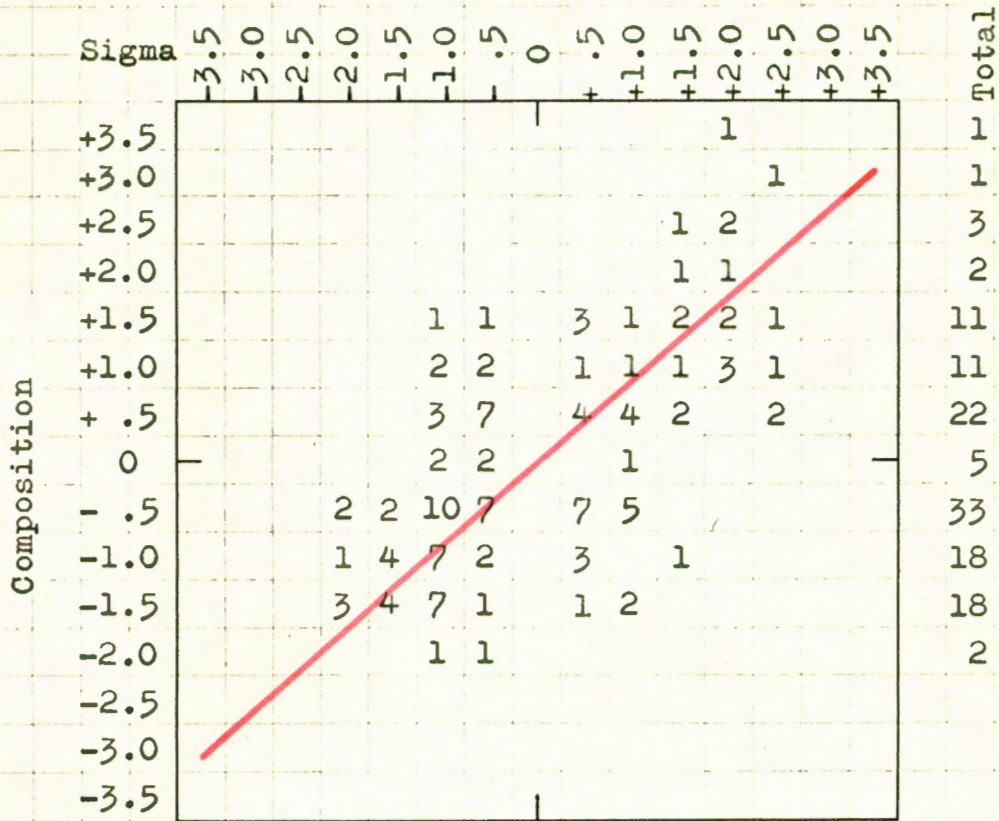
The combined composition scores acquired as a result of the standards described above are shown in Table IX. Of the 140 scores, eighty are below the mean (44), and only fifty-five are above it. The median (42) and mode (40) are slightly below the mean, indicating a skewed distribution much like the one described in Chapter III. Forty-two per cent (59) of the scores are clustered within one standard deviation below the mean, and 55 per cent (78) within one and one-half standard deviations. All minus scores are within two standard deviations. Those scores above the mean show no such concentration, but are distributed broadly across a range exceeding three standard deviations.

Comparison of this distribution to the results of the expression portion of the Iowa test listed in Table I

TABLE IX  
 FREQUENCY DISTRIBUTION OF COMPOSITE  
 COMPOSITION SCORES

Interval	Frequency	Standard deviations
81-86	1	3.5
75-80	1	3.0
69-74	3	2.5
63-68	2	2.0
57-62	13	1.5
51-56	12	1.0
45-50	23	.5
44	5	0
38-43	37	- .5
32-37	22	-1.0
26-31	19	-1.5
20-25	2	-2.0
14-19		-2.5
8-13		-3.0
2- 7		-3.5
Total	140	
Mean	44	
Median	42	
Mode	40	
Sigma	12	





Totals            6 103323    1914 8 9 5            127  
 Correlation coefficient .87

Iowa Expression

FIGURE 1

CORRELATION OF COMPOSITION SCORES AND  
 SCORES ON IOWA TEST OF EDUCATIONAL  
 DEVELOPMENT TEST NUMBER THREE,  
 CORRECTNESS OF EXPRESSION

shows marked similarities. A correlation of these factors indicating a strong tendency for the expression scores on the standardized test to vary with the composition scores acquired by the criteria listed above is presented in Figure 1. The correlation coefficient of .87 (computed by the Pearson Product Moment formula) is graphically shown by the regression line, which was plotted from the formula. This substantially confirms the validity of the criteria employed in evaluating the subjects' compositions for this study.

Whether each criterion was located in the best category and whether each was given the proper emphasis in the scoring procedure is problematical. The criteria do make possible a scoring system based upon the presence of fixed qualities and thereby take the scoring of the subjects' writings out of the realm of personal impressions.

## CHAPTER VI

### CORRELATIONS OF THE INDIVIDUAL TESTS AND GENERAL RESULTS

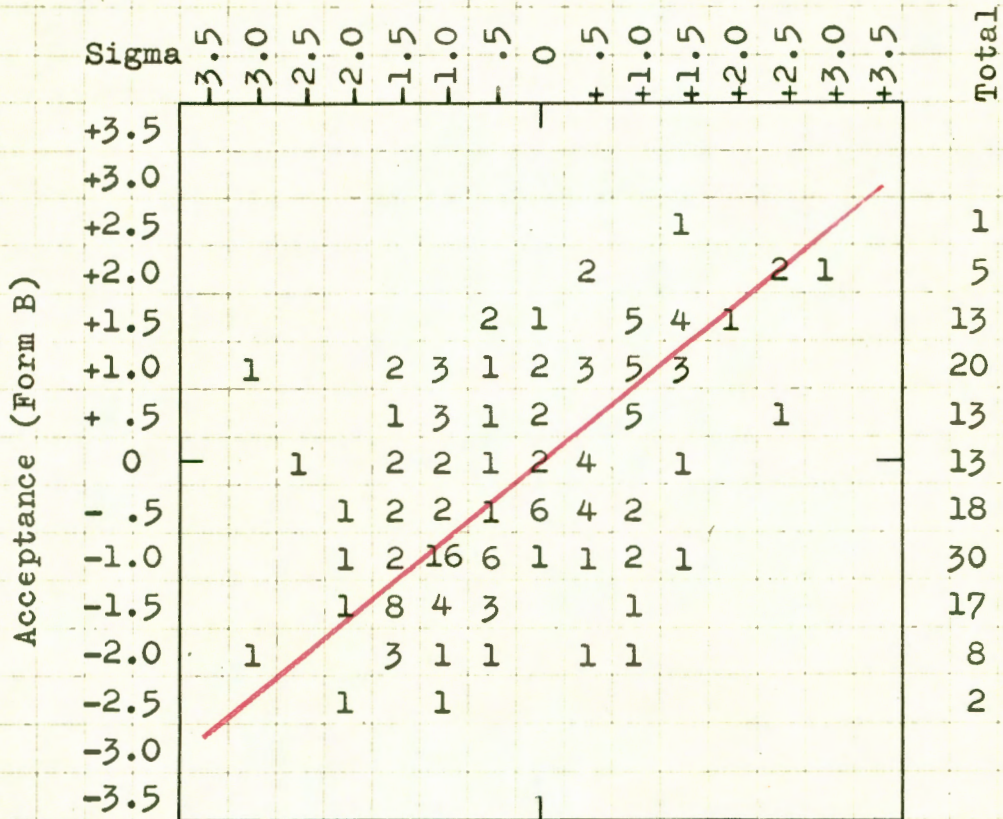
The preceding chapters described the methods employed in acquiring the data upon which this chapter was based. In this chapter the accumulated data will be correlated. And the results of these correlations will be discussed.

#### I. THE INDIVIDUAL CORRELATIONS

Methods. The following correlations were all computed by the Pearson Product Moment  $r$  formula, and all correlation coefficients so calculated are rounded off to the nearest hundredth point. Regression lines plotted on the scatter diagrams were drawn from the computed coefficients, and they represent the relationship of the vertical to the horizontal scale.

Two acceptance forms. The correlation of forms A and B of the acceptance test is displayed in Figure 2. The scatter diagram of the two forms shows a distribution familiar from the Iowa scores discussed in Chapter III--a concentration of scores within two standard deviations below, and a broad dispersion of fewer scores above the means. Also obvious are the slightly bimodal tendency and the narrower





Totals 2 1 4 20 32 16 14 15 21 10 1 3 1 140  
 Correlation coefficient .82

Acceptance (Form A)

FIGURE 2

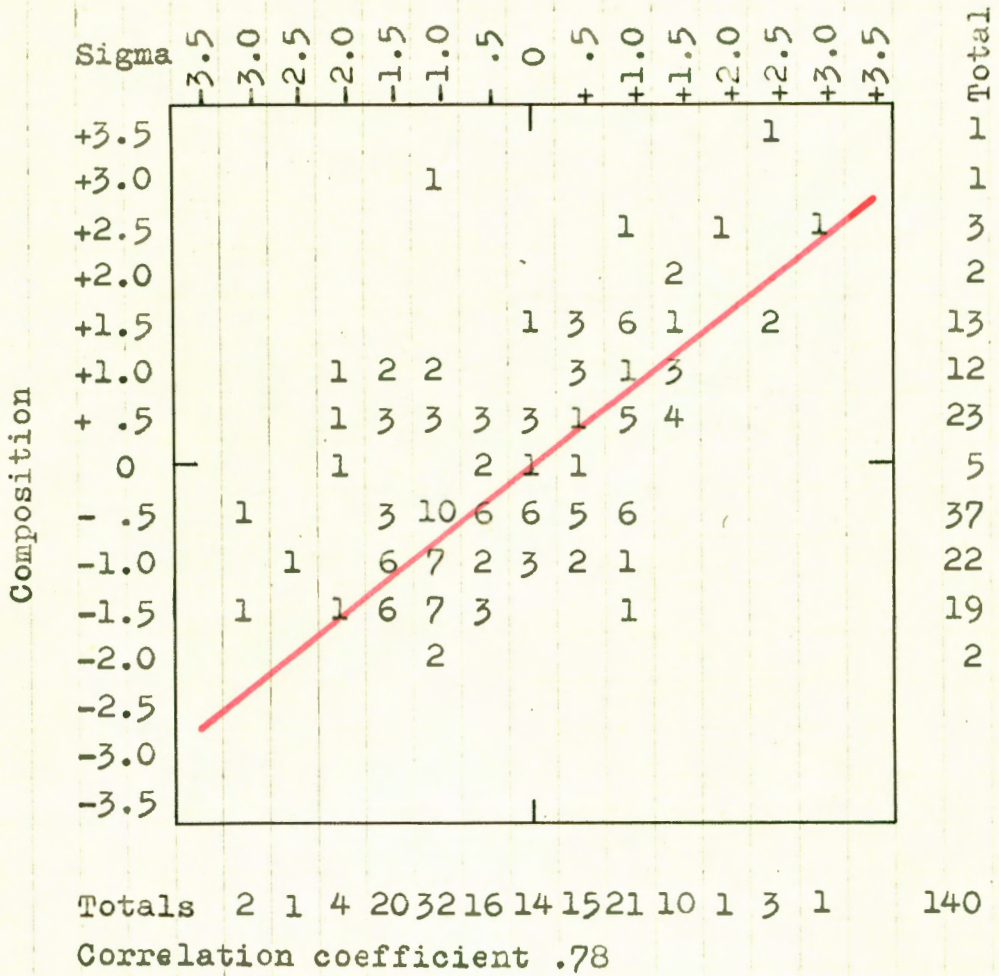
CORRELATION OF ACCEPTANCE SCORES,  
 FORM A AND FORM B

range present in form B.

The correlational coefficient of .82 is not high for separate forms of the same test, but these forms were not designed to be compared directly; they were to determine which method was more highly correlated and also to provide a means of avoiding error. The differences in administering the materials and the differences in instructions were certain to show in such a comparison. This correlation, then, reflects the tendency of the two forms to vary together in spite of the variations in method, and in this respect, the coefficient is not surprisingly low, but rather, remarkably high--even when considering the similarity of materials.

Composition and form A. The scores on this simpler form of the acceptance test when juxtaposed with the composition scores produced the expected results. Figure 3 emphasizes the strong correlation of acceptance and writing ability expressed in the coefficient, .78 (.776). Prominent in this scattergram is the familiar bimodal pattern which resulted in two clusters of scores along the regression line, both within one and one-half standard deviations above and below the means. However, by far the greater number (50) are in the group below, and a relatively small number (27) above the means. Those scores outside the higher group are widely dispersed along the regression line, while those below tend to deviate from it.





Acceptance (Form A)

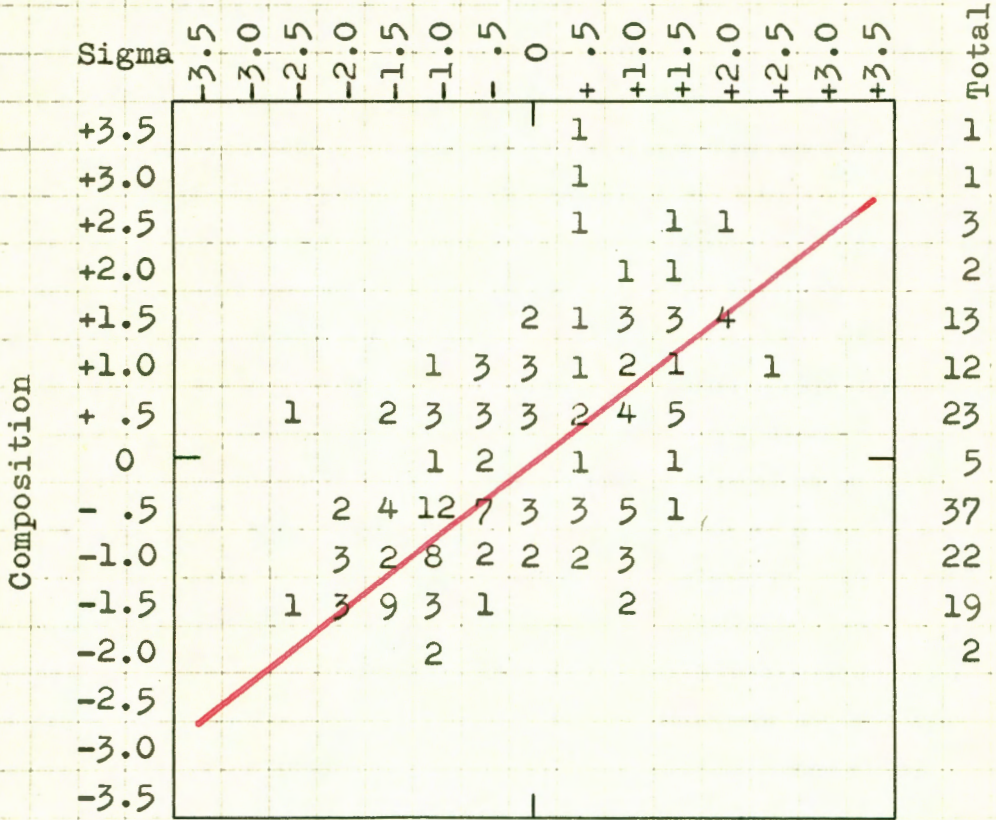
FIGURE 3

CORRELATION OF COMPOSITION SCORES  
AND SCORES ON ACCEPTANCE  
TEST, FORM A

Composition and form B. The .78 (.782) correlation coefficient of form B with composition is very similar to that of form A, indicating that the different methods didn't greatly affect the relationship. However, this form of the acceptance test was supposed to reflect more conservative judgment of grammaticality than was form A. A comparison of the scattergrams in Figures 3 and 4 points out clearly that the extreme scores have indeed been influenced by the differing directions and methods. The one extremely high (plus three standard deviations) composition score which fell below the acceptance mean in Figure 3 has shifted to just above the mean in Figure 4. Also, the other high composition scores in Figure 3 have moved toward the acceptance mean on form B. Apparently the extremes were more influenced by the altered methods than were the middle scores. The general configurations remained the same; the slight bimodal tendency in form A was retained in form B.

Composition and form AB. The sum of forms A and B was the score which had been originally designed to more accurately reflect the subjects' acceptance of the materials submitted for judgment. Determining its relationship to composition is then the primary purpose of the study. Figures 3 and 4, already discussed, present the weighted forms of the test. Theoretically, a combination of the two





Totals            2   8 17 30 18 13 13 20 13 5 1            140  
 Correlation coefficient .78

Acceptance (Form B)

FIGURE 4  
 CORRELATION OF COMPOSITION SCORES  
 AND SCORES ON ACCEPTANCE  
 TEST, FORM B

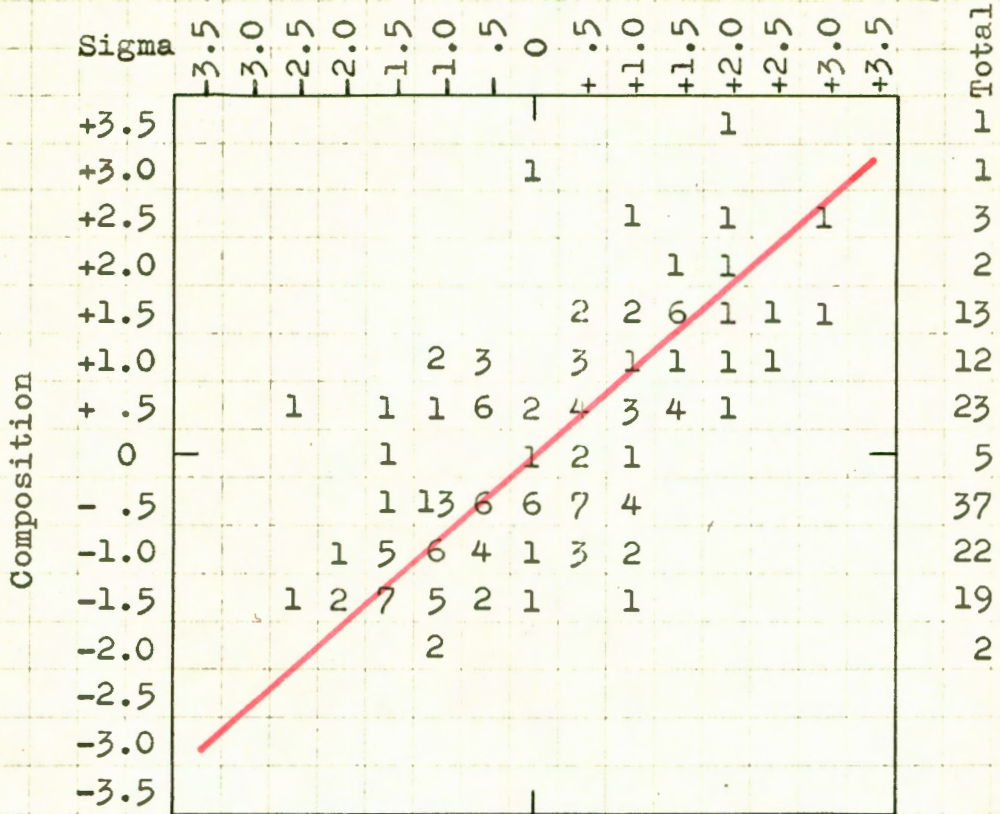


would nullify the errors, and the combined score would more closely parallel the factor measured by both.

The frequency distribution of both forms suggests that the subjects responded to the biased directions in the expected fashion; this correlation confirms it. The correlation expressed in Figure 5 indicates that while the two forms separately correlate substantially with composition, the combined scores do so more highly still: the correlation rose to .87.

A comparison of Figures 4 and 5 shows that the highest scores, which were apparently more subject to method changes, were nearly restored to their original (form A) positions near the regression line; the extremely low scores were relatively unaffected by the combination of the scores. Another notable change brought about by the merger of the two forms is the change in the distribution from slightly bi-modal to near normal (although still skewed to the left).

The Iowa and form AB. The Iowa Test of Educational Development, originally employed as a measure of the subjects' abilities and as a means of validating the composition criteria, affords an opportunity to add a further correlation pertinent to the study. Both the Iowa test and the acceptance test correlate .87 with composition--a fact which invites their comparison. The data represented in Figure 6 make it apparent that there is a far less



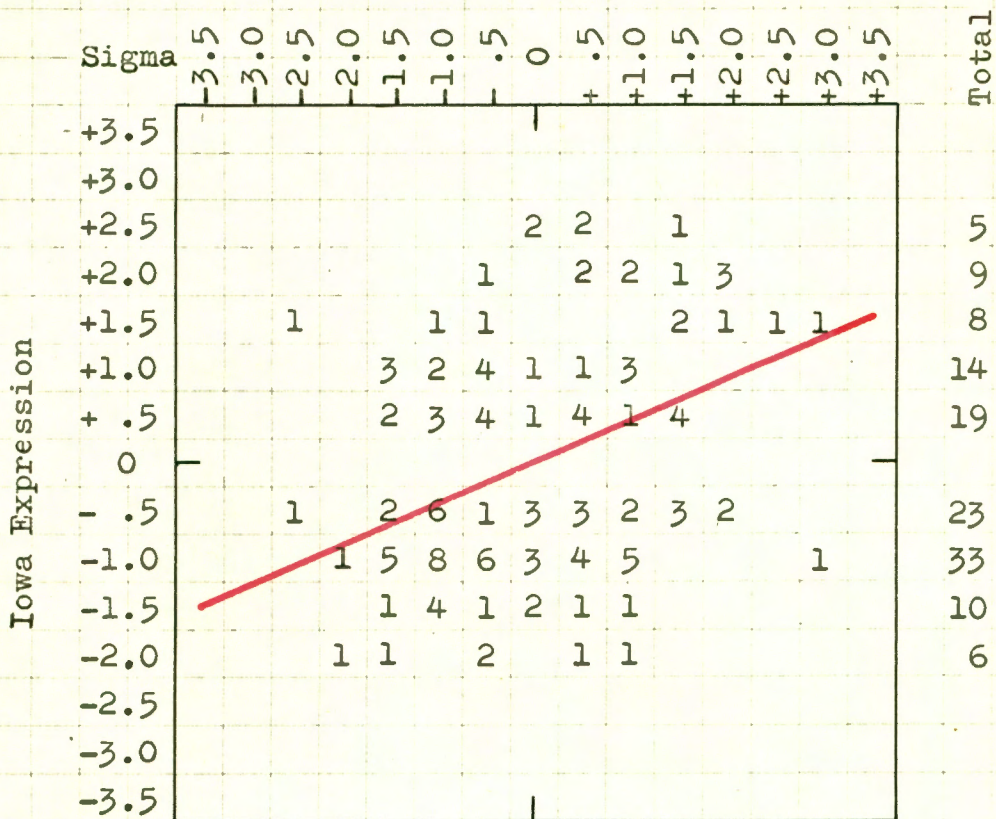
Totals        2 3 15 29 21 12 21 15 12 6 2 2        140  
 Correlation coefficient .87

Acceptance (Form AB)

FIGURE 5

CORRELATION OF COMPOSITION SCORES  
 AND SCORES ON ACCEPTANCE  
 TEST, FORM AB





Totals 2 2 14 24 20 12 18 15 11 6 1 2 127  
 Correlation coefficient .43

Acceptance (Form AB)

FIGURE 6

CORRELATION OF ACCEPTANCE SCORES FORM  
 AB AND COMPOSITE SCORES ON IOWA  
 TEST OF EDUCATIONAL  
 DEVELOPMENT

significant relationship between these two measuring instruments than there is between each and what it measures. This observation is born out by the coefficient of correlation-- a moderate .43. Still apparent in the distribution is the clustering of lower scores and dispersion of the higher ones; this was true of all distributions in the series-- apparently indicative of the below average characteristics of the group.

## II. SOME GENERAL CONCLUSIONS

The correlations of the various data substantially support the original postulate--that those subjects who write well will also accept a wider range of relatively meaningless utterances as grammatical. The results of the separate and combined acceptance scores--distributions and their own interrelationships, as well as their high correlations with composition--all serve to confirm both the means of determining acceptance and the original assumption; and the Iowa tests support the composition criteria.

## CHAPTER VII

### SUMMARY AND CONCLUSIONS

#### I. SUMMARY

Two facets of language behavior, achievement in written composition and acceptance of certain word sequences as grammatical, were postulated to be responses governed by the grammatical sense used by native speakers to form their own utterances and to judge the utterances of others. It was further postulated that if both responses are indeed influenced by the same sense, the two responses will be similar in some ways; for example, since the subjects' compositions indicated a wide range of abilities in writing, a similar range of acceptance scores was tentatively predicted. This investigation was intended to determine whether the subjects would respond to the test sequences in a pattern or range which was similar to that of their writing abilities, and also to what extent the two response patterns are inter-related. But first, means of measuring both responses were necessary.

The necessity of measuring the subjects' writing ability and acceptance with a reasonable degree of accuracy required that certain standards be established. Arbitrary criteria were employed as guidelines in an attempt to ensure as much reliability as was possible in scoring the composi-

tions, but even these provide only moderate assurance that the scoring was consistent, obviously. And the resulting scores cannot be assumed to be absolute indications of the subjects' writing abilities; but by totaling each subject's scores on three separate compositions, individually only measures of performance on specific assignments, a reasonably accurate approximation of the subjects' writing abilities was possible. The strong (.87) correlation of the subjects' composite scores arrived at by these criteria and their scores on the Iowa Correctness of Expression Test substantiates to some degree the validity of the criteria, and the scores themselves.

In an effort to simplify the grammaticality tests, each subject's acceptance of the sequences rather than his ability to discriminate between grammatical and ungrammatical sequences was measured. In accordance with Chomsky and Lenneberg's views, relatively meaningless sequences were employed. These theoretically encouraged intuitive rather than rational judgments, and reduced the influence of meaning upon the individual judgments. The number of sequences each subject accepted was interpreted to reflect his "acceptance"--his willingness to accept those utterances as grammatical.

Just as would be expected, a range of acceptance scores resulted. But what was actually measured by this

test? It really measured only the subjects' acceptance of a specific group of nonsensical word sequences. However, for the purposes of this study a subject's acceptance was considered to be indicative of his acceptance of all such utterances. Further, the range of acceptance scores appeared to suggest that a range of sensitivity to, or ability to discern, grammaticality in the sequences may exist. Clearly, more need be learned before these assumptions can be stated with any confidence, but the implication is there.

Correlation of the composition and acceptance scores indicated that a strong relationship exists between the two factors. Individually, the acceptance tests correlate .78 (.776 and .782) with composition; together, the coefficient is considerably higher (.87). The strength of the correlations confirms the original postulate that the two language responses are similar, and closely related.

The composite coefficient also confirms the methods employed in administering the separate forms of the acceptance test; that it was higher than the individual coefficients also justifies employing the test-retest method using different tests. Although the use of two forms is usually designed to test the reliability of the instrument, the distinctly different forms precluded such a test. But it did accomplish the desired result: two separate measurements which, together, compensated for some of the subjects'

uncertainty--probably of the criterion, grammaticality--and achieved a more accurate measure as a result. This assumes that no higher correlation than that which exists--in the specific group--will be measurable, while faulty experimental design may certainly result in a lower coefficient. But although no reliability measure was attempted, the correlation of the subjects' scores on form A with form B resulted in a coefficient (.82) which supports the earlier supposition that the range of acceptance scores may indicate either a varying propensity to accept such sequences, or a range of sensitivities to the nuances of language, i.e., an ability to detect or recognize the grammaticality of utterances.

Although the different forms of the acceptance test served their purpose, their use proved to be a disadvantage in another way. The several changes in both the directions and the administering of the test items obscure the causes of the changes in response which occurred on form B. Had one factor been manipulated at a time, the shifting of the extreme scores might have been explained; but, this exceeds the scope of the present study.

## II. CONCLUSIONS

This study confirms, in many respects, the findings of Maclay and Sleator in that it, also, shows that people respond differently to language stimuli. But, contrary to



their suggestions, the differences in the responses are not due to mere chance. A pattern has emerged which suggests strongly that the ability to detect grammaticality varies widely from one individual to another.

The wide range of acceptance scores is one indication of this. Several subjects accepted as few as six of the test sequences. Others accepted as many as twenty-five. This range indicates that language judgments do vary; the correlation coefficient confirms that they vary in a very consistent pattern: a subject's acceptance of the test sequences tends to change in proportion to his ability in written composition. The original postulate that the two are closely related is confirmed, and this, in turn, supports the view that an individual's language behavior is governed by his sense of grammaticality--that this sense may be his guide in writing, as well as in speech.

The ranges of scores and the correlation coefficients which witness the similarity of the distribution of those scores bear out the additional conclusion that a range of ability to recognize the grammaticality of an utterance may exist. The results of this study direct the inquiry away from the utterance as the sole stimulus of the response and toward the subjects as individuals. Those subjects who write well also tended to accept more of the sequences than those who do not. The subject who scored highest in compo-

sition also accepted the highest number of sequences, twenty-five--over four times as many as several poorer writers.

Whether the former is more sensitive to the cues intrinsic in the sequences than are the latter cannot be absolutely determined from the present data; but it is abundantly clear that the subjects have within them the determining factor which influences their judgments, the factor which makes them more or less willing to accept the test sequences. This is certainly suggested by the data, and the divergent judgments which have as yet remained unexplained may be clarified in this fashion. The views discussed earlier concerning grammatical judgments are oriented toward the utterances. They assume only that some factor within the utterance influences acceptance or rejection. In this view, the deviant judgments are explained by the tendency of subjects to see meaning even where it is not, and to judge the utterance on that basis. As stated before, meaning as a basis for judgment only explains the acceptance of ungrammatical utterances and the rejection of grammatical but meaningless ones; it cannot explain the rejection of normal utterances. And that view cannot explain the correlation of writing with grammatical judgments without postulating a general range of language ability, for judgments based upon chance meanings derived from the utterances will

not likely correlate so highly with writing skill. Contextual and semantic properties cannot explain the judgments of the thirty-two "naive native speakers" in the present study who rejected the sequence "steel yourself to the water" as ungrammatical; nor the twenty-two who rejected "a collision made the scene a shambles" and "we usually have a battery of full voices" (see Table III, page 22). And similar responses were recorded by Maclay and Sleator .

In the absence of an explanation for such diverse judgments, Maclay and Sleator have proposed the use of statistical procedures to compensate for the subjects' inconsistencies. But while statistics may compensate for such judgments, it fails to explain them. And linguistic significance based upon the manipulation of numbers is unconvincing when such inconsistencies remain unexplained. Another proposed variable by which the deviant judgments are explained away is the various interpretation of the criterion, grammaticality, that the subjects may not interpret the criterion in the same way. It is, no doubt, very true that this is a problem in this type of study and that it would appear to logically explain such judgments--until the correlation of judgments and writing is introduced. It is extremely unlikely that those who misinterpret the criterion by chance will pattern so consistently. But it is little wonder that varied interpretations should occur, since

grammaticality is a complex idea that few "naive native speakers" would fully understand.

Of what value are grammatical judgments to the study of language? More data is necessary before that question can be answered. Perhaps further study should be subject oriented; perhaps the subjects should be selected according to their ability in other areas of language behavior. Those who write well, for instance, seem to be more aware of grammaticality. And selection would probably reduce the varied interpretations of the criterion and isolate those with more fully developed "sets of rules." For although a select group of informants contradicts the tenets of statistical procedure, judges of jurisprudence are not selected by lot, but by experience in their profession.

But even these judgments would probably require statistical treatment, and are therefore unsatisfactory; statistical relevance seems a poor criterion by which to test a proposed grammar which can theoretically generate all grammatical utterances and no ungrammatical ones. It appears to the writer that the "all" and the "no" in the above are absolute terms which preclude any statistical "relevancies."

This study has linked grammatical judgments firmly with written compositions. By this association it suggests strongly that such judgments are governed by the same factors

that influence writing, and that those judgments will thus vary from individual to individual much as writing ability does.

Additional research is necessary in the area of intuitive grammar. More sophisticated studies with adequate controls, single variables, and balanced groups of subjects should add a great deal to our knowledge of language. But a more immediate benefit may result. While much remains to be done, it is conceivable that additional research may produce a simple and rapid means of determining writing ability, based upon judgments of grammaticality. (And the simplicity of machine-scoring yes/no answers needs no elaboration here.) The value of grammatical judgments to the precise study of language, however, appears to be extremely limited. Clearly, the questions raised by Hill and Fries in this respect are, as yet, unanswered.

## BIBLIOGRAPHY

## BIBLIOGRAPHY

1. Braddock, Richard, Richard Lloyd-Jones, and Lowell Schoer. Research in Written Composition. Champaign, Illinois: National Council of Teachers of English, 1963. 142 pp.
2. Chomsky, Noam. "Some Methodological Remarks on Generative Grammar," Readings in Applied English Linguistics. 2d ed. Harold B. Allen, editor. New York: Appleton-Century-Crofts, 1964. Pp. 173-192.
3. \_\_\_\_\_. Syntactic Structures. The Hague: Mouton and Company, 1957. (Janua Linguarum, IV) 118 pp.
4. Francis, W. Nelson. The English Language An Introduction. New York: W. W. Norton and Company, 1963. 273 pp.
5. Fries, Charles C. Linguistics and Reading. New York: Holt, Rinehart and Winston, Inc., 1962. 264 pp.
6. Hill, Archibald A. "Grammaticality," Readings in Applied English Linguistics. 2d ed. Harold B. Allen, editor. New York: Appleton-Century-Crofts, 1964. Pp. 163-172.
7. Lenneberg, Eric H. "Language Evolution and Purposive Behavior," Culture in History: Essays in Honor of Paul Radin, Stanley Diamond, editor. New York: Columbia University Press, 1960. Pp. 869-893.
8. Lloyd, Donald. "Structure in Language," College English, 24:598-602, May, 1963.
9. Maclay, Howard, and Mary D. Sleator. "Responses to Language: Judgments of Grammaticalness," International Journal of American Linguists, 26:275-282, April, 1960.
10. Sledd, James. A Short Introduction to English Grammar. Chicago: Scott, Foresman and Company, 1959. 346 pp.
11. "Your Scores on the Iowa Tests of Educational Development and What They Mean." 4th ed. Pamphlet, catalog number 7-773, Science Research Associates, Chicago, 1958. 1 p.